Microbiota-based models enhance detection of colorectal cancer

by

Nielson T. Baxter

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Microbiology and Immunology)
in the University of Michigan
2016

Doctoral Committee:

Associate Professor Patrick D. Schloss, Chair
Assistant Professor Grace Y. Chen
Professor Thomas M. Schmidt
Associate Professor Vincent B. Young

“Do not let yourselves become tainted by a deprecating and barren skepticism. Do not let yourselves be discouraged by the sadness of certain hours which pass over nations. Live in the serene peace of laboratories and libraries. Say to yourselves first: 'What have I done for my instruction?' and, as you gradually advance, 'What have I done for my country?' until the time comes when you may have the immense happiness of thinking that you have contributed in some way to the progress and to the good of humanity.”

- Louis Pasteur

ACKNOWLEDGEMENTS

This work was made possible by an excellent group of colleagues, collaborators, friends, and family. Thank you first and foremost to my thesis advisor Pat Schloss for his leadership and mentorship. Pat gave me the independence to thrive in my research, while always being available for guidance at a moment's notice. I am especially grateful that he had the confidence to involve me in many diverse projects and that he trusted me enough to let me abandon projects when I felt they weren't worthwhile.

I am grateful to all of the members of the Schloss lab past and present. Thank you to Joe Zackular, whose thesis work and mentorship laid the foundation for this dissertation. Thank you to Matt Jenior, Jhansi Leslie, Alyx Schubert, and Kathryn Iverson. Their intellectual support and friendships have been invaluable for the last 5 years and hopefully for years to come. Thank you to Jim Kozich, Hamide Sinani, and Charlie Koumpourus for their technical assistance with my research. Thanks to all of the other Schloss lab members who have made this such an enjoyable journey and for never letting science get in the way of a good lab meeting.

I have been very fortunate to be a part of an exemplary institution and department. I benefitted from outstanding mentorship from the likes of Michele Swanson, Nicole Koropatikin, Joel Swanson, and many others. Thank you to those who have collaborated with me on various projects, namely Phil Myers, Simeone Marino,

Alex Rickard, and Nicole Koropatikin. Thank you to Mack Ruffin, Mary Rogers, and Grace Chen who were essential contributors to the work in this dissertation. I am especially grateful for the guidance of my thesis committee, Vince Young, Grace Chen, and Tom Schmidt. Thank you to Grace for introducing the Schloss lab to the field of colorectal cancer. Her technical and clinical expertise have been critical to our work. Thank you to Tom for his continued mentorship in both undergraduate education and microbiological research. Thank you also to those who introduced me to biomedical research at Indiana University, especially Cheng Kao, Ranjith Kumar, and Rob Vaughan.

Finally, thank you to all of the friends, family, teachers, and mentors who have helped and guided me throughout my life. Thank you to my parents, Stewart and Linda Baxter, whose selfless hard work has provided me with the opportunities to pursue any career I could choose. Thank you also to Mike and Lorie Heneghan for their constant guidance and support. Thank you to my wife, Sarah, for her love and friendship and for providing the benchmark of personal, academic, and professional excellence that has inspired me to work to my fullest potential.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: Introduction

The gut microbiota and colorectal cancer

The human microbiota is the collection of microorganisms that live in and on the human body. Only recently has the scientific community begun to appreciate the important roles that these bacteria, archaea, viruses, and fungi and other eukaryotes play in regulating human health. The gastrointestinal tract is the most densely populated part of the human body, and one of the highest densities of microbial life on the planet [1]. This diverse collection of organisms harbors an even more diverse catalog of genes that encode for roughly 100 times more unique genes than the human genome. This functional diversity allows the gut microbiota to compliment its host in a number of ways, including synthesis of vitamins, degradation of resistant starches, immune system maturation, and resistance to pathogens [2–5]. Due to its important role in human health, disruptions to the microbiota have been associated with a variety of diseases, such as Crohn's disease, ulcerative colitis, enteric infections, diabetes, obesity, malnutrition, and of particular interest to this dissertation, colorectal cancer (CRC) [5–9].

A growing body of evidence has demonstrated that the gut microbiota plays an important role in CRC. The gut microbiota is known to influence cancer-related functions like cell proliferation, angiogenesis, and apoptosis [10–12]. It is also linked to diet, obesity and inflammatory bowel disease, which are known risk factors of CRC [6, 8, 13–

15]. Moreover, at least thirty studies in the past decade have identified alterations in the microbiota associated with CRC [16]. Animal models have shown that cancer-associated changes in the microbiota can promote tumorigenesis, and that several bacterial species are capable of promoting colon tumorigenesis on their own [9, 17–19].

Microbiota composition influences tumorigenesis

To better understand the role of the microbiota in CRC, our lab turned to a chemically induced mouse model of colitis-associated CRC. In this model, mice are given a single intraperitoneal injection of the carcinogen azoxymethane (AOM), followed by three five-day rounds of 2% dextran sulfate sodium (DSS) administered ad libitum in the drinking water. This model works well because it is quick and the mutations and tumor progression mirror those of human CRC [20, 21]. Our lab hypothesized that the community as a whole, more so than any one species, is responsible for modulating colon tumorigenesis. Using the AOM/DSS mouse model, we observed shifts in the composition of the microbiota during tumorigenesis [9]. To test whether such shifts influenced tumorigenesis, tumor-associated or normal stool was transferred to germ-free mice that were then subjected to AOM/DSS treatment. Mice inoculated with the tumor-associated microbiota developed significantly more tumors than those inoculated with a healthy microbiota [9]. This phenotype occurred in the absence of any known tumorigenic pathogens and suggested that the microbiota could modulate tumorigenesis. More recently, we showed that manipulation of the microbiota using different combinations of antibiotics could vary the number of tumors mice developed, and a three-antibiotic cocktail of vancomycin, streptomycin, and metronidazole could all but block tumorigenesis entirely [22]. Furthermore, we could predict the number of

tumors mice would develop based on the composition of the microbiota during antibiotic treatment. These findings suggested that the composition of the microbiota influences susceptibility to tumorigenesis.

Using antibiotics to manipulate the microbiota is limited in two important ways. First, treating mice with antibiotics primarily alters the abundance of populations already present in the community. Membership should be less affected, especially for mice living in an isolated environment in which they are exposed to a limited number of species. Second, antibiotics can directly impact the immune system of the host, independent of the microbiota [23]. This could in turn influence tumorigenesis. To address these concerns while further investigating how the composition of the microbiota influences tumorigenesis, we turned to a germ-free mouse model. Chapter 2 of this dissertation describes an experiment in which germ-free mice were inoculated with stool from one of six human donors and then subjected to the AOM/DSS model of CRC [24]. This method allowed us to study highly distinct microbiota structures without the confounding effects of antibiotics. Furthermore, it allowed us to investigate microbial populations of human origin, which are likely of greater clinical relevance. The findings from that study demonstrated that the initial composition of the microbiota may influence an individual's susceptibility to CRC.

CRC Screening

The observation that microbiota composition differs in patients with CRC opens up the possibility of using shifts in the microbiota as a biomarker of tumor development. Early detection of tumors is key for CRC prevention and treatment. Patient's in whom CRC is detected at stage I, have greater than 90% of survival, whereas patients in whom

tumors are detect at stage IV have less than 13% chance of survival [25]. Great progress has been made in reducing CRC incidence and mortality through increased surveillance, yet CRC remains the second leading cause of death among cancers in the United States [26]. This is due in part to a continued lack of adherence to screening guidelines, as a third of individuals fail to receive appropriate screening [27]. Screening adherence is drastically lower among those who are uninsured, suggesting that cost is one barrier to screening [27, 28]. Patients also commonly cite fear and/or discomfort as deterrents to undergo structural exams like colonoscopy and sigmoidoscopy [29]. Therefore, there is a need for screening alternatives that are inexpensive and noninvasive.

For many years the most common noninvasive screening method for CRC was the guaiac-based fecal occult blood test (gFOBT), which detects blood in stool. The gFOBT is largely being replaced by the newer more accurate fecal immunochemical test (FIT), which also detects blood, but is specific to the human globin protein [30]. Both gFOBT and FIT are inexpensive and noninvasive, however neither is able to reliably detect early stage tumors, particularly precancerous adenomas. The relatively new multitarget DNA test (MT-sDNA) combines FIT with host-associated DNA markers (i.e. KRAS mutations, aberrant methylation patterns, β-actin) and has improved sensitivity compared to FIT or gFOBT [31]. Unfortunately MT-sDNA is much more expensive than FIT or gFOBT, costing as much as $649 in out-of-pocket expenses (http://www.cologuardtest.com/what-to-expect-with-cologuard/faq/cologuard-cost-how-much-is-cologuard). Thus, there remains a need for inexpensive, noninvasive tests that

can reliably detect early stage tumors. The microbiota could potentially be source of biomarkers for such a test.

Individual species as biomarkers of CRC

Many bacteria have been associated with CRC, but only a few have been shown to be potential drivers of tumorigenesis. One such bacterium is enterotoxigenic *Bacteroides fragilis* (ETBF). ETBF differs from other *B. fragilis* strains by the secretion of a toxin, called fragilysin. ETBF is capable of promting Wnt/β-catenin signaling by cleaving E-cadherin, inducing NF-κB signaling, and initiating a Th17 immune response, all of which enhance colon tumorigenesis [17, 32]. Inoculation of $APC^{Min/+}$ mice with ETBF results in accelerated tumorigenesis and shifts the location of tumors from the small intestine to the distal colon [17]. Evidence for ETBF as a biomarker for CRC is limited, but one study detected the fragilysin gene in Bacteroides isolates from the stool of 38% of CRC patients compared to 12% of controls [33]. Another study isolated Bacteroides from mucosal samples and found that fragilysin positive isolates were more common in CRC patients compared to controls [34]. However, both studies were limited by having a small numbers of patients and by relying on culturing isolates for detection.

Similar to ETBF, certain strains of *E. coli* may have the potential to promote colon tumorigenesis. Some *E. coli* harbor a polyketide synthase (*pks*) genotoxic pathogenicity island, which is capable of causing DNA double-stranded breaks in eukaryotic cells. Arthur et al. showed that *pks*+ *E. coli* increased DNA damage and increased tumor multiplicity in Il10-/- mice treated with AOM [18]. Deletion of the *pks* island reduced DNA damage and tumor multiplicity, but not inflammation. In the same study, they isolated the *pks*+ *E. coli* in 14 of 21 CRC patients and 14 of 35 IBD patients, compared to 5 of 24

controls, suggesting *pks*+ *E. coli* could be a marker of CRC or colonic inflammation in general.

In the many studies comparing the microbiota of healthy individuals to those with CRC, *Fusobacterium nucleatum* is the most commonly enriched species in CRC patients. Kostic et al. found multiple *Fusobacterium* species enriched in colorectal carcinoma tissue biopsies relative to adjacent healthy tissue [35] and an enrichment of *F. nucleatum* in the stool of patients with colorectal adenomas relative to healthy controls [19]. They went on to show that *F. nucleatum* can accelerate small intestinal tumorigenesis in $APC^{Min/+}$ mice via recruitment of myeloid cells and activation of NF-κB. Concurrently Rubinstein et al. found that binding of the *F. nucleatum* adhesion FadA to E-cadherin activates β-catenin signaling and downstream activation of NF-κB [36]. Together these studies demonstrate that *F. nucleatum* is capable of promoting intestinal tumorigenesis. *F. nucleatum* also shows promise as a biomarker for CRC. Using qPCR, Kostic et al. detected *F. nucleatum* in the stool of 27 out of 27 cancer patients with carcinomas, 24 of 28 patients with adenomas, and 15 out of 31 healthy individuals. Many other studies have observed an enrichment of *F. nucleatum* in CRC patients relative to controls, making it arguably the best single-species biomarker for CRC [37–40].

Multi-species models for detecting CRC

Attempts to use individual species as biomarkers for CRC have had limited success, as none of the strains mentioned above is present in the majority of CRC cases. To address this problem, several groups have attempted to use a combination of bacterial species to differentiate healthy individuals from those with CRC. Zackular et al.

demonstrated that logit models using 5 or 6 bacterial species could detect carcinomas just as well as a model using traditional risk factors (age, race, and BMI) and could detect adenomas better than those risk factors [40]. They also showed that combining bacterial species with patient risk factors and/or gFOBT could improve the detection of CRC over any one method by itself. Among the most predictive populations were ones associated with *Fusobacterium* and *Porphyromonas*. Limitations of this study were that it included only 90 patients (30 normal, 30 adenoma, 30 carcinoma) and did not perform any validation of the models.

Shortly thereafter, Zeller et al., with a larger patient cohort, developed a model using the relative abundance of 22 bacterial species that could detect CRC better than gFOBT [39]. Again, they found that combining a microbiota-based model with gFOBT improved detection of CRC over either method alone. Importantly, they validated their model by applying it to two external datasets. The bacterial species that were most predictive of CRC were *Porphyromonas asaccharolytica*, *Petpostreptococcus stomatis*, and two strains of *F. nucleatum*. Also of interest, was the observation that models based on metagenomic markers were no better than models based on 16S rRNA gene abundances.

In 2015 Yu et al. developed a model based on 20 microbial gene markers [37]. Genes associated with *F. nucleatum*, *Peptostreptococcus stomatis*, and *Parvimonas micra*, were among the most predictive biomarkers. Building from these findings, Chapter 3 of this dissertation describes microbiota-based models we developed for detecting CRC based on the microbiota [41]. We showed that combining bacterial abundances with fecal hemoglobin concentrations from FIT could further improve the sensitivity for CRC,

especially early stage lesions. Like the aforementioned studies, we found that oral species, including *F. nucleatum*, *P. asaccharolytica*, *P. micra*, and *P. stomatis*, were the most overrepresented in CRC samples.

Combining FIT with other biomarkers is not a new idea. The multitarget stool DNA test combines FIT with several other host-associated biomarkers into a model for determining whether an individuals has CRC. One barrier to combining microbiota-based screening with other tests is the need to collect and store an additional stool sample for microbiota characterization. Patients who perceive stool collection as inconvenient or indelicate may be dissuaded from adhering to screening guidelines. Furthermore, stool collection could have added financial costs from collection materials, processing, and storage. Thus, an alternative source of material for microbiota analyses is needed.

Sinha et al. compared several sampling methods, including used gFOBT cards and rectal swabs. Despite some biases between methods, they found that gFOBT cards offered a high level of stability and reproducibility as source of fecal material for 16S rRNA gene sequencing [42]. This showed that microbiota analyses and blood detection could be carried out from a single convenient sample. However, gFOBT is largely being replaced by the more accurate FIT method. Therefore an alternative would be to utilize the residual fecal material contained within the FIT sampling cartridges after they have been used for hemoglobin quantification. In Chapter 4 of this dissertation, we tested whether the residual buffer from FIT cartridges could be used for microbiota characterization. We found that, indeed, bacterial DNA isolated from FIT cartridges

recapitulated the composition of the fecal microbiota and could be used to identify patients with CRC with the same accuracy as DNA isolated directly from stool.

Regardless of whether the shifts in the microbiota are the cause or effect of tumorigenesis, such shifts show potential as biomarkers for CRC. As we continue to better understand the role of the microbiota in CRC, we may be able to develop improved methods for screening, accelerating the downward trend in CRC incidence and mortality. Chapters 2, 3, and 4 of this dissertation advance the field through three important findings: 1.) An individual's baseline microbiota may influence their susceptibility to tumorigenesis. 2.) Changes in an individual's microbiota can be used in conjunction with FIT to more accurately detect lesions. 3.) DNA from FIT cartridges are an alternative source of material for microbiota-based screening. Chapter 5 will discuss the implications of these findings and the future of microbiota-based screening for CRC.

References

1. Whitman WB, Coleman DC, Wiebe WJ: **Prokaryotes: The unseen majority**. *Proceedings of the National Academy of Sciences*. 1998, **95**:6578–6583.

2. Bentley R, Meganathan R: **Biosynthesis of vitamin k (menaquinone) in bacteria.** *Microbiological reviews*. 1982, **46**:241.

3. Flint HJ, Scott KP, Duncan SH, Louis P, Forano E: **Microbial degradation of complex carbohydrates in the gut**. *Gut microbes*. 2012, **3**:289–306.

4. Gaboriau-Routhiau V, Rakotobe S, Lécuyer E, Mulder I, Lan A, Bridonneau C, Rochet V, Pisi A, De Paepe M, Brandi G, others: **The key role of segmented filamentous bacteria in the coordinated maturation of gut helper t cell responses**. *Immunity*. 2009, **31**:677–689.

5. Buffie CG, Pamer EG: **Microbiota-mediated colonization resistance against intestinal pathogens**. *Nature Reviews Immunology*. 2013, **13**:790–801.

6. Kostic AD, Xavier RJ, Gevers D: **The microbiome in inflammatory bowel disease: Current status and the future ahead**. *Gastroenterology*. 2014, **146**:1489–1499.

7. Koren O, Goodrich JK, Cullender TC, Spor A, Laitinen K, Bäckhed HK, Gonzalez A, Werner JJ, Angenent LT, Knight R, others: **Host remodeling of the gut microbiome and metabolic changes during pregnancy**. *Cell*. 2012, **150**:470–480.

8. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V, Henrissat B, Bain JR, others: **Gut microbiota from twins discordant for obesity modulate metabolism in mice**. *Science*. 2013, **341**:1241214.

9. Zackular JP, Baxter NT, Iverson KD, Sadler WD, Petrosino JF, Chen GY, Schloss PD: **The gut microbiome modulates colon tumorigenesis**. *MBio*. 2013, **4**:e00692–13.

10. Cheesman SE, Neal JT, Mittge E, Seredick BM, Guillemin K: **Epithelial cell proliferation in the developing zebrafish intestine is regulated by the wnt pathway and microbial signaling via myd88.** *Proc Natl Acad Sci U S A*. 2011:4570–4577.

11. Dolara P, Caderni G, Salvadori M, Morozzi G, Fabiani R, Cresci A, Orpianesi C, Trallori G, Russo A, Palli D: **Fecal levels of short-chain fatty acids and bile acids as determinants of colonic mucosal cell proliferation in humans**. *Nutrition and cancer*. 2002, **42**:186–190.

12. Stappenbeck TS, Hooper LV, Gordon JI: **Developmental regulation of intestinal angiogenesis by indigenous microbes via paneth cells**. *Proceedings of the National Academy of Sciences*. 2002, **99**:15451–15455.

13. Turnbaugh PJ, Bäckhed F, Fulton L, Gordon JI: **Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome**. *Cell host & microbe*. 2008, **3**:213–223.

14. Huxley RR, Ansary-Moghaddam A, Clifton P, Czernichow S, Parr CL, Woodward M: **The impact of dietary and lifestyle risk factors on risk of colorectal cancer: A quantitative overview of the epidemiological evidence**. *International journal of cancer*. 2009, **125**:171–180.

15. Jess T, Rungoe C, Peyrin–Biroulet L: **Risk of colorectal cancer in patients with ulcerative colitis: A meta-analysis of population-based cohort studies**. *Clinical Gastroenterology and Hepatology*. 2012, **10**:639–645.

16. Keku TO, Dulal S, Deveaux A, Jovov B, Han X: **The gastrointestinal microbiota and colorectal cancer**. *American Journal of Physiology-Gastrointestinal and Liver Physiology*. 2015, **308**:G351–G363.

17. Wu S, Rhee K-J, Albesiano E, Rabizadeh S, Wu X, Yen H-R, Huso DL, Brancati FL, Wick E, McAllister F, others: **A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses**. *Nature medicine*. 2009, **15**:1016–1022.

18. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan T-J, Campbell BJ, Abujamel T, Dogan B, Rogers AB, others: **Intestinal inflammation targets cancer-inducing activity of the microbiota**. *science*. 2012, **338**:120–123.

19. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, Clancy TE, Chung DC, Lochhead P, Hold GL, others: **Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment**. *Cell host & microbe*. 2013, **14**:207–215.

20. Tanaka T, Kohno H, Suzuki R, Yamada Y, Sugie S, Mori H: **A novel inflammation-related mouse colon carcinogenesis model induced by azoxymethane and dextran sodium sulfate**. *Cancer science*. 2003, **94**:965–973.

21. De Robertis M, Massi E, Poeta ML, Carotti S, Morini S, Cecchetelli L, Signori E, Fazio VM, others: **The aOM/DSS murine model for the study of colon**

carcinogenesis: From pathways to diagnosis and therapy studies. *Journal of carcinogenesis*. 2011, **10**:9.

22. Zackular JP, Baxter NT, Chen GY, Schloss PD: **Manipulation of the gut microbiota reveals role in colon tumorigenesis**. *mSphere*. 2015, **1**:e00001–e00015.

23. Morgun A, Dzutsev A, Dong X, Greer RL, Sexton DJ, Ravel J, Schuster M, Hsiao W, Matzinger P, Shulzhenko N: **Uncovering effects of antibiotics on the host and microbiota using transkingdom gene networks**. *Gut*. 2015:gutjnl–2014.

24. Baxter NT, Zackular JP, Chen GY, Schloss PD: **Structure of the gut microbiome following colonization with human feces determines colonic tumor burden**. *Microbiome*. 2014, **2**:1.

25. Siegel R, DeSantis C, Jemal A: **Colorectal cancer statistics, 2014**. *CA: a cancer journal for clinicians*. 2014, **64**:104–117.

26. Group UCSW, others: **United states cancer statistics: 1999–2010 incidence and mortality web-based report**. *Atlanta: US Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute*. 2013.

27. Centers for Disease Control and Prevention: **Vital signs: colorectal cancer screening test use–United States, 2012.** *MMWR Morbidity and mortality weekly report*. 2013, **62**:881.

28. Hsia J, Kemper E, Kiefe C, Zapka J, Sofaer S, Pettinger M, Bowen D, Limacher M, Lillington L, Mason E, others: **The importance of health insurance as a determinant**

of cancer screening: evidence from the Women's Health Initiative. *Preventive medicine*. 2000, **31**:261–270.

29. Jones RM, Devers KJ, Kuzel AJ, Woolf SH: **Patient-reported barriers to colorectal cancer screening: a mixed-methods analysis**. *American journal of preventive medicine*. 2010, **38**:508–516.

30. Rex DK, Johnson DA, Anderson JC, Schoenfeld PS, Burke CA, Inadomi JM: **American College of Gastroenterology guidelines for colorectal cancer screening 2008**. *The American journal of gastroenterology*. 2009, **104**:739–750.

31. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, Ahlquist DA, Berger BM: **Multitarget stool DNA testing for colorectal-cancer screening**. *New England Journal of Medicine*. 2014, **370**:1287–1297.

32. Soler AP, Miller RD, Laughlin KV, Carp NZ, Klurfeld DM, Mullin JM: **Increased tight junctional permeability is associated with the development of colon cancer**. *Carcinogenesis*. 1999, **20**:1425–1432.

33. Ulger Toprak N, Yagci A, Gulluoglu B, Akin M, Demirkalem P, Celenk T, Soyletir G: **A possible role of bacteroides fragilis enterotoxin in the aetiology of colorectal cancer**. *Clinical microbiology and infection*. 2006, **12**:782–786.

34. Boleij A, Hechenbleikner EM, Goodwin AC, Badani R, Stein EM, Lazarev MG, Ellis B, Carroll KC, Albesiano E, Wick EC, others: **The bacteroides fragilis toxin gene is prevalent in the colon mucosa of colorectal cancer patients**. *Clinical Infectious Diseases*. 2014:ciu787.

35. Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, Ojesina AI, Jung J, Bass AJ, Tabernero J, others: **Genomic analysis identifies association of Fusobacterium with colorectal carcinoma**. *Genome research*. 2012, **22**:292–298.

36. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW: **Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating e-cadherin/β-catenin signaling via its fadA adhesin**. *Cell host & microbe*. 2013, **14**:195–206.

37. Yu J, Feng Q, Wong SH, Zhang D, Liang Q yi, Qin Y, Tang L, Zhao H, Stenvang J, Li Y, others: **Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer**. *Gut*. 2015:gutjnl–2015.

38. Warren RL, Freeman DJ, Pleasance S, Watson P, Moore RA, Cochrane K, Allen-Vercoe E, Holt RA: **Co-occurrence of anaerobic bacteria in colorectal carcinomas**. *Microbiome*. 2013, **1**:16.

39. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Böhm J, Brunetti F, Habermann N, others: **Potential of fecal microbiota for early-stage detection of colorectal cancer**. *Molecular systems biology*. 2014, **10**:766.

40. Zackular JP, Rogers MA, Ruffin MT, Schloss PD: **The human gut microbiome as a screening tool for colorectal cancer**. *Cancer Prevention Research*. 2014, **7**:1112–1121.

41. Baxter NT, Ruffin MT, Rogers MA, Schloss PD: **Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions**. *Genome Medicine*. 2016, **8**:1.

42. Sinha R, Chen J, Amir A, Vogtmann E, Inman KS, Flores-Munguia R, Sampson JN, Knight R, Chia N, others: **Collecting fecal samples for microbiome analyses in epidemiology studies**. *Cancer Epidemiology Biomarkers & Prevention*. 2015:cebp–0951.

CHAPTER 2:

Structure Of The Gut Microbiome Following Colonization With Human Feces Determines Colonic Tumor Burden

The contents of this chapter have been published as:

Baxter NT*, Zackular JP*, Chen GY, Schloss PD. (2014). Structure of the gut microbiome following colonization with human feces determines colonic tumor burden. *Microbiome* 2(1):20. *Equal Contributon

Abstract

A growing body of evidence indicates that the gut microbiome plays a role in the development of colorectal cancer (CRC). CRC patients harbor gut microbiomes that are structurally distinct from those of healthy individuals; however without the ability to track individuals during disease progression, it has not been possible to observe changes in the microbiome over the course of tumorigenesis. Mouse models have demonstrated that these changes can further promote colonic tumorigenesis. However, these models have relied upon mouse-adapted bacterial populations and so it remains unclear which human-adapted bacterial populations are responsible for modulating tumorigenesis. We transplanted fecal microbiota from three CRC patients and three healthy individuals into germ-free mice, resulting in six structurally distinct microbial communities. Subjecting

these mice to a chemically induced model of CRC resulted in different levels of tumorigenesis between mice. Differences in the number of tumors were strongly associated with the baseline microbiome structure in mice, but not with the cancer-status of the human donors. Partitioning of baseline communities into enterotypes by Dirichlet multinomial mixture modeling resulted in 3 enterotypes that corresponded with tumor burden. The taxa most strongly positively correlated with increased tumor burden were members of the *Bacteroides*, *Parabacteroides*, *Alistipes*, and *Akkermansia*, all of which are Gram-negative. Members of the Gram-positive Clostridiales, including multiple members of *Clostridium* Group XIVa were strongly negatively correlated with tumors. Analysis of the inferred metagenome of each community revealed a negative correlation between tumor counts and the potential for butyrate production, and a positive correlation between tumor counts and the capacity for host glycan degradation. Despite harboring distinct gut communities, all mice underwent conserved structural changes over the course of the model. The extent of these changes was also correlated with tumor incidence. Our results suggest that the initial structure of the microbiome determines the susceptibility to colonic tumorigenesis. There appear to be opposing roles for certain Gram-negative (Bacteroidales and Verrucomicrobia) and Gram-positive (Clostridiales) bacteria in tumor susceptibility. Thus, the impact of community structure is potentially mediated by the balance between protective, butyrate producing populations and inflammatory, mucin degrading populations.

Background

Colorectal cancer (CRC) is the second leading cause of cancer-related deaths in the United States each year [1]. Recent evidence suggests that the community of microbes inhabiting the gastrointestinal tract plays an important role in the development and progression of CRC [2-4]. This community, termed the gut microbiome, is known to influence cancer-related functions including cell proliferation, angiogenesis, and apoptosis, and it is strongly linked to diet, obesity and inflammation, which are known risk factors of CRC [5-9]. Using a mouse model of CRC, we have shown that structural changes to the microbiome occur during tumorigenesis and result in a gut microbiome with an increased tumorigenic capacity [10]. These findings demonstrate that the gut microbiome has a causal role in the development and progression of CRC.

Several survey-based studies have shown that CRC patients harbor microbial communities that are structurally distinct from those of healthy individuals [11-15]. However, there has been no consensus among these studies as to which bacterial populations are important. In mouse models several gut commensals have been shown to promote tumorigenesis in the colon. Both enterotoxigenic *Bacteroides fragilis* (ETBF) and strains of *Escherichia coli* that carry the *pks* pathogenicity island can promote tumorigenesis by the production of toxins [3, 4]. *Fusobacterium nucleatum* has also been shown to potentiate tumorigenesis in mouse models and cell culture experiments by stimulating inflammation via myeloid cell recruitment and/or activation of β-catenin signaling [2, 16]. *Fusobacterium* was also found to be enriched in a subset of human colon adenomas [15]. Although there is increasing evidence that *Fusobacterium* is

involved in CRC cases, it was detected in less than half of adenomas, which suggests that other bacterial populations are capable of potentiating tumorigenesis [2]. In fact, it may be that CRC is a polymicrobial disease requiring combinations of these or other populations to influence tumorigenesis.

While individual bacterial species have been associated with some human CRC cases, in other cases the capacity of the microbiome to modulate tumorigenesis could be determined by the structure of the community as a whole rather than the presence or absence of individual populations [4, 17]. The potentially polymicrobial influence of the gut microbiome on this disease necessitates the disentangling of the complex interactions between bacterial populations in the gut. Understanding these interactions requires investigation of the relationship between the microbiome and tumorigenesis under a diverse set of community structures. Unfortunately, mechanistic studies typically rely on experiments with conventionally reared inbred mouse living in homogenous, controlled environments, leading to relatively little variation in microbiome structure between individual animals. Although, experiments in conventional mice are useful for understanding the mechanisms by which the microbiome modulates tumorigenesis, they are limited by investigating only those bacterial strains found in laboratory mice, many of which are absent in humans. It is reasonable to expect that incorporating human-associated microbial populations into these experiments would increase the ability to translate results to humans.

To investigate the role of microbiome structure in tumorigenesis, we combined the advantages of the high interpersonal variation among humans and the convenience of a mouse model. We inoculated germ-free mice with microbiota from human subjects harboring distinct microbiomes. This technique enabled us to test the effect of different baseline microbiome communities with variation beyond what is seen in conventionally reared mice. The transfer of human microbiota to germ-free mice, sometimes referred to as “humanization”, has been employed to study the microbiome in the context of several other diseases. In studies of diabetes, obesity, and malnutrition, colonization with human feces has been reported to recapitulate the phenotype of the human donors in the recipient mice [18-21]. Thus, in addition to searching for tumor-modulating community structures, we sought to determine whether this strategy could be used to recapitulate the tumor-promoting capacity of CRC patients’ microbiota in mice.

Methods

Mouse experiments

Fecal samples from three healthy individuals and three patients found to harbor carcinomas were obtained through the Early Detection Research Network (Table 2.1). Diagnoses were determined based on colonoscopy and histology. All six samples were PCR-negative for the ETBF toxin and the *E. coli pks* island [4, 22]. Inocula were prepared by mixing 200 mg of each sample in 5ml of PBS. Age-matched (6-10 weeks), male, germ-free C57BL/6 mice were inoculated by oral gavage with 100 µl of inoculum (n=10 for groups H1 and C1, n=5 for others). Mice were housed 5 mice per cage. Three

weeks after inoculation mice received a single intra-peritoneal injection of azoxymethane (AOM) (10 mg/kg of body weight). Five days later mice were subjected to the first of three five-day rounds of 2% dextran sulfate sodium (DSS) administered *ad libitum* in the drinking water (Figure 2.1). Sixteen days of recovery separated each round of DSS. Three weeks after the third and final round of DSS mice were euthanized and colonic tumors were enumerated. With this model mice consistently develop noninvasive adenomas with dysplastic changes [23, 24]. Throughout the experiment the mice were housed in germ-free isolators at the University of Michigan Germ-free Facility. This animal experiment was approved by the University Committee on Use and Care of Animals at the University of Michigan.

DNA extraction and 16S rRNA gene sequencing

Mouse fecal samples were collected throughout the experiment and frozen at -20˚C. Genomic DNA from samples collected on days 0 and 73 and the human inocular were isolated using the PowerSoil-htp 96 Well Soil DNA isolation kit (Mo Bio) using an EpMotion 5075 automated pipetting system. The V4 region of the 16S rRNA gene was amplified using custom barcoded primers and sequenced as described previously using an Illumina MiSeq sequencer [25]. All fastq files and the MIMARKS spreadsheet are available at http://www.mothur.org/human_mouse_aomdss.

Sequence curation and analysis

16S rRNA gene sequences were curated using the mothur software package as described previously [25, 26]. Briefly, paired end reads were assembled into contigs and

aligned to the SILVA 16S rRNA sequence database [27]. Sequences that failed to align or were flagged as possible chimeras by UCHIME were removed [28]. Each sequence was classified using a Naïve Bayesian classifier trained against a 16S rRNA gene training set provided by the Ribosomal Database Project (http: //sourceforge.net/projects/rdp-classifier/) [29]. Finally, sequences were grouped based on their taxonomic classification or clustered into operational taxonomic units (OTUs) based on a 97% similarity cutoff. The number of sequences in each sample was rarefied to 3,306 sequences per sample to minimize the effects of uneven sampling. Parallel sequencing and processing of a mock community indicated that the error rate of the curated sequences was 0.085%.

The dissimilarity in community structure between samples was calculated using the Θ_{YC} metric [30]. The Θ_{YC} distances between samples were used for ordination analysis by non-metric dimensional scaling (NMDS) in two dimensions. Ten iterations were performed and the resulting ordination that had the lowest stress was used for data visualizations. Dirichlet multinomial mixture models were generated to group samples into enterotypes based on the abundance of bacterial genera in each sample [31]. To identify conserved changes that occurred over the course of the AOM/DSS model, the samples from each mouse on day 0, and the samples collected at the end of the model were grouped into “baseline” and “endpoint” categories respectively. The R randomForest package was used to identify the OTUs that best distinguished between the two categories based on their importance for the classification model [32, 33].

The Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) software package was used to infer the metagenomic content of each sample, based on the taxonomy and abundance of each OTU [34]. Although this method is limited by the number of available genomes, it has been shown to replicate metagenomes to a high degree of accuracy, especially for human-adapted bacterial communities. The weighted Nearest Sequenced Taxon Index (NSTI) for our samples was 0.056 +/- 0.01. In general, NSTI values below 0.06, suggest that closely related reference genomes were available the dataset (http://picrust.github.io/picrust/tutorials/quality_control.html). From the inferred metagenomes, we identified KEGG orthologs that could be used as markers for butyrate production or host-glycan degradation. Because either butyrate kinase or butyryl CoA:acetate CoA transferase is required for butyrate production in the gut, the KEGG orthologs chosen as markers for butyrate production were K00929 (butyrate kinase [EC:2.7.2.7]), K01034 (acetate CoA-transferase alpha subunit [EC:2.8.3.8]), K01035 (acetate CoA-transferase beta subunit [EC:2.8.3.8]) [35]. To choose markers for glycan degradation, we found all of the KEGG orthologs annotated as sialidases, fucosidases, sulfatases, or members of the glycoside hydrolase family 18, as these classes of enzymes are necessary, and moderately specific for host glycan degradation [36, 37]. Ten such KEGG orthologs were found in the metagenomes, and used as markers; K01138 (uncharacterized sulfatase [EC:3.1.6.-]), K01130 (arylsufatase [EC:3.1.6.1]), K01135 (arylsufatase B [EC:3.1.6.12]), K01137 (N-acetylglucosamine-6-sulfatases [EC:3.1.6.14]), K01134 (arylsufatase A [EC:3.1.6.8]), K01186 (sialidase-1 [EC:3.2.1.18]), K01206 (alpha-L-fucosidase [EC:3.2.151]), K01183 (1,4-beta-poly-N-

acetylglucosaminidase [EC:3.2.1.14]), K01205 (alpha-N-acetylglucosaminidase [EC:3.2.1.50]), and K05970 (sialate O-acetylesterase [EC:3.1.1.53]). Finally, we calculated the Spearman correlation coefficients between tumor counts and these KEGG orthologs.

Statistical Analyses

Differences in tumor counts between DMM partitions were examined using a Wilcoxon rank-sum test. To test whether there was a significant difference in tumor counts between groups that received healthy or cancer-associated inocula, we rank transformed the tumor counts to correct for heteroscedasticity and performed a nested ANOVA. Differences in community structure were examined using Analysis of Molecular Variance (AMOVA) in mothur.

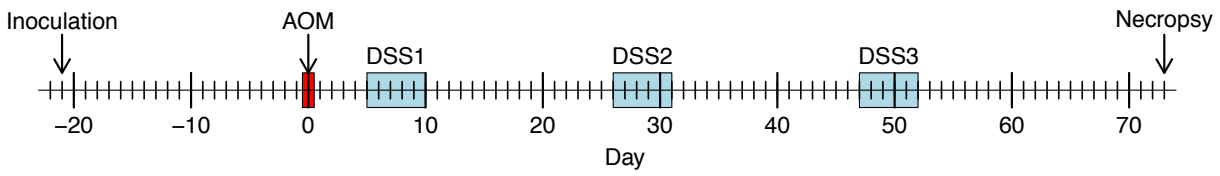

Figure 2.1. **Experimental Design.** Germ-free mice were inoculated by oral gavage with one of six human inocula. Twenty-one days later (day 0) they received a single intra-peritoneal injection of AOM (10 mg/kg). Mice were subsequently administered three five-day rounds of 2% DSS in the drinking water, with sixteen days of rest in between. Mice were euthanized seventy-three days after the AOM injection for enumeration of colonic tumors. The inocula and samples collected on day 0 and day 73 were used for 16S rRNA sequencing.

Patient	Age	Gender	Race	Height (cm)	weight (kg)	FOBT result	Medications	Diagnosis	Stage	Location	Min tumor size(cm)	Max tumor size (cm)
C1	74	f	white	157	79	-	Exelon, Furosemide,Trazodone	Cancer	IIIc	rectal	9.5	9.5
C2	63	m	asian	167	86	+	nexium, metformin	Cancer	I	rectal	1.2	2.1
C3	60	m	white	182	109	+	Claritan	Cancer	IIIb	rectal	2	3.1
H1	52	f	white	162	84	-	requip	Normal	n/a	n/a	n/a	n/a
H2	69	f	white	157	46	-	none	Normal	n/a	n/a	n/a	n/a
H3	55	f	asian	157	73	-	cyclobenzaprine HCL	Normal	n/a	n/a	n/a	n/a

Table 2.1. Metadata for the six inoculum donors.

Results

Colonization of germ-free mice with human microbiota

We colonized germ-free mice with human feces from six individuals to determine whether different initial community structures would yield different numbers of tumors after going through the AOM/DSS model. Three of the donors had healthy colons (H1, H2, H3) and three had colonic carcinomas (C1, C2, C3). Samples were chosen because they represented broad variation in community structure (Figure 2.2A). Following gavage and a 21-day colonization period, groups showed varying levels of similarity to their inocula based on phylum level relative abundances and the Θ_{YC} distances calculated from OTU abundances (Figure 2.2B). Low Θ_{YC} distances between mice within groups suggested that individual communities were consistent within each group, while large Θ_{YC} distances between groups suggest that each group harbored a gut microbiome that was structurally distinct from the others. Pairwise AMOVA between groups revealed that colonization with different inocula resulted in significantly different community structures ($p<0.01$, Benjamini-Hochberg correction). These results suggest that although mice do not closely resemble their inoculum, all sets of mice developed stable, structurally distinct communities.

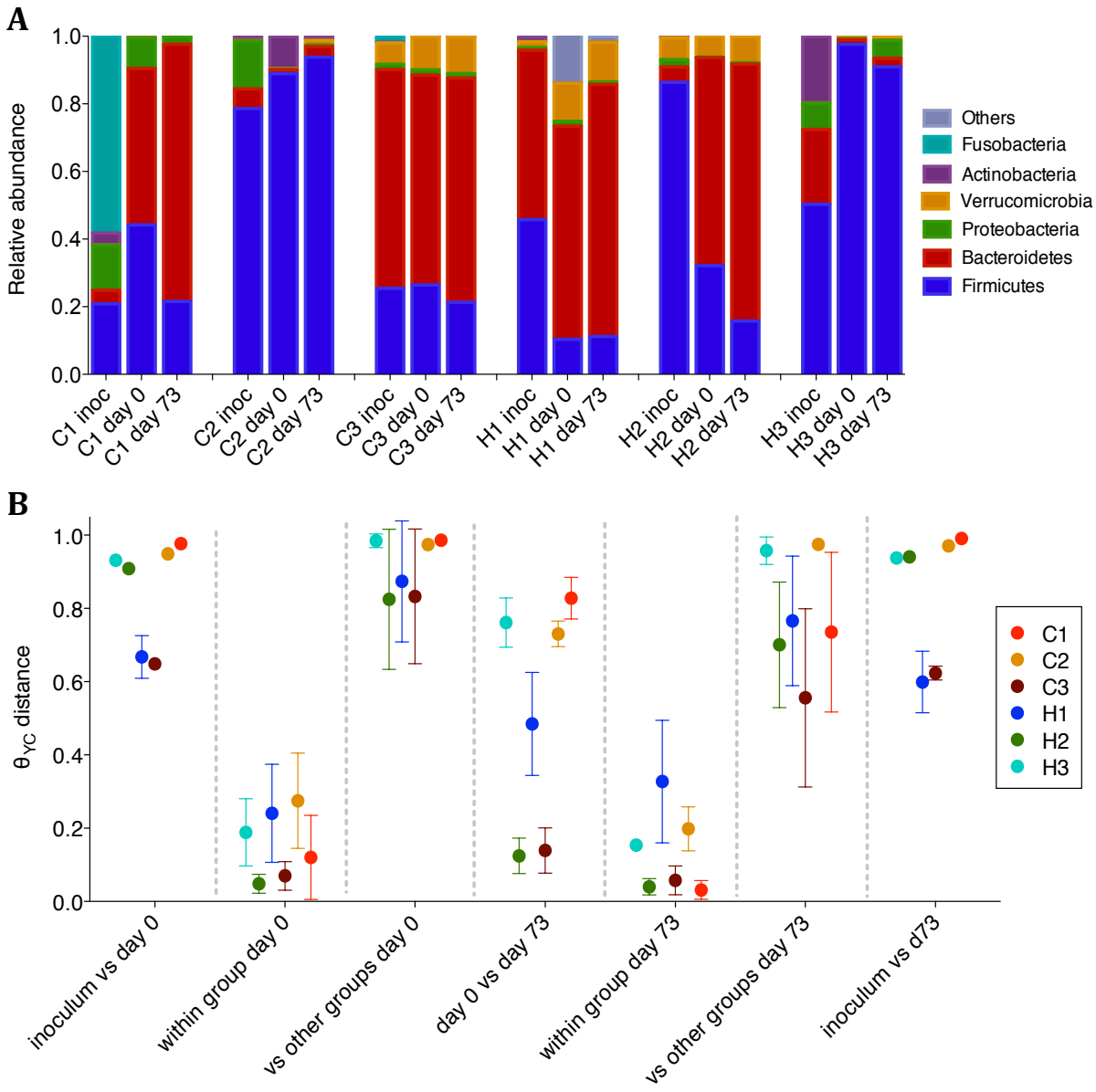

Figure 2.2. Taxonomic composition and beta diversity across treatment groups and time. (A) Phylum level relative abundance of the fecal microbiome of each group and in its inoculum. **(B)** Average Θ_{YC} distances(+/- SEM) within and between groups at various time points; between each group and its inoculum, within each group at day 0, each group compared to others at day 0, between day 0 and day 73 for each group, and each group compared to others at day 73.

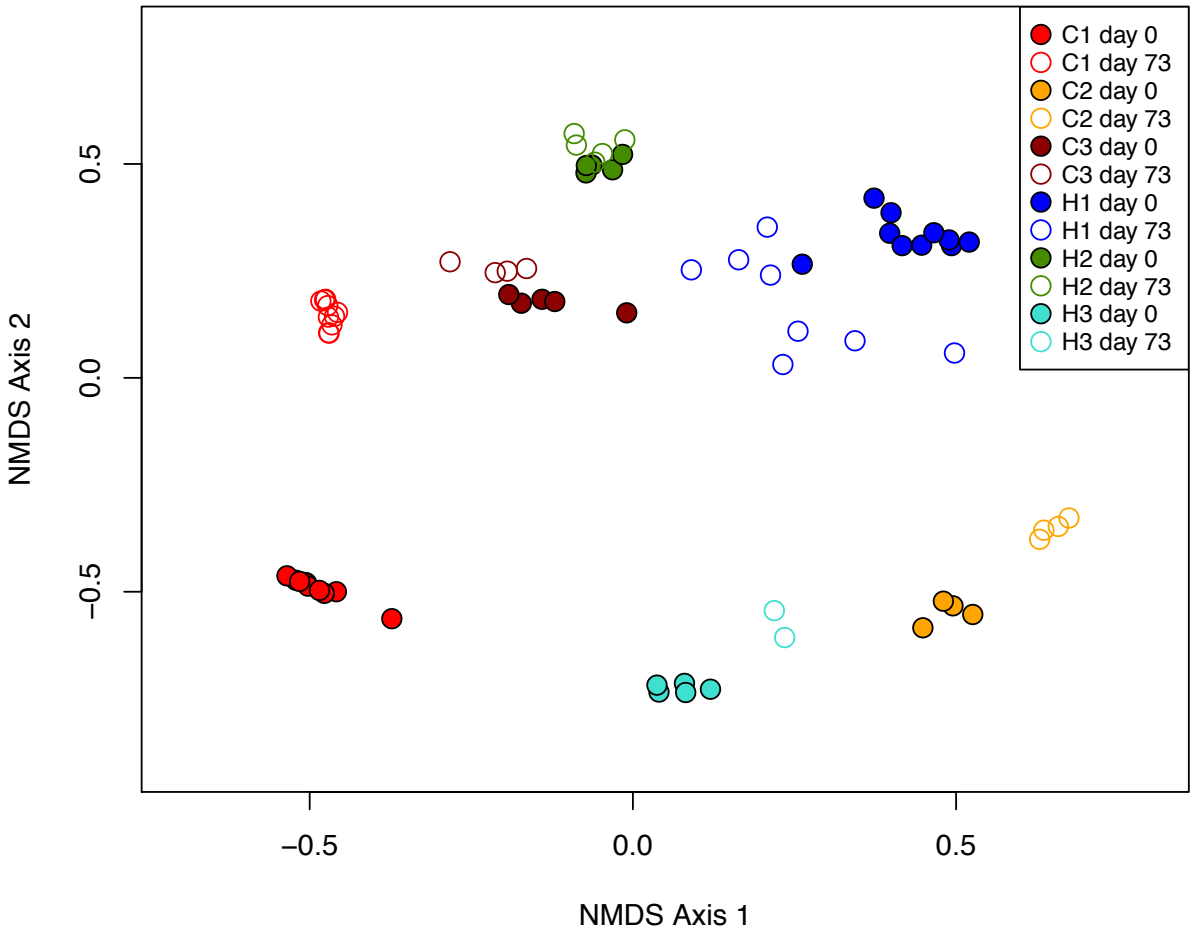

Figure 2.3. Temporal changes in community structure. NMDS ordination based OTU abundances between samples on day 0 and day 73. Distances were calculated with Θ_{YC}.

Tumor incidence is linked to initial community structure

Once colonized, mice were subjected to the AOM/DSS model of CRC. We observed significant variation in the number of tumors between mice (Figure 2.4A). These differences were associated with the inoculum they received, but not the cancer status of the human donor (Nested ANOVA $p<0.0005$). Thus the phenotype of the human subject was not transferred to their mouse counterparts. Ordination of the communities revealed an association between the community structure of each group at the beginning of the AOM/DSS model and their median tumor counts (Figure 2.4B). To test for cage effects groups H1 and C1 were each inoculated into duplicate cages of 5 mice each (n=10 per inoculum). There was no significant difference in microbiome structure ($p>0.05$, AMOVA) or tumor counts ($p>0.05$, Wilcoxon test) between cages within each group.

To determine which OTUs were driving this trend, we generated a biplot using the NMDS axes generated from the Θ_{YC} distances between samples collected at the time of AOM injection (day 0) (Figure 2.4B). Among the OTUs most strongly correlated with high tumor counts were two OTUs from the genus *Bacteroides* (OTUs 1 and OTU 4). More detailed characterization of these OTUs indicated that OTU 1 was closely affiliated with *B. uniformis* and OTU 4 was affiliated with a mixture of a mixture of *Bacteroides* spp. including *B. fragilis*, *B. ovatus*, *B. xylanisolvens* and *B. thetaiotaomicron*. Both of these OTUs were found in all six cohorts of mice and their initial abundances were positively correlated with tumor counts ($\rho=0.47$ and 0.49, respectively; both $p<0.005$; Spearman correlation). Interestingly, all samples were PCR-negative for the ETBF toxin

gene, suggesting that OTU 4 was not ETBF. Other OTUs associated with high tumor counts were affiliated with the genera *Parabacteroides* (OTU 18), *Alistipes* (OTU 19), and *Bacteroides* as well as an OTU affiliated with the species *Akkermansia muciniphila* (OTU 11). In addition, several OTUs associated with the *Clostridium* Group XIVa (OTUs 7, 9, 15, and 17), *Clostridium* Group IV (OTU 49), and unclassified members of the Lachnospiraceae (OTUs 67 and 13) were correlated with lower tumor counts. These results indicate that the relative abundance of specific OTUs in the starting community could be associated with tumor counts.

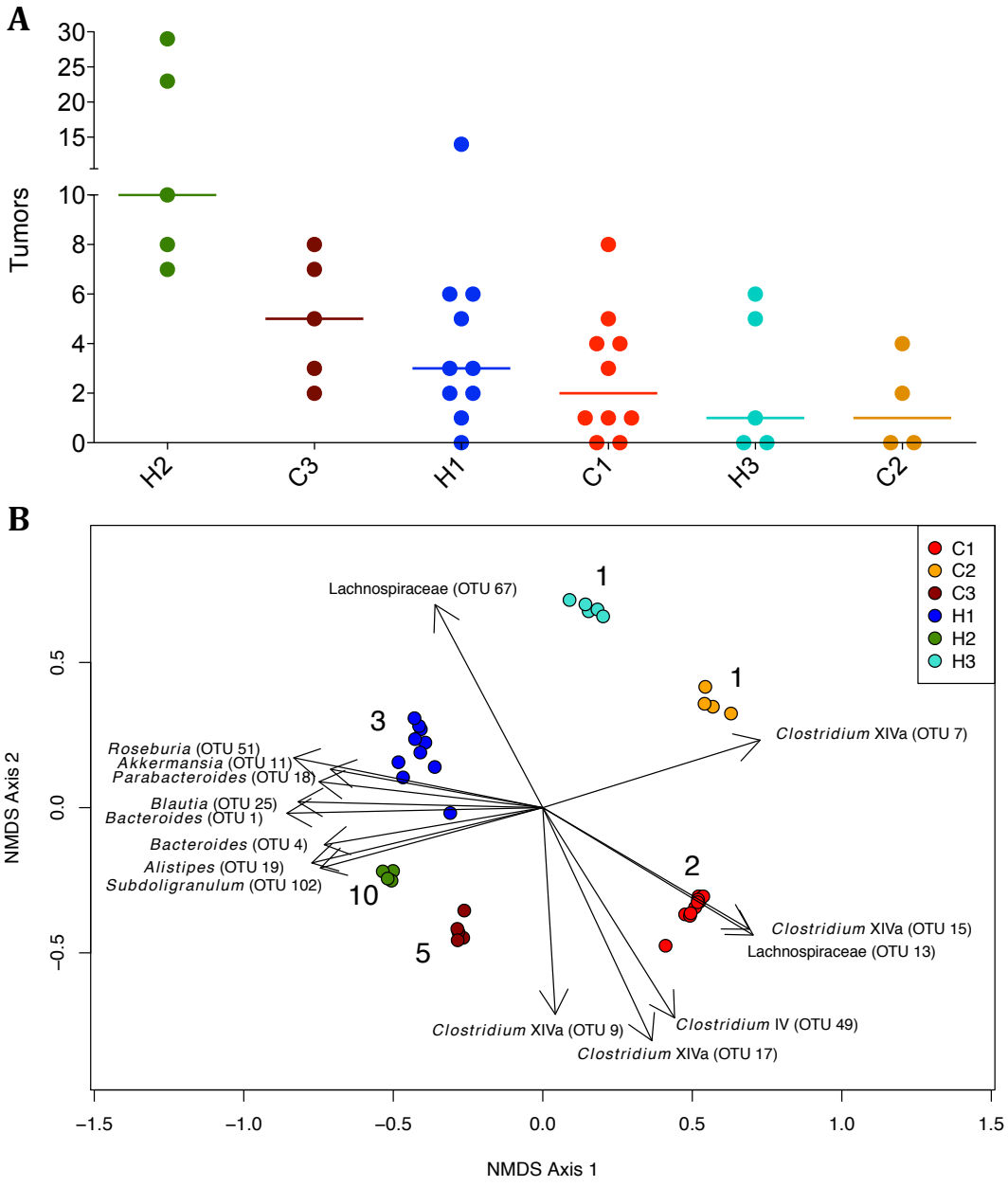

Figure 2.4. Correlation of tumor incidence with initial gut community structure. **(A)** Stripchart of tumor counts (with line at median) for each group. **(B)** NMDS plot based on Θ_{YC} distances between samples at day 0 with biplot of the 15 OTUs most strongly correlated with the NMDS axes (stress = 0.21). Median tumor counts for each group are adjacent to their corresponding dots.

To further test the association between the starting community structure and tumor incidence we clustered samples into community types using Dirichlet multinomial mixture (DMM) models based on the abundance of bacterial genera found in the mice. This approach allowed us to quantify the association between the starting community structure and tumor burden in an unbiased manner. The DMM model with the highest likelihood partitioned the samples into three enterotypes (Figure 2.5A). Enterotype 1 was composed exclusively of samples from the three treatment groups with the highest tumor counts (H2, C3, H1). Enterotype 2 was composed of samples from C1, which had the third lowest tumor counts. Enterotype 3 was composed entirely of samples from the two groups with the lowest tumor counts (C2, H3). As a result, mice in Enterotype 1 had significantly more tumors than the other two partitions ($p<0.05$, Wilcoxon test; Figure 2.5B). Consistent with the OTU analysis, the DMM partition with the highest tumor counts was enriched for the genus *Bacteroides* (Figure 2.5C). In addition, other genera within the order Bacteroidales (*Parabacteroides* and *Alistipes)*, as well as *Akkermansia* were enriched in Enterotype 1. An unclassified member of the Porphrymonodaceae, was enriched in Enterotype 2, which had significantly fewer tumors than Enterotype 1. Enterotype 3, which had the fewest tumors, was enriched for several genera within the order Clostridiales (*Clostridim* Group XIVa*, Clostridium* Group XI*, Clostridium* Group XVIII*, Flavonifractor*, and unclassified Lachnospiraceae). These data suggest a potentially tumorigenic role for certain members of Bacteroidales and a protective role for certain members of Clostridiales.

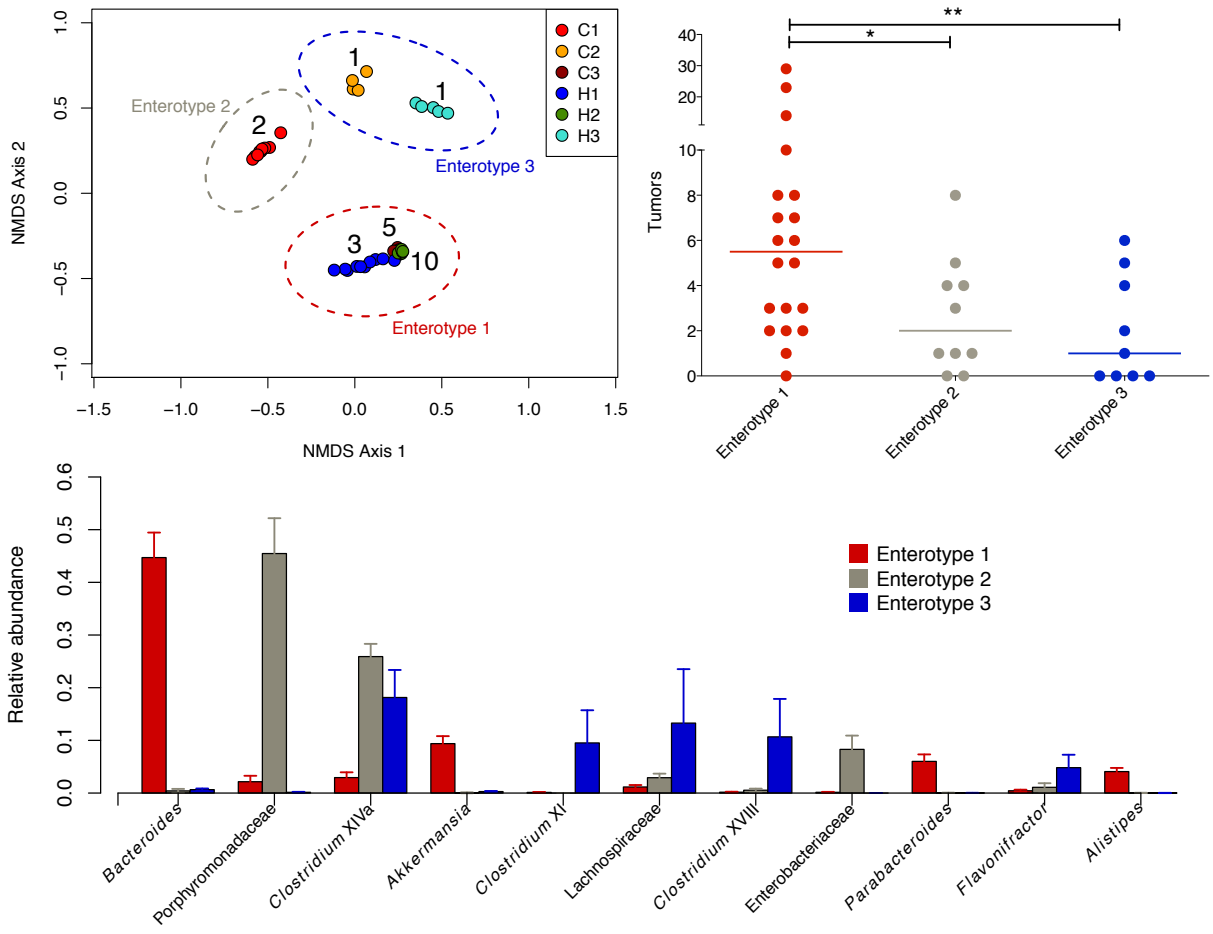

Figure 2.5. Correlation of enterotypes with tumor incidence. (A) NMDS plot based on genus level abundances with median tumor counts for each group (stress = 0.13) . Samples are circled based on their DMM enterotype. **(B)** Tumor counts for the mice in each DMM enterotype (* p<0.05, **p<0.01, Wilcoxon rank-sum test) **(C)** Relative abundance of the genera with the largest differences between enterotypes.

Changes in the microbiome during the AOM/DSS model

To determine the extent to which the microbiomes of each group changed over the course of the AOM/DSS model we calculated the Θ_{YC} distances between the communities in mice at the time of AOM injection and at the end of the experiment. Interestingly, the two groups with the highest tumor counts (H2, C3) changed very little over time (Θ_{YC}=0.12 and 0.14), while the microbiomes of the three groups with the lowest tumor counts (C2, H3, C1) changed substantially (Θ_{YC} = 0.73, 0.76, 0.83) (Figure 2.2A). Thus the closer the initial community of each group was to the tumor-associated endpoint community, the more tumors those mice developed.

To identify which OTUs changed over time, we combined samples from all six treatment groups and used the Random Forest machine-learning algorithm to identify the OTUs that allowed us to differentiate between the samples from the beginning and end of the model, regardless of the inoculum. The resulting model was able to distinguish between the baseline and endpoint samples with 98.6% accuracy. The OTUs that provided the greatest mean decrease in accuracy when removed from the analysis were affiliated with *Turicibacter* (OTU 36), *Bacteroides* (OTU 4), Porphyromonadaceae (OTU 59), and several genera within the Clostridiales (OTUs 113, 25, 28, 127, 144, 42, and 17; Figure 2.6). Despite harboring drastically different community structures, all treatment groups underwent conserved changes in microbiome structure over the course of the model.

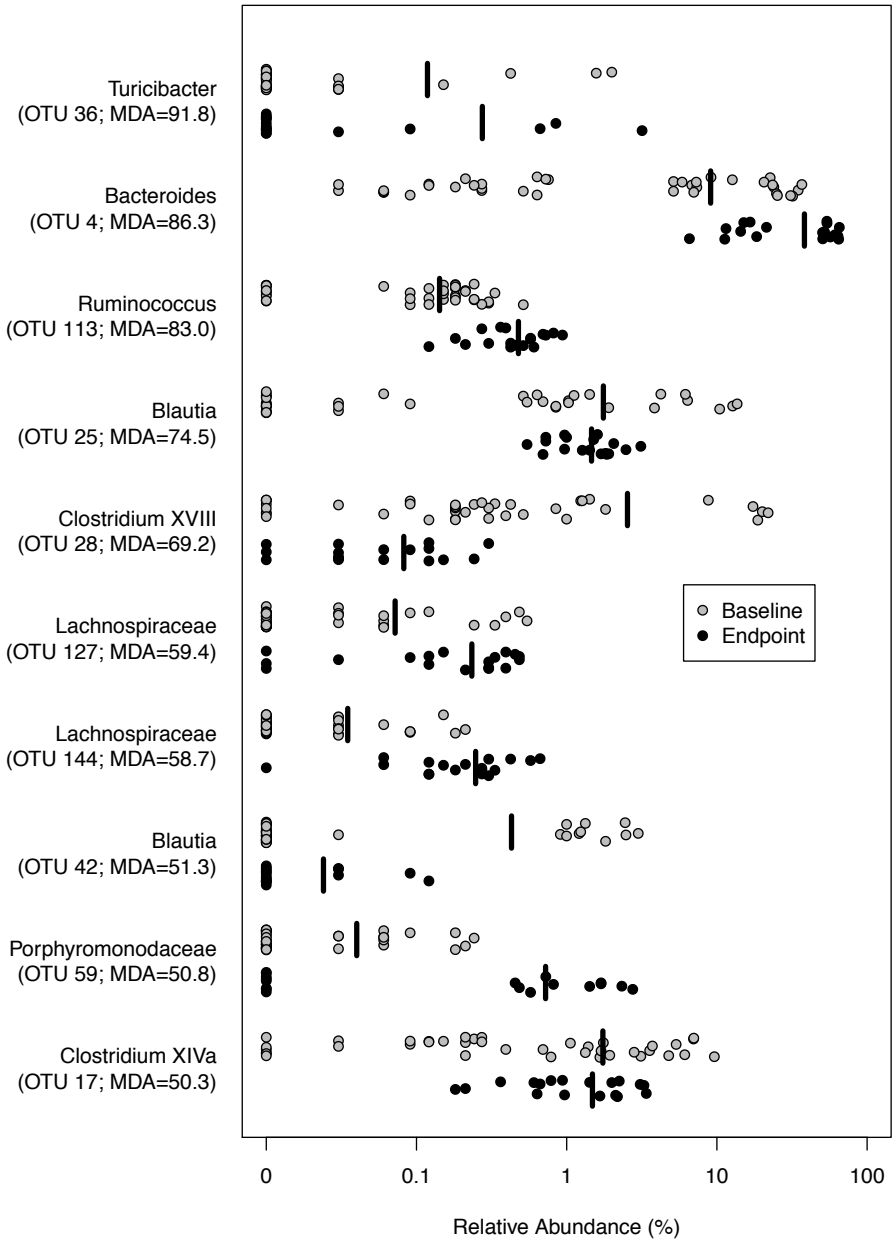

Figure 2.6. Temporal changes in the microbiome are conserved between groups. Strip chart showing the relative abundances of the 10 OTUs with the highest importance for distinguishing between baseline (day 0) and endpoint (day 73) communities by random forest as measured by the mean decrease accuracy (MDA) when the OTU was removed from the model. Each dot represents a single mouse. The black lines represent the mean relative abundance for all mice.

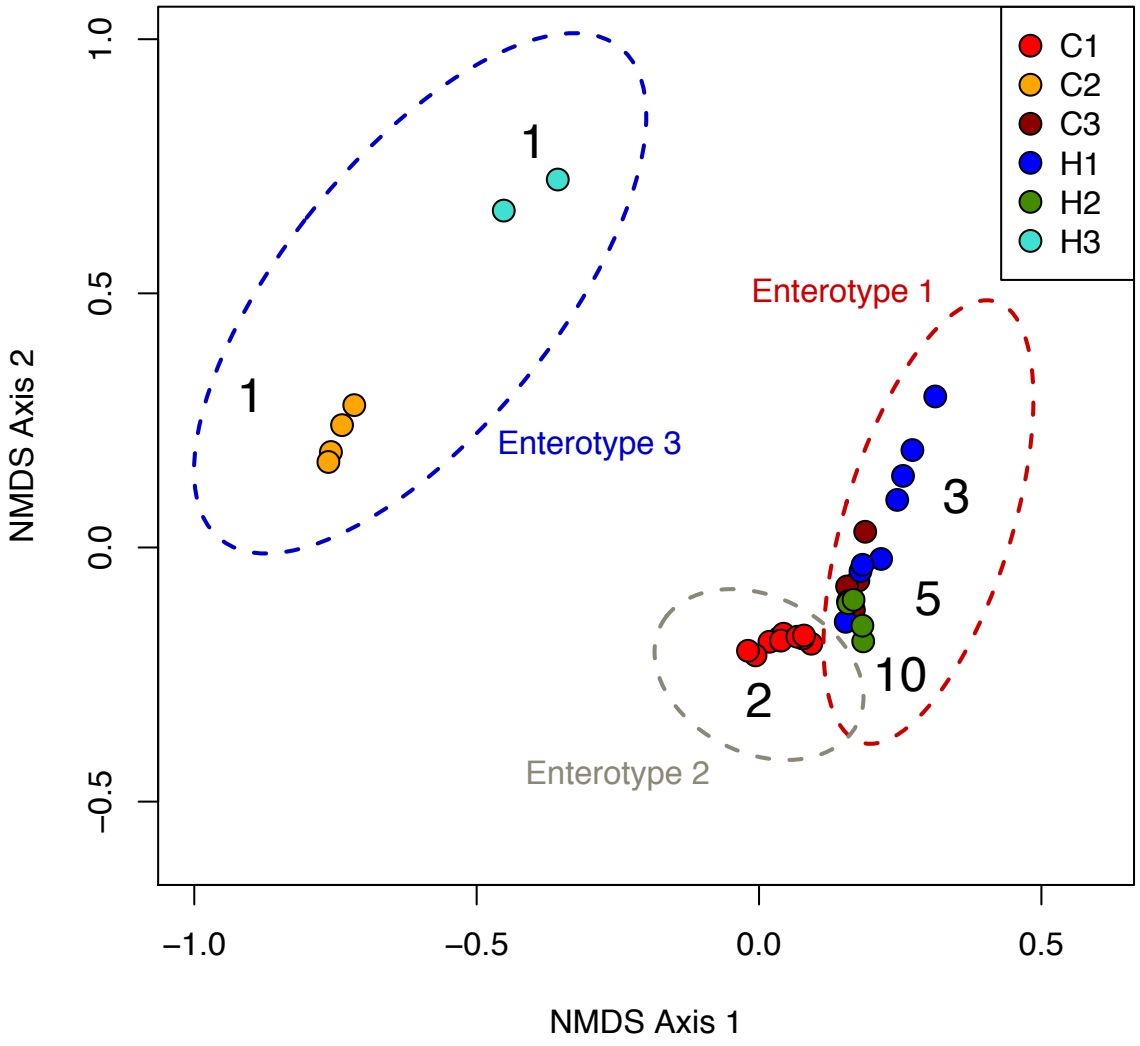

Figure 2.7. Samples remain in same enterotypes over the course of the model. NMDS ordination showing DMM enterotypes generated based on genus level abundances on day 73. Distances were calculated with Θ_{YC}. Despite changes in OTU abundance over the course of the model, all mice cluster into the same enterotypes on day 73 as they did on day 0.

Tumor incidence is linked to butyrate production and host glycan degradation.

Our experiments suggested that Clostridiales, Bacteroidales and Akkermansia played a role in modulating tumorigenesis. Members of the Clostridiales, especially *Clostridium* Group XIVa*,* are the predominate producers of intestinal butyrate, an important anti-inflammatory and anti-tumorigenic metabolite in the gut [35, 38, 39] *Bacteroides* and *Akkermansia*, on the other hand, are known to breakdown host-derived glycans, especially mucin. Mucin degradation has been linked to intestinal inflammation and can facilitate colonization of intestinal pathogens [40-42]. To test whether the genomic potential for these metabolic activities is linked to tumor incidence, we used the PICRUSt software package to predict the metagenomic content for each sample at the time of AOM injection. Butyrate production in the gut requires either butyryl-CoA:acetate-CoA transferase or butyrate kinase [30]. KEGG Orthologs (KOs) of the alpha and beta subunits of butyryl-CoA:acetate-CoA transferase were negatively correlated with tumor incidence ($r<-0.35$, $p<0.05$,). Butyrate kinase had the same trend, but the correlation was not statistically significant ($r=-0.30$, $p=0.08$). Next, we identified KOs for sialidases, fucosidases, sulfatases, and N-acetylglucosaminidases, which are indicative of host glycan degradation [36, 37]. Of the 10 such KOs found in our metagenomes, 7 (two arylsulfatases, an uncharacterized sulfatase, alpha-L-fucosidase, sialate O-acetylesterase, alpha-N-acetylglucosaminidase, 1,4-beta-poly-N-acetylglucosaminidase) were positively correlated with tumor count ($r>0.47$, $p <0.01$). None of the 3 remaining KOs were correlated with tumors. Together, these data suggest that the correlation between tumor incidence and the microbiome may be dependent on metabolic activity rather than bacterial phylogeny.

Discussion

The results of this study demonstrate that the structure of the gut microbiome is important for determining susceptibility to inflammation-associated tumorigenesis. We observed strong correlations between initial community structure of the gut microbiome and tumor multiplicity. This relationship is driven primarily by two distinct groups of bacteria. In general we found that members of the Bacteroidales (*Bacteroides, Parabacteroides, Alistipes,* and Porphrymonodaceae) were associated with a higher rate of tumorigenesis, while members of the Clostridiales, especially *Clostridium* Group XIVa, were associated with a decreased rate of tumorigenesis. There were exceptions to this pattern however, as a few OTUs associated with Clostridiales (OTUs associated with *Roseburia, Blautia,* and *Subdoligranulum*) were enriched in the groups with higher tumor counts (Figure 2.4B). However, these OTUs were less abundant (<0.7% mean abundance) than those Clostridales that were negatively correlated with tumors (~2% mean abundance). Therefore the data generally support a model in which susceptibility to colonic tumorigenesis is determined by the balance between the abundance of members of Bacteroidales and Clostridiales. One limitation of this study is that we only assayed the fecal communities. While this was necessary for correlating baseline community structure with the numbers of tumors that developed, characterization of the mucosal microbiota could potentially yield additional associations with tumor burden. It is also important to note that, although we observed variation in the number of tumors within inoculum groups, we were unable to correlated these differences with any differences in their microbiomes.

Based on our predicted metagenomic analysis, the roles of Clostridiales and Bacteroidales could be dependent on specific metabolic activities. Members of *Clostridium* Group XIVa are the predominant producers of butyrate in the gut [35]. Given the anti-inflammatory and anti-tumorigenic properties of intestinal butyrate, its production by members of *Clostridium* Group XIVa could explain the association with lower susceptibility to colon tumorigenesis [38, 39]. This hypothesis is supported by our predicted metagenomic data, which correlated the increased potential for butyrate production with decreased tumorigenesis. *Bacteroides* and *Akkermansia* were the two genera most strongly correlated with higher rates of tumorigenesis. Both are known mucin degraders, and several genes linked to mucin degradation were positively correlated with tumor incidence. Additionally, previous studies have linked mucin degradation by *Bacteroides* and *Akkermansia* with intestinal inflammation [40-42]. It is possible that an overabundance of these or other mucin degraders could undermine the integrity of the mucosal barrier, leading to increased inflammation. Such a mechanism could be an alternative to the ETBF-based model of tumorigenesis as we were unable to detect the gene for the ETBF toxin in any of our samples. While we cannot exclude the possibility of a novel toxin in the *Bacteroides* populations in our experiment, the additional correlation with *Akkermansia muciniphila* supports a model in which inflammation is induced by mucin degradation. If further experiments confirm this model, blocking mucin degradation could be a used as a therapeutic for preventing or slowing the progression of tumorigenesis.

In this study we observed a relationship between tumor multiplicity and the extent to which the microbiome shifted over the course of the model. The gut community of mice with high tumor counts changed very little over the course of the model, while the microbiome of groups with low tumor counts changed drastically. Thus the more similar the baseline community was to the endpoint community, the more tumors the host developed. We hypothesize that the microbiome of these mice was not significantly altered by the AOM/DSS model since it was already in a state of dysbiosis. Therefore, there was a greater exposure to a tumorigenic microbiome. Similarly, in a previous study we colonized germ-free mice with the feces of conventional mice that had already gone through the model [10]. These mice developed more tumors than germ-free mice colonized with feces from normal mice. Thus, in addition to needing a dysbiotic community to exacerbate tumorigenesis, the length of exposure to that community is important to tumor formation.

In contrast to earlier studies where human feces were used to colonize germ-free mice, we were unable to recapitulate the structures of the human microbiota donors, as numerous members of the donor community failed to colonize the recipients and others colonized in different abundances. For example, one of the donor communities (C1) was dominated by *Fusobacterium spp* (58% relative abundance). Another inoculum (C3), contained *F. nucleatum* at 2% relative abundance [2]. However we did not recover any sequences from the Fusobacteria phylum in the recipient mice. We were also unable to culture it from the original human stool sample, suggesting it may not have survived the freezing and thawing of the sample or was never alive in the stool. While

we did not fully recapitulate the community structure or phenotype of the human donors, colonizing mice with human fecal communities did serve as a useful tool for generating novel community structures to test the influence of specific bacterial populations on tumorigenesis. This strategy also allowed us to investigate the role of human microbiota, which should be more clinically relevant, while maintaining the tractability of a mouse model.

Conclusions

In this study we found that the process of colonizing germ-free mice with human fecal communities did not recapitulate the phenotype of the human donors in this particular mouse model of CRC. Nonetheless, our findings demonstrate the importance of the initial microbiome structure in determining the rate of tumorigenesis. Furthermore, we identified several bacterial populations correlated with tumor incidence in the context of six distinct gut communities. Multiple OTUs associated with the order Bacteroidales and the species *Akkermansia muciniphila* were correlated with exacerbated tumorigenesis, while several OTUs associated with *Clostridium* Group XIVa and other Clostridiales were correlated with protection. Based on inferred metagenomes of the baseline communities, we provided evidence that the positive correlations between *Akkermansia* and Bacteroidales and tumor incidence could be a result of their ability to degrade mucin, and the negative correlation between the Clostridiales and tumor incidence could be due to the production of butyrate. The results are consistent with a model in which susceptibility is determined by the balance between mucin degradation and short chain

fatty acid production. More studies are needed to confirm theses results and to test the mechanisms by which these or other bacterial populations influence colon tumorigenesis. A better understanding of microbiome structures with a propensity to promote or inhibit tumorigenesis could lead to the development of prebiotic or probiotic therapies to prevent or slow the development and progression of CRC.

References

1. American Cancer Society: **Colorectal Cancer Facts & Figures 2011-2013**. Atlanta; 2013.
2. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, Clancy TE, Chung DC, Lochhead P, Hold GL *et al*: **Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment**. *Cell Host Microbe* 2013, **14**(2):207-215.
3. Wu S, Rhee K-J, Albesiano E, Rabizadeh S, Wu X, Yen H-R, Huso DL, Brancati FL, Wick E, McAllister F *et al*: **A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses**. *Nat Med* 2009, **15**(9):1016-1022.
4. Arthur JC, Perez-Chanona E, Muhlbauer M, Tomkovich S, Uronis JM, Fan T-J, Campbell BJ, Abujamel T, Dogan B, Rogers AB *et al*: **Intestinal inflammation targets cancer-inducing activity of the microbiota**. *Science* 2012, **338**(6103):120-123.
5. Cheesman SE, Neal JT, Mittge E, Seredick BM, Guillemin K: **Epithelial cell proliferation in the developing zebrafish intestine is regulated by the Wnt pathway and microbial signaling via Myd88**. *Proc Natl Acad Sci U S A* 2011, **108 Suppl 1**:4570-4577.
6. Dolara P, Caderni G, Salvadori M, Morozzi G, Fabiani R, Cresci A, Orpianesi C, Trallori G, Russo A, Palli D: **Fecal levels of short-chain fatty acids and bile acids as determinants of colonic mucosal cell proliferation in humans**. *Nutr Cancer* 2002, **42**(2):186-190.

7. Stappenbeck TS, Hooper LV, Gordon JI: **Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells**. *Proc Natl Acad Sci U S A* 2002, **99**(24):15451-15455.

8. Huxley RR, Ansary-Moghaddam A, Clifton P, Czernichow S, Parr CL, Woodward M: **The impact of dietary and lifestyle risk factors on risk of colorectal cancer: a quantitative overview of the epidemiological evidence**. *Int J Cancer* 2009, **125**(1):171-180.

9. Chambers WM, Warren BF, Jewell DP, Mortensen NJM: **Cancer surveillance in ulcerative colitis**. *Br J Surg* 2005, **92**(8):928-936.

10. Zackular JP, Baxter NT, Iverson KD, Sadler WD, Petrosino JF, Chen GY, Schloss PD: **The gut microbiome modulates colon tumorigenesis**. *mBio* 2013, **4**(6):e00692-00613.

11. Wang T, Cai G, Qiu Y, Fei N, Zhang M, Pang X, Jia W, Cai S, Zhao L: **Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers**. *ISME J* 2012, **6**(2):320-329.

12. Chen H-M, Yu Y-N, Wang J-L, Lin Y-W, Kong X, Yang C-Q, Yang L, Liu Z-J, Yuan Y-Z, Liu F *et al*: **Decreased dietary fiber intake and structural alteration of gut microbiota in patients with advanced colorectal adenoma**. *Am J Clin Nutr* 2013, **97**(5):1044-1052.

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

14. Shen XJ, Rawls JF, Randall T, Burcal L, Mpande CN, Jenkins N, Jovov B, Abdo Z, Sandler RS, Keku TO: **Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas**. *Gut Microbes* 2010, **1**(3):138-147.

15. Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, Ojesina AI, Jung J, Bass AJ, Tabernero J *et al*: **Genomic analysis identifies association of Fusobacterium with colorectal carcinoma**. *Genome Res* 2012, **22**(2):292-298.

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

17. Toprak NU, Yagci A, Gulluoglu BM, Akin ML, Demirkalem P, Celenk T, Soyletir G: **A possible role of Bacteroides fragilis enterotoxin in the aetiology of colorectal cancer**. *Clin Microbiol Infec* 2006, **12**(8):782-786.

18. Alpert C, Sczesny S, Gruhl B, Blaut M: **Long-term stability of the human gut microbiota in two different rat strains**. *Current issues in molecular biology* 2008, **10**(1-2):17-24.

19. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI: **An obesity-associated gut microbiome with increased capacity for energy harvest**. *Nature* 2006, **444**(7122):1027-1031.

20. Koren O, Goodrich JK, Cullender TC, Spor A, Laitinen K, Backhed HK, Gonzalez A, Werner JJ, Angenent LT, Knight R *et al*: **Host remodeling of the gut**

microbiome and metabolic changes during pregnancy. *Cell* 2012, **150**(3):470-480.

21. Smith MI, Yatsunenko T, Manary MJ, Trehan I, Mkakosya R, Cheng J, Kau AL, Rich SS, Concannon P, Mychaleckyj JC *et al*: **Gut microbiomes of Malawian twin pairs discordant for kwashiorkor**. *Science* 2013, **339**(6119):548-554.
22. Franco AA, Cheng RK, Goodman A, Sears CL: **Modulation of bft expression by the Bacteroides fragilis pathogenicity island and its flanking region**. *Mol Microbiol* 2002, **45**(4):1067-1077.
23. Chen GY, Shaw MH, Redondo G, Nunez G: **The Innate Immune Receptor Nod1 Protects the Intestine from Inflammation-induced Tumorigenesis**. *Cancer Res* 2008, **68**(24):10060-10067.
24. Tanaka T, Kohno H, Suzuki R, Yamada Y, Sugie S, Mori H: **A novel inflammation-related mouse colon carcinogenesis model induced by azoxymethane and dextran sodium sulfate**. *Cancer Sci* 2003, **94**(11):965-973.
25. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD: **Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform**. *Appl Environ Microbiol* 2013, **79**(17):5112-5120.
26. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ *et al*: **Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities**. *Appl Environ Microbiol* 2009, **75**(23):7537-7541.

27. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glockner FO: **SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB**. *Nucleic Acids Res* 2007, **35**(21):7188-7196.

28. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R: **UCHIME improves sensitivity and speed of chimera detection**. *Bioinformatics* 2011, **27**(16):2194-2200.

29. Wang Q, Garrity GM, Tiedje JM, Cole JR: **Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy**. *Appl Environ Microbiol* 2007, **73**(16):5261-5267.

30. Yue JC, Clayton MK: **A similarity measure based on species proportions**. *Commun Stat-Theor M* 2005, **34**(11):2123-2131.

31. Holmes I, Harris K, Quince C: **Dirichlet Multinomial Mixtures: Generative Models for Microbial Metagenomics**. *Plos One* 2012, **7**(2).

32. Breiman L: **Random forests**. *Mach Learn* 2001, **45**(1):5-32.

33. Liaw A, Wiener M: **Classification and Regression by randomForest**. *R News* 2002, **2**:18-22.

34. Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R *et al*: **Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences**. *Nat Biotechnol* 2013, **31**(9):814-821.

35. Pryde SE, Duncan SH, Hold GL, Stewart CS, Flint HJ: **The microbiology of butyrate formation in the human colon**. *FEMS Microbiol Lett* 2002, **217**(2):133-139.

36. Corfield AP, Wagner SA, Clamp JR, Kriaris MS, Hoskins LC: **Mucin degradation in the human colon: production of sialidase, sialate O-acetylesterase, N-acetylneuraminate lyase, arylesterase, and glycosulfatase activities by strains of fecal bacteria**. *Infect Immun* 1992, **60**(10):3971-3978.

37. Martens EC, Chiang HC, Gordon JI: **Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont**. *Cell Host Microbe* 2008, **4**(5):447-457.

38. Segain JP, Raingeard de la Bletiere D, Bourreille A, Leray V, Gervois N, Rosales C, Ferrier L, Bonnet C, Blottiere HM, Galmiche JP: **Butyrate inhibits inflammatory responses through NFkappaB inhibition: implications for Crohn's disease**. *Gut* 2000, **47**(3):397-403.

39. Hague A, Elder DJ, Hicks DJ, Paraskeva C: **Apoptosis in colorectal tumour cells: induction by the short chain fatty acids butyrate, propionate and acetate and by the bile salt deoxycholate**. *Int J Cancer* 1995, **60**(3):400-406.

40. Bloom SM, Bijanki VN, Nava GM, Sun L, Malvin NP, Donermeyer DL, Dunne WM, Allen PM, Stappenbeck TS: **Commensal Bacteroides species induce colitis in host-genotype-specific fashion in a mouse model of inflammatory bowel disease**. *Cell Host Microbe* 2011, **9**(5):390-403.

41. Ng KM, Ferreyra JA, Higginbottom SK, Lynch JB, Kashyap PC, Gopinath S, Naidu N, Choudhury B, Weimer BC, Monack DM *et al*: **Microbiota-liberated**

host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* 2013, **502**(7469):96-99.

42. Ganesh BP, Klopfleisch R, Loh G, Blaut M: **Commensal Akkermansia muciniphila Exacerbates Gut Inflammation in Salmonella Typhimurium-Infected Gnotobiotic Mice**. *Plos One* 2013, **8**(9).

CHAPTER 3:

Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions

The contents of this chapter have been published as:

> Baxter NT, Ruffin MT IV, Rogers MAM, Schloss PD. (2016). Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. *Genome Medicine* 8(1):37.

Abstract

Colorectal cancer is the second leading cause of death among cancers in the United States. Although individuals diagnosed early have a greater than 90% chance of survival, more than one-third of individuals do not adhere to screening recommendations partly because the standard diagnostics, colonoscopy and sigmoidoscopy, are expensive and invasive. Thus, there is a great need to improve the sensitivity of non-invasive tests to detect early stage cancers and adenomas. Numerous studies have identified shifts in the composition of the gut microbiota associated with the progression of colorectal cancer, suggesting that the gut microbiota may represent a reservoir of biomarkers that would complement existing non-invasive methods such as

the widely used fecal immunochemical test (FIT). We sequenced the 16S rRNA genes from the stool samples of 490 patients. We used the relative abundances of the bacterial populations within each sample to develop a random forest classification model that detects colonic lesions using the relative abundance of gut microbiota and the concentration of hemoglobin in stool. The microbiota-based random forest model detected 91.7% of cancers and 45.5% of adenomas while FIT alone detected 75.0% and 15.7%, respectively. Of the colonic lesions missed by FIT, the model detected 70.0% of cancers and 37.7% of adenomas. We confirmed known associations of *Porphyromonas assaccharolytica*, *Peptostreptococcus stomatis*, *Parvimonas micra*, and *Fusobacterium nucleatum* with CRC. Yet, we found that the loss of potentially beneficial organisms, such as members of the Lachnospiraceae, was more predictive for identifying patients with adenomas when used in combination with FIT. These findings demonstrate the potential for microbiota analysis to complement existing screening methods to improve detection of colonic lesions.

Background

Colorectal cancer mortality has steadily declined in recent decades, due in large part to increased screening [1]. Yet current screening tests, the fecal immunochemical test (FIT) and the multitarget DNA test, have a sensitivity of 7.6% and 17.2%, respectively, for detecting non-advanced adenoma – just the type of early lesion that screening is meant to identify [2]. Although structural exams including colonoscopy and sigmoidoscopy are able to detect both adenomas and carcinomas, the high cost and invasive nature are barriers for many people. Fear, discomfort, and embarrassment are

among the most cited reasons patients choose to forego CRC screening [3]. Likewise the large disparity in screening rates between those with and without health insurance highlights the need for inexpensive screening methods [1, 4, 5]. Unfortunately cheaper, less invasive stool-based tests like guaic fecal occult blood test and FIT are unable to reliably detect adenomas [6]. The newly introduced stool DNA panel has improved accuracy compared to FIT, but is still limited in its ability to accurately detect adenomas [2]. Thus there is need for novel screening methods that are inexpensive and capable of detecting both cancer and adenomas.

The gut microbiota, the collection of microorganisms that inhabit the gastrointestinal tract, are one potential source of biomarkers for detecting colonic lesions. Numerous studies have observed alterations in the gut bacterial communities of patients with CRC [7–12]. Experiments in animal models have demonstrated that such alterations have the potential to accelerate tumorigenesis [13]. Furthermore, several members of the gut microbiota have been shown to potentiate both the development and progression of CRC by a variety of mechanisms [14–16]. Although each of these organisms may play a role in certain cases of CRC, none of them is present in every case. Therefore we postulate that no one organism is an effective biomarker on its own and that focusing on a single bacterial population excludes the potential that the microbial etiology of the disease is actually polymicrobial.

Two recent studies used statistical models that take into account the abundances of multiple bacterial species and the results of guaic fecal occult blood test (gFOBT) to distinguish healthy individuals from those with CRC [17, 18]. The analysis by Zackular et al. [17] used samples from a limited number of subjects (N=30 normal, 30 adenoma,

and 30 carcinoma), while that of Zeller et al [18] had a larger cohort from multiple clinical sites (N=156 and N=335). A shortcoming of the Zeller study was the pooling of subjects with non-advanced adenomas with control subjects as well as the exclusion of subjects with advanced adenomas. A limitation of both studies was that they relied on gFOBT rather than FIT to detect hemoglobin in stool. FIT provides a quantitatve measure of hemoglobin concentrations and has largely replaced gFOBT clinically because of its improved sensitivity. Regardless of their weaknesses, these studies demonstrated the feasibility of using microbiome data identify subjects with colonic lesions.

In the present study, we demonstrate the potential for microbiota analysis to complement FIT for improved detection of colonic lesions, especially adenomas. We utilized the random forest algorithm, which is a decision tree-based machine learning algorithm for classification that accounts for non-linear data and interactions among features and includes an internal cross-validation to prevent overfitting [19]. With this method we identified bacterial populations that could distinguish healthy individuals from those with adenomas or carcinomas. In doing so, we confirmed previously observed associations of certain bacterial taxa with CRC. Many lesions detected using the microbiota were distinct from those detected by FIT, suggesting the microbiota could complement FIT to improve sensitivity. By incorporating data on hemoglobin and bacterial abundances into a single model (labeled the Multitarget Microbiota Test or MMT), we were able to improve the sensitivity for adenomas and cancer compared to FIT alone.

Methods

Study Design/Patient sampling. Eligible patients for this study were at least 18 years old, willing to sign informed consent, able to tolerate removal of 58 mL of blood, and willing to collect a stool sample. Patient age at the time of enrollment ranged from 29 to 89 with a median of 60. All patients were asymptomatic and were excluded if they had undergone surgery, radiation, or chemotherapy for current CRC prior to baseline samples or had inflammatory bowel disease, known hereditary non-polyposis CRC, or familial adenomatous polyposis. Colonoscopies were performed and fecal samples were collected from subjects in 4 locations: Toronto (Ontario, Canada), Boston (Massachusetts, USA), Houston (Texas, USA), and Ann Arbor (Michigan, USA). Patient diagnoses were determined by colonoscopic examination and histopathological review of any biopsies taken. Patients with an adenoma greater than 1cm, more than three adenomas of any size, or an adenoma with villous histology were classified as advanced adenoma. Whole evacuated stool was collected from each patient either prior to colonoscopy preparation or 1-2 weeks after colonoscopy. This has been shown to be sufficient time for the microbiota to recover from colonoscopy preparation [20]. Stool samples were packed in ice, shipped to a processing center via next day delivery and stored at -80˚C. The University of Michigan Institutional Review Board approved this study, and all subjects provided informed consent. This study conformed to the guidelines of the Helsinki Declaration.

Fecal Immunochemical Tests. Fecal material for FIT was collected from frozen stool aliquots using OC FIT-CHEK sampling bottles (Polymedco Inc.) and processed using an

OC-Auto Micro 80 automated system (Polymedco Inc.). Hemoglobin concentrations were used for generating ROC curves for FIT and for building the MMT.

16S rRNA Gene Sequencing. DNA was extracted from approximately 50 mg of fecal material from each subject using the PowerSoil-htp 96 Well Soil DNA isolation kit (MO BIO Laboratories) and an epMotion 5075 automated pipetting system (Eppendorf). The V4 region of the bacterial 16S rRNA gene was amplified using custom barcoded primers and sequenced as described previously using an Illumina MiSeq sequencer [21]. The 490 samples were divided into three sequencing runs to increase the per sample sequencing depth. Although the same percentage of samples from the three groups were represented on each sequencing run, samples were randomly assigned to the sequencing runs to avoid confounding our analysis based on diagnosis or demographics.

Sequence Curation. The 16S rRNA gene sequences were curated using the mothur software package (v1.36), as described previously [21, 22]. Briefly, paired-end reads were merged into contigs, screened for quality, aligned to SILVA 16S rRNA sequence database, and screened for chimeras. Sequences were classified using a naive Bayesian classifier trained against a 16S rRNA gene training set provided by the Ribosomal Database Project (RDP) [23]. Curated sequences were clustered into operational taxonomic units (OTUs) using a 97% similarity cutoff with the average neighbor clustering algorithm. Species-level classifications for OTUs of interest were determined by blasting the predominant sequences within each OTU to the NCBI 16S rRNA database. The putative species was only reported for OTUs with greater than 99% sequence identity to a single species in the database; otherwise the consensus

RDP classification was used. The number of sequences in each sample was rarefied to 10,000 per sample to minimize the effects of uneven sampling. Only the 335 OTUs present in at least 5% of samples were included in the feature selection for the random forest models.

Statistical Methods. All statistical analyses were performed using R (v.3.2.0). Random Forest models were generated using the AUCRF package [24]. All ROC curves presented for random forest models are based on the out-of-bag (OOB) error rates. For each model, leave-one-out and 10-fold cross-validations were performed to further estimate the generalization error of the model. The AUC of ROC curves were compared using the method described by DeLong et al. [25]. The optimal cutoff for the MMT was determined using Youden's *J* statistic [26]. This cutoff was determined using the ROC curve for differentiating cancer from normal. Comparisons of sensitivities of FIT and the MMT at the same specificity were performed using the method developed by Pepe et al. with 1000 bootsrap replicates [27]. All of the aforementioned statistics for analyzing ROC curves were performed using the pROC package in R [28]. To control for diagnosis while testing the effects of sex on the microbiome we used PERMANOVA as implemented in the adonis function in the vegan R package [29].

Results

Complementary detection of lesions by FIT and the microbiota. We characterized the bacterial communities of stool samples from 490 patients using 16S rRNA gene sequencing. Among these patients, 120 had CRC, 198 had adenomas, and 172 had no colonic lesions. In addition to characterizing the bacterial community, we tested each

sample for the concentration of hemoglobin using FIT. With these data we compared the ability to detect lesions using FIT to using a microbiota-based model. First we developed a random forest classification model for differentiating healthy individuals from those with adenomas based on the relative abundance of bacterial populations in stool. We determined the optimal model using the AUC-RF algorithm for maximizing the area under the curve (AUC) of the receiver operating characteristic (ROC) curve for a random forest model [24]. The optimal model utilized 22 bacterial populations (Fig. 3.2A). The vast majority of OTUs in the model (17 out of 22) belonged to the order Clostridales, 4 were associated with the genus *Bacteroides*, and one OTU was unclassified at the phylum level (Fig. 3.2B). The AUC for this and subsequent random forest models were generated based on the out-of-bag (OOB) probabilities for each sample. Additional leave-one-out and 10-fold cross validations showed no significant difference in AUC compared to the OOB AUC (Fig. 3.3A). The AUC for the microbiota model (0.673) was significantly different from a random assignment ($p<0.001$), but not significantly different from that of FIT (FIT AUC:0.639, $p>0.05$, Fig. 3.1A). At the 100 ng/ml cutoff FIT detected 15.7% of adenomas with a specificity of 97.1%. Setting the microbiota model to the same 97.1% specificity resulted in 18.2% sensitivity for adenomas. When comparing the results of the tests for each sample, only 2.5% of adenomas were detected by both tests, while 28.8% were detected by only one of the two tests (Fig. 3.1B). Thus, the two tests detected small but distinct subsets of adenomas.

Next we generated a random forest model for differentiating normal individuals from those with cancer using the relative abundance of 34 bacterial populations (Fig. 3.4A,

Fig. 3.4B). Consistent with previous observations, the bacteria most strongly associated with CRC belonged to taxa commonly associated with periodontal disease [18, 30, 31]. These include OTUs associated *Pophyromonas assaccharolytica* (OTU105), *Fusobacterium nucleatum* (OTU264), *Parvimonas micra* (OTU281), *Peptostreptococcus stomatis* (OTU310), *Gemella spp.* (OTU356), and an unclassified *Prevotella* (OTU57) (Fig. 3.4C). The ROC curve for the model had an AUC of 0.847, which was similar to AUCs reported for other microbiota-based models for CRC [17, 18]. The AUC of this model was significantly better than a random assignment ($p<0.001$), but was significantly lower than that of FIT (FIT AUC:0.929, $p=0.005$, Fig. 3.1C). As with the adenoma versus normal model, we confirmed the OOB AUC with leave-one-out cross validation and 100 iterations of 10-fold cross validation (Fig. 3.3B). At the manufacturer recommended cutoff of 100 ng/ml FIT detected 75.0% of cancers with a specificity of 97.1%. At the same specificity the microbiota model detected 51.7% of cancers. Although more cancers were detected by FIT, the microbiota model was able to detect 33.3% of cancers missed by FIT (Fig. 3.1D).

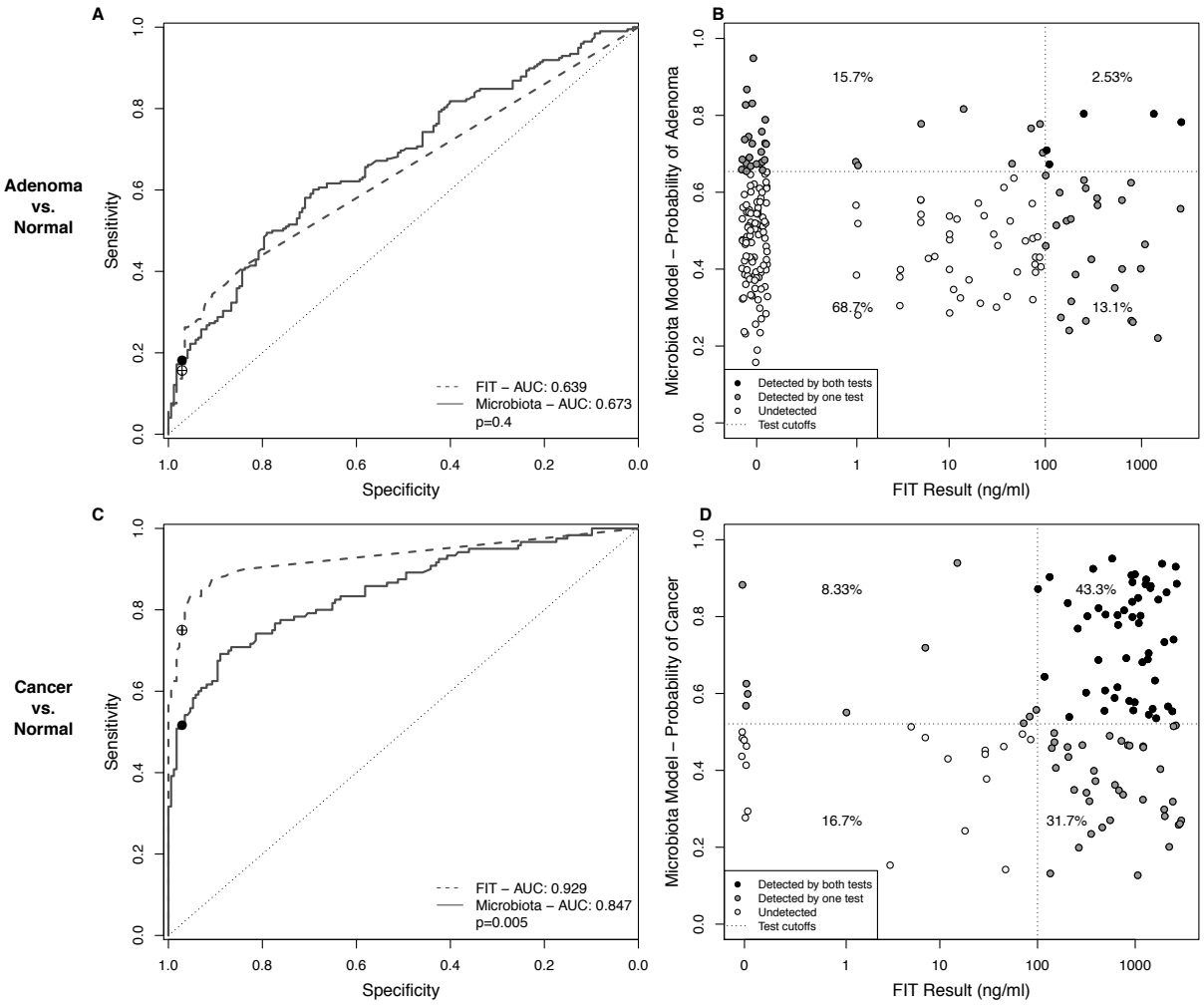

Figure 3.1. Microbiota-based models can complement FIT. (A,C) ROC curves for distinguishing healthy patients from those with adenoma (A) or cancer (C) based on FIT or a microbiota-based random forest model. Open circles show the sensitivity and specifity of FIT with a 100 ng/ml cutoff. Black points show the sensitivity and specificity of the microbiota-based models at the chosen cutoffs. (B,D) Results of FIT and a microbiota-based model for each adenoma (B) or cancer (D) sample. Dotted lines represent the cutoffs for each test. Points are shaded based on whether the lesion was detected by both tests (black), one of the two tests (grey), or neither test (white).

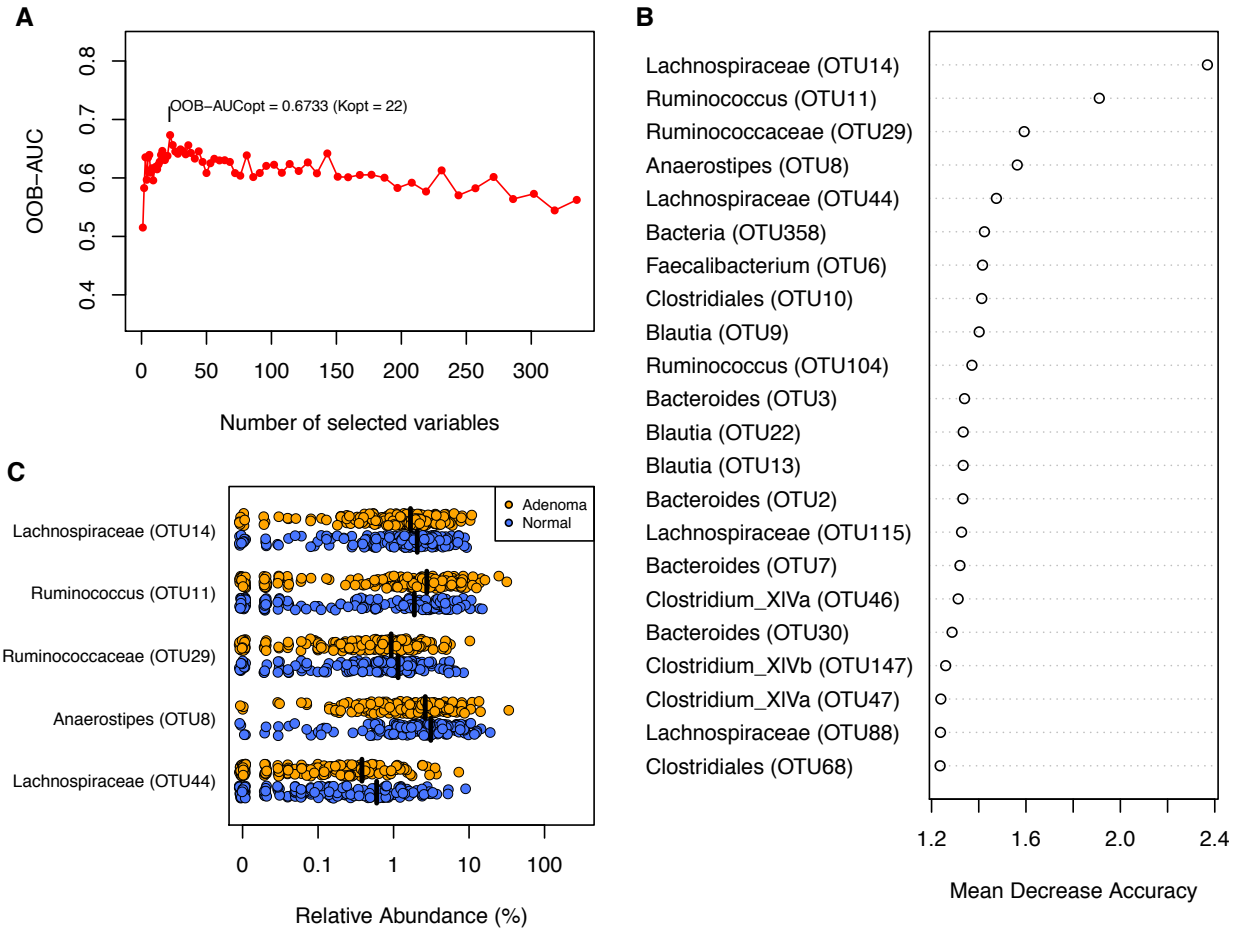

Figure 3.2. Random forest feature selection for detecting adenomas. (A) Change in AUC with varying number of variables in the random forest model. The model with the highest AUC contained 22 OTUs. (B) Importance of each OTU in the model as measured by mean decrease accuracy when the OTU is removed from the model. (C) Relative abundance of the most discriminatory OTUs in adenoma and normal samples.

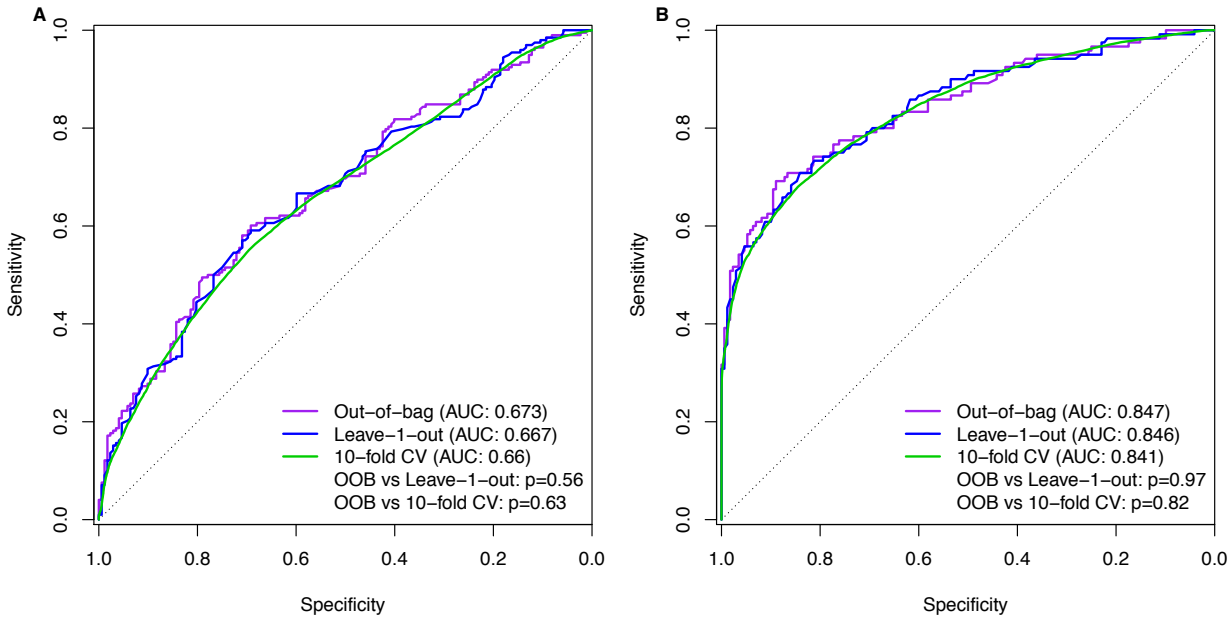

Figure 3.3. Cross validation of OTU random forest models. ROC curves for the (A) adenoma versus normal OTU model and (B) cancer versus normal OTU model based on OOB estimates, leave-one-out cross validation, and ten-fold cross validation.

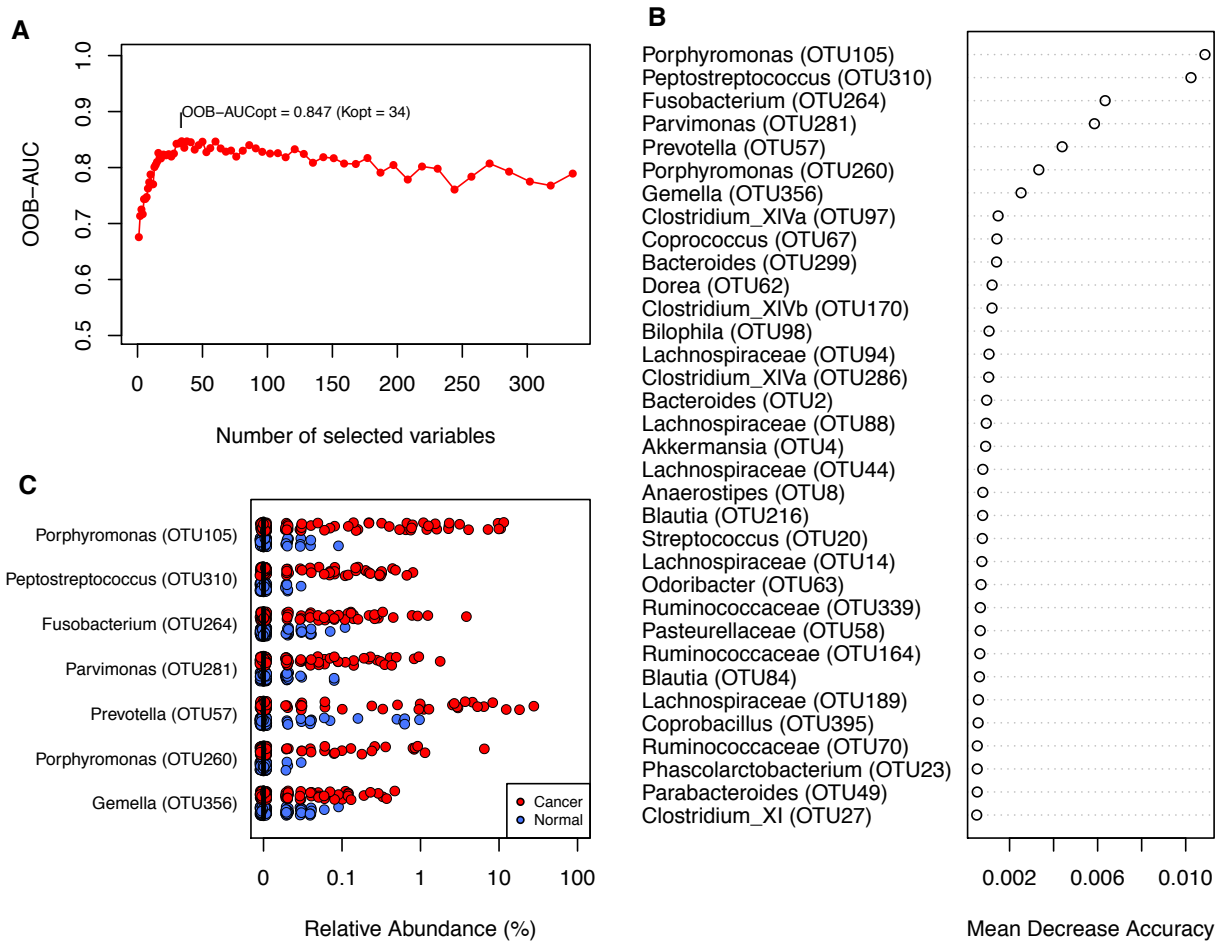

Figure 3.4. Random forest feature selection for detecting cancers. (A) Change in AUC with varying number of variables in the random forest model. The model with the highest AUC contained 34 OTUs. (B) Importance of each OTU in the model as measured by mean decrease accuracy when the OTU is removed from the model. (C) Relative abundance of the most discriminatory OTUs in cancer and normal samples.

Multitarget Microbiota Test for colonic lesions. Many of the adenomas and some of the carcinomas were detected by the microbiota models, but not FIT, suggesting that the two screening methods could complement each other if combined into a single test. Based on these observations, we developed a random forest model using both the microbiota and FIT that would differentiate normal individuals from those with any type of colonic lesion (i.e. adenoma or carcinoma). The optimal model, referred to as the Multitarget Microbiota Test (MMT), used the relative abundances of 23 OTUs and the concentration of hemoglobin as determined by FIT. Of those OTUs, 16 were members of the Firmicutes phylum, including 3 from the Ruminococcaceae family and 10 from the Lachnospiraceae family (Fig. 3.5). Three OTUs were associated with the genus *Bacteroides*. The remaining OTUs were associated with *Porphyromonas*, *Parabacteroides*, *Collinsella*, and Enterobacteriaceae. The OTU associated with *Porphyromonas* was most closely related to *Porphyromonas asaccharolytica*, which has been previously shown to be predictive of CRC [17, 18, 32]. Interestingly the majority of OTUs used in the model, especially the Lachnospiraceae, were enriched in normal patients (Fig. 3.5), suggesting that a loss of beneficial organisms in addition to the emergence of pathogens may be indicative of CRC development. As with the previous random forest models we performed leave-one-out cross validation and 100 iterations of 10-fold cross validation and found no difference in AUC compared to the OOB estimates (Fig. 3.6).

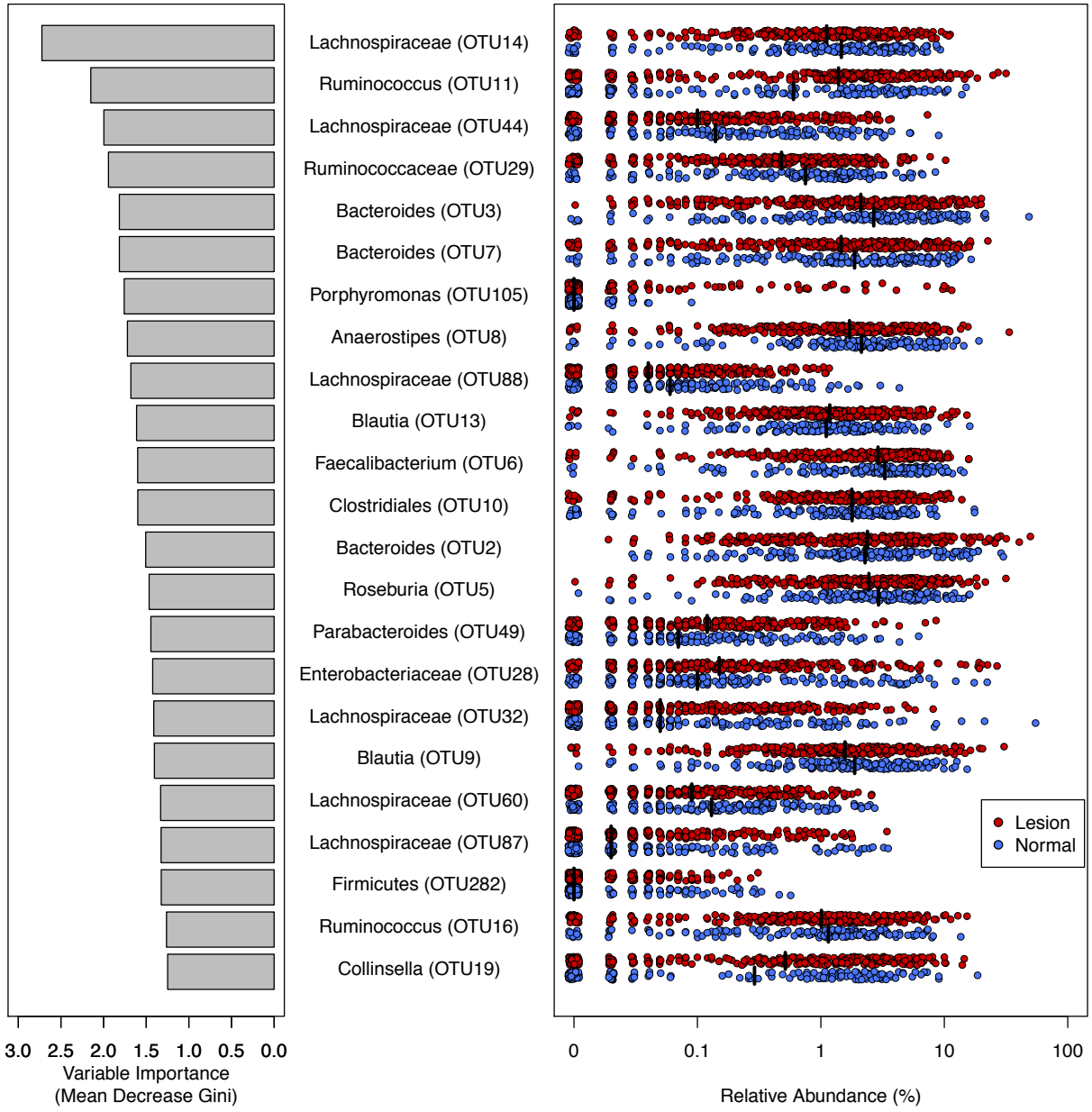

Figure 3.5. Bacterial OTUs in MMT. (left) Importance of each OTU used in the MMT as measured by the mean decrease in the Gini index when the OTU is removed from the model. (right) Stripchart of the relative abundances of each OTU in the MMT with black lines at the medians.

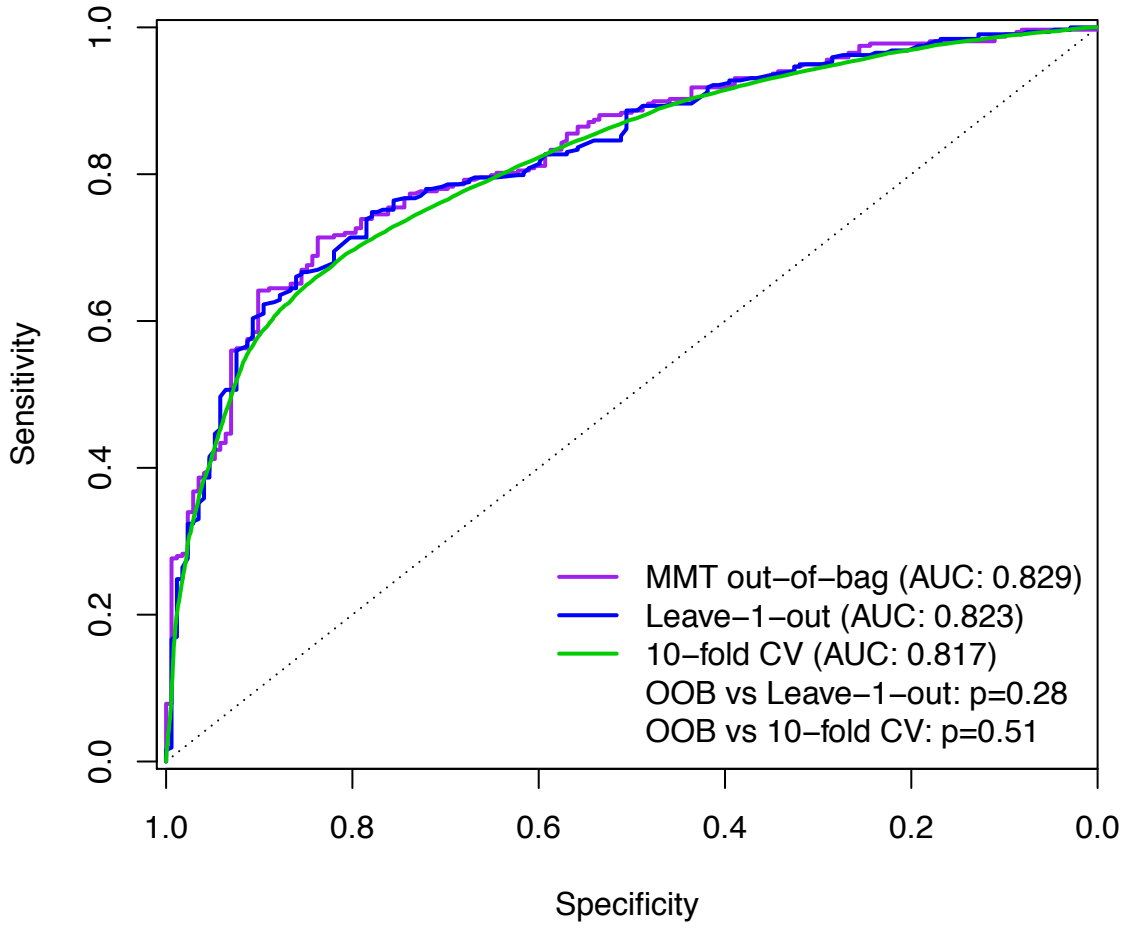

Figure 3.6. Cross validation of MMT. ROC curves for the MMT model based on OOB estimates, leave-one-out cross validation, and ten-fold cross validation.

Comparing MMT to FIT. To determine whether microbiota sequence data could be used to complement FIT, we compared the performance of the MMT to FIT. For differentiating any lesions from normal, the AUC for the MMT was significantly higher than FIT (MMT AUC:0.829, FIT AUC:0.749, $p<0.001$, Fig. 3.7A). Subdividing the lesions, detecting adenomas by the MMT (AUC:0.755) was significantly better than FIT (AUC:0.639, $p<0.001$), but not for differentiating cancer from normal (MMT AUC:0.952, FIT AUC:0.929, $p=0.09$). To generate a categorical prediction from the MMT, we determined the model's optimal threshold for detecting cancer (0.57 probability of a lesion) using Youden's J statisitc [26]. Samples scoring above this cutoff were classified as lesions, and those below the cutoff were classified as normal. We then compared the sensitivity and specificity of the MMT to those of FIT using a threshold of 100 ng/ml of hemoglobin. At these cutoffs the MMT detected 91.7% of cancers and 45.5% of adenomas compared to 75.0% and 15.7% for FIT (Table 1, Fig. 3.7B, Fig. 3.7C). When adenomas and cancers were pooled together, the MMT detected 62.9% of lesions, while FIT only detected 38.1%. However, the increased sensitivity of the MMT was accompanied by a decrease in specificity (90.1%) compared to FIT (97.1%).

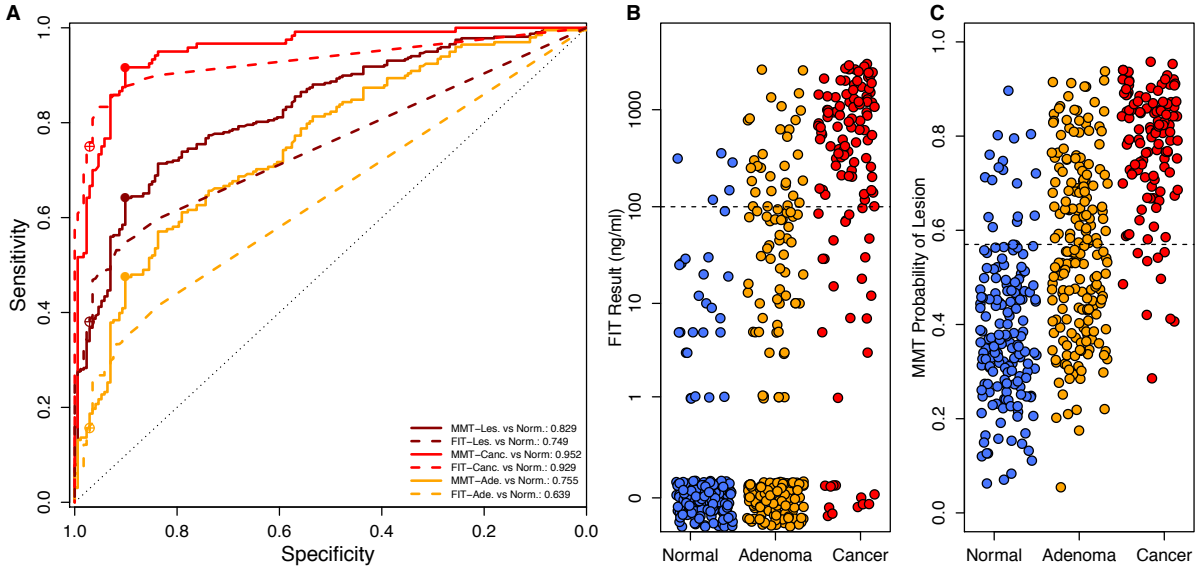

Figure 3.7. Comparing MMT to FIT. (A) ROC Curves for the MMT (solid lines) or FIT (dashed lines) for distinguishing normal from any lesion (dark red), normal from cancer (red) and normal from adenoma (orange). Filled dots show the sensitivity and specificity of the MMT at the optimal cutoff (0.622). Open dots show the sensitivity and specificity of FIT at the 100 ng/ml cutoff. (B,C) Stripcharts showing the results for FIT (B) and the MMT (C). Dashed lines show the cutoff for each test. Points with a FIT result of 0 are jittered to improve visibility.

Diagnosis		Fecal Immunochemical Test		Multitarget Microbiota Test	
		True Positives	Sensitivity (95% CI)	True Positives	Sensitivity (95% CI)
Cancer	n=120	90	**75.0** (67.5-82.5)	110	**91.7** (86.7-958)
Adenoma	n=198	31	**15.7** (10.6-20.7)	90	**45.5** (38.4-52.5)
Any Lesions	n=318	121	**38.1** (32.7-43.4)	200	**62.9** (57.2-67.9)
		True Negatives	Specificity (95% CI)	True Negatives	Specificity (95% CI)
Normal	n=172	167	**97.1** (94.2-99.4)	155	**90.1** (85.5-94.2)

Table 3.1. Sensitivities and specificities for FIT and MMT. The 95% confidence intervals were computed with 2000 stratified bootstrap replicates.

To better understand the relationship between the MMT and FIT, we compared the results of the two tests for each sample (Fig. 3.8A). All but one of the samples that tested positive by FIT also tested positive by the MMT. However the MMT was able to detect 70.0% of cancers and 37.7% of adenomas that FIT had failed to detect, while maintaining a specificity of 92.8% (Fig. 3.8B). This result demonstrated that incorporation of data from a subject's microbiota could complement FIT to improve its sensitivity.

To make a fairer comparison of the sensitivities of these two tests, we reduced the cutoff for FIT to 7 ng/ml to match the 90.1% specificity of the MMT. At the lower cutoff for FIT there was no significant difference in sensitivity for cancer between the two tests ($p=0.2$), but the MMT remained significantly more sensitive for detecting adenomas ($p=0.02$) and all lesions grouped together ($p=0.04$, Fig. 3.9).

The purpose of screening is to identify asymptomatic individuals with early stage disease (i.e., true positives). Therefore, we estimated the number of true positives captured through FIT and MMT in the recommended screening population in the United States (adults ages 50-75 years). The prevalence of lesions in an average-risk population was obtained through a previously published meta-analysis [33]. Based on sensitivities of FIT and MMT in our dataset, we estimate that MMT would detect approximately 40 thousand additional cancers, 1.3 million additional advanced adenomas, and 5.1 million additional non-advanced adenomas compared to using FIT (Table 2). Thus the improved sensitivity of the MMT would increase the total number of true positives identified in the recommended screening population of the United States by approximately 6.5 million. However, due to the lower specificity of MMT, it would also

result in an estimated 4.3 million additional false positives compared to FIT. Further studies would be needed to determine whether detection of 6.5 million additional lesions (mostly non-advanced adenomas) would outweigh the added cost of 4.3 million additional false positives.

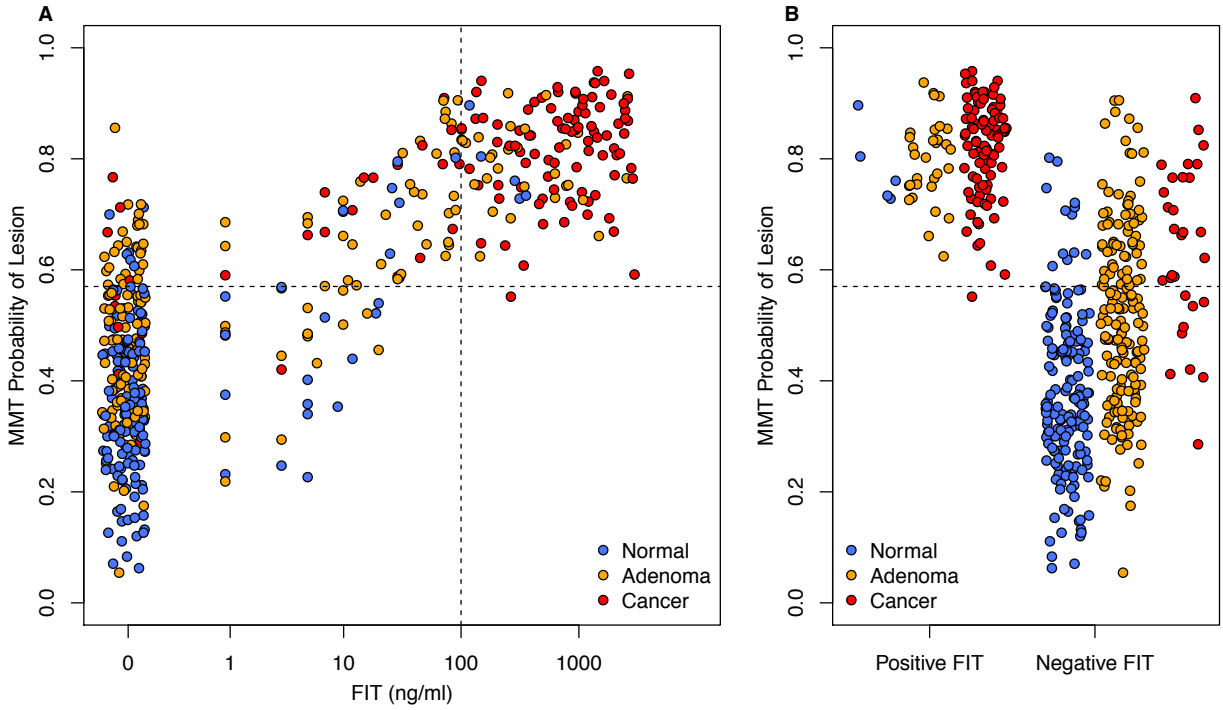

Figure 3.8. Relationship between FIT and MMT for each sample. (A) Scatterplot of MMT and FIT results for each sample. Dashed lines show the cutoff for each test. Points with a FIT result of 0 are jittered to improve visibility. (B) Stripchart of MMT results for samples separated by binary FIT result.

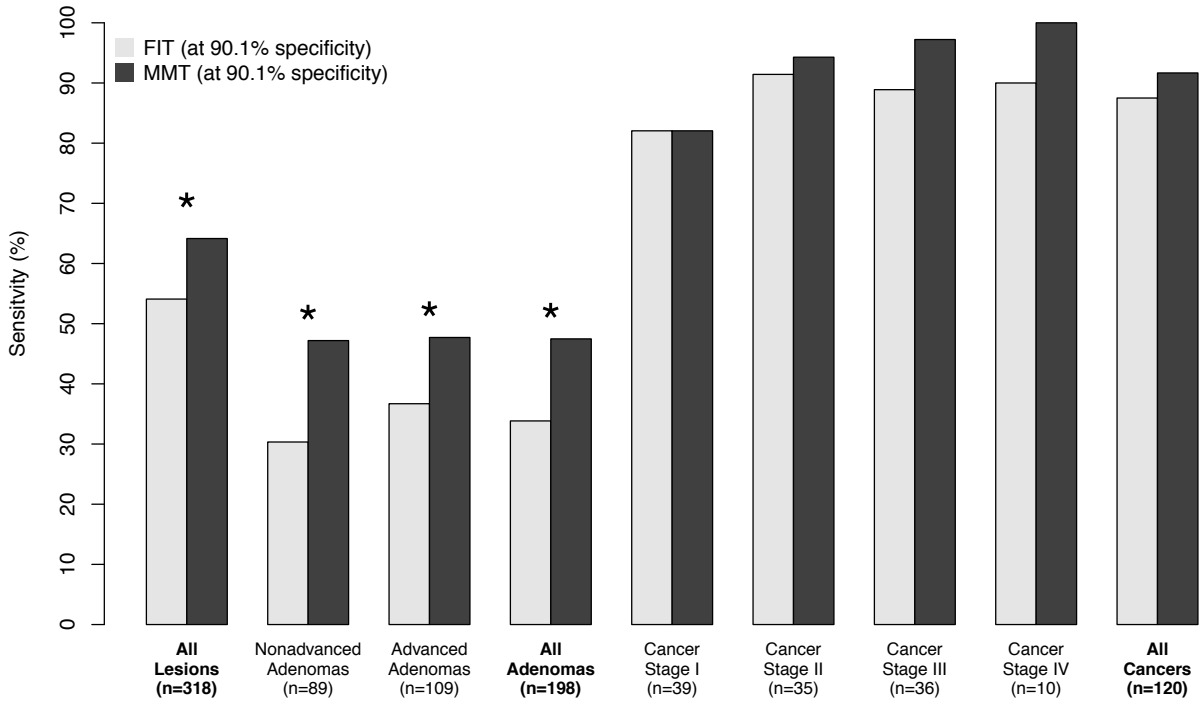

Figure 3.9. Sensitivities for FIT and MMT for each stage of tumor development with matching specificities. The cutoff for FIT was reduced to 7 ng/ml to match the specificity of the MMT. Sensitivities were compared using the method proposed by Pepe et al. (* = $p<0.05$, 1000 bootstrap replicates).

Condition	Prevalence	Number of Persons, ages 50-75 years, with Condition	True Positives identified by FIT	True Positives identified by MMT
Cancer	0.3%	241,483	181,112	221,359
Advanced Adenoma	5.7%	4,588,174	883,960	2,188,854
Non-advanced Adenoma	17.7%	14,247,488	1,600,841	6,723,534

*Number of persons in the United States in 2010, 50-75 years of age, was 80,494,283.

Table 3.2. Estimated number of true positives detected in average risk population.

Number of true positives identified through FIT and MMT in the United States in adults 50-75 years of age, based on published estimates of CRC prevalence. The sensitivities for FIT (100 ng/ml cutoff) on advanced and non-advanced adenomas were 19.3% and 11.2%, respectively.

Effect of patient characteristics on model performance. Previous studies have identified differences in diagnostic test performance for certain demographic groups or for people taking certain medications [34–36]. Therefore we tested whether the MMT performance differed between patient populations. We found no difference in model performance according to age, BMI, NSAID usage, diabetes, smoking, or previous history of polyps (all $p>0.05$). However the model was significantly better at differentiating normal from lesion for females than for males ($p=0.02$; Fig. 3.10). For females the model detected 63.6% of lesions with a specificity of 94.6%. For males the model detected 64.5% of lesions with a much lower specificity of 82%. The MMT detected 51.2% of adenomas in females and 44.9% in males. Consistent with the lower specificity for males, the MMT had a higher sensitivity for cancer among males (98.5%) than females (82.7%). The discrepancy appeared to be due to differences in FIT results rather than differences in the microbiome. After correcting for diagnosis, there was a significant effect of sex on FIT result ($p=0.006$, two-way ANOVA), but not on the overall structure of the microbiome (PERMANOVA: $p=0.07$). The lower specificity and higher sensitivity for cancer among males is consistent with previous observations that males have a higher positive rate for FIT [34, 35].

We have previously shown that incorporating patient metadata into microbiome-based diagnostic models can improve screening accuracy [17]. To test whether the same was true for the MMT we generated a random forest model that combined patients' age, BMI, sex, and smoking status with the OTUs and FIT result from the MMT. The AUC of the ROC curve for this model (0.869) was not significantly different from that of the MMT (AUC: 0.829, $p=0.11$, Fig. 3.11). When the model with patient metadata was set to the

same specificity as the MMT (90.1%), it did not improve the sensitivity for lesions (63.4%) compared to MMT (62.9%, p=0.9). Thus, contrary to our previous findings, incorporation of patient metadata did not significantly improve the MMT.

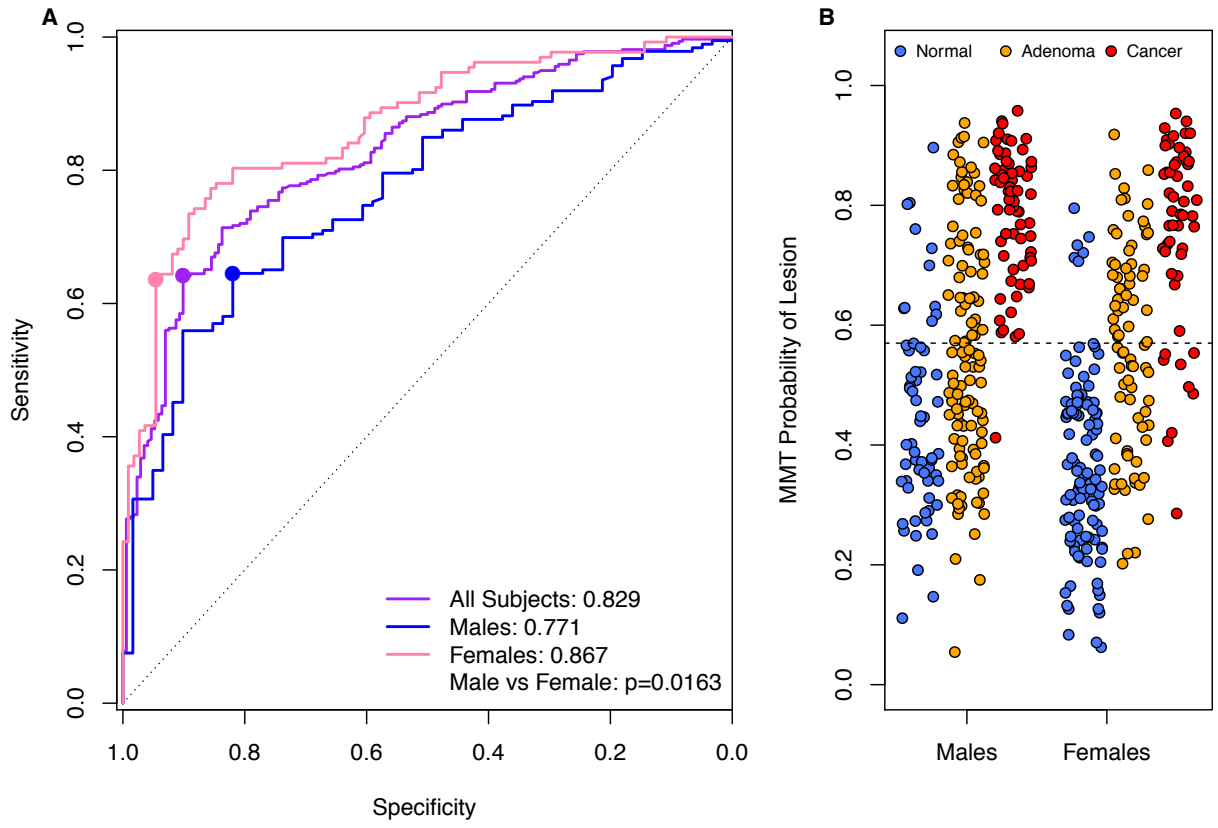

Figure 3.10. MMT performance by sex. ROC curves (left) and stripchart (right) of MMT results separated by sex.

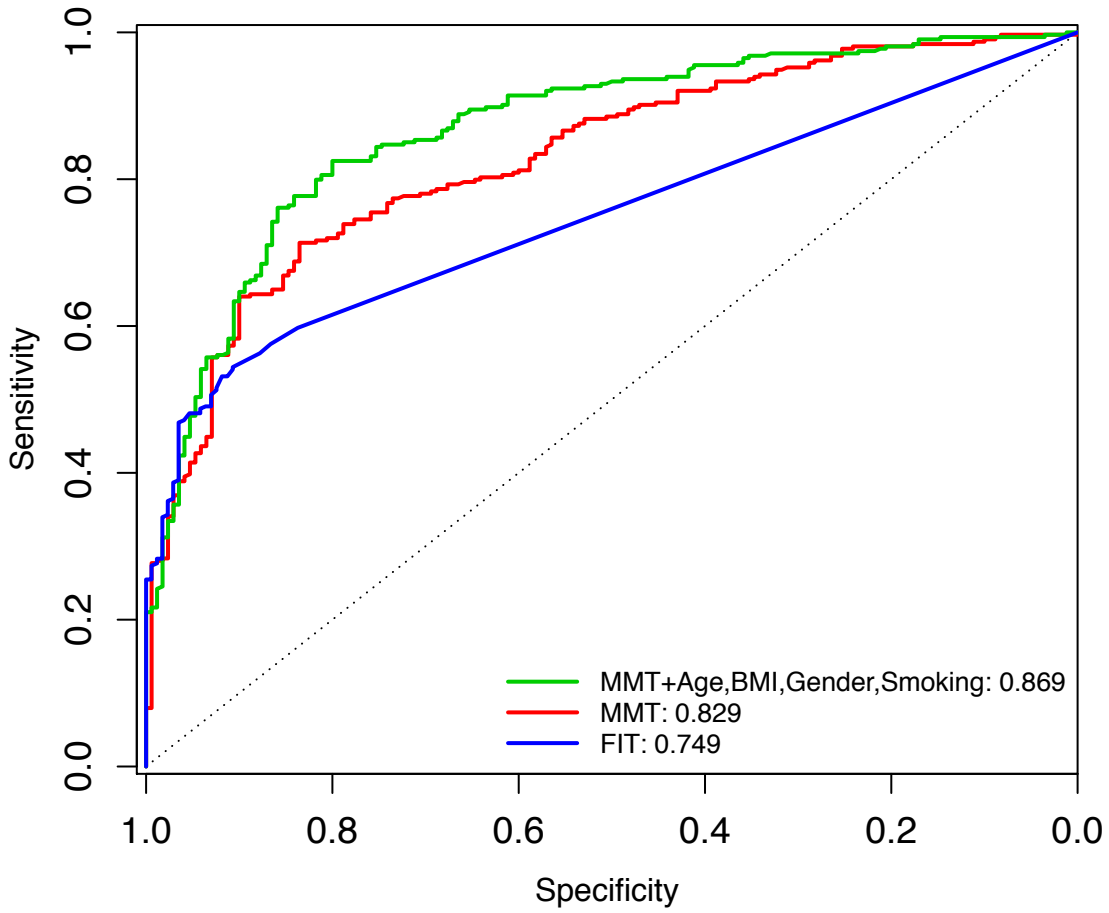

Figure 3.11. MMT with patient metadata. ROC curves for distinguishing normal from lesion using FIT, the MMT, or the MMT with metadata.

Discussion

We confirmed previous findings that the gut microbiota can be used to differentiate healthy individuals from those with colonic lesions. Although FIT was better at detecting cancers than a model using only the microbiota, microbiota-based models detected a subset of lesions that were not detected by FIT. This suggested that the two methods could complement each other. Based on this observation we developed a cross-validated random forest model that combined both FIT and the microbiota to detect colonic lesions. The resulting MMT had higher sensitivity than FIT for detecting lesions, especially adenomas. The MMT was also able to detect the majority of cancers missed by FIT. However, the increased sensitivity of MMT was accompanied by a decrease in specificity compared to FIT. With a false positive rate more than three times higher than FIT (9.9% versus 2.9%), an annual MMT would result in more colonoscopies than using FIT as the primary screening test. However, the higher sensitivity of the MMT might make it possible to reduce the frequency of screening, thereby offsetting the difference in the number of colonoscopies. Additional studies would be needed identify the appropriate screening inverval and to determine whether the increased number of true positives identified by MMT justify the increased number of false positives.

It was recently shown that when FIT was combined with host-associated DNA biomarkers, the ability to detect adenomas and carcinomas was significantly improved over FIT alone [2]. The sensitivity of the host-associated DNA screen was 92.3% for cancer and 42.4% for adenomas with a specificity of 89.8%, all very similar to what we observed with our MMT. Such results support the assertion that because of the large

interpersonal variation in markers for adenomas and carcinomas, it is necessary to employ a panel of biomarkers and to use a model that integrates the biomarkers. The accuracy of our model may be further improved by incorporating additional indicators such as host-associated biomarkers or those targeting specific genes involved in the underlying mechanism of tumorigenesis such as bacterial toxins [15, 16, 18]. More generally, predictive and diagnostic models for other diseases with a microbial etiology may benefit from a similar approach. For example, we recently demonstrated the ability to detect *Clostridium difficile* infection based on the composition of the microbiota [37]. Such models are likely to be useful as microbiota sequencing gains traction as a tool for characterizing health.

Surprisingly most of the OTUs that work well for identifying cancers, including *Fusobacterium nucleatum* (OTU264), *Peptostroptococcus stomatis* (OTU310), and *Parvimonas micra* (OTU281), were excluded from the MMT. This is likely due to these OTUs being positively correlated with FIT (all $p<0.001$, Spearman correlation), meaning they add little information when used in combination with FIT. Instead the MMT is enriched for OTUs that help detect adenomas. Thus the MMT model relies primarily on FIT for detecting cancer, and uses the microbiota to help identify adenomas undetectable by FIT alone. It is also interesting that most of the OTUs used in the MMT were enriched in normal individuals, suggesting that a loss of beneficial organisms in addition to the emergence of pathogens may be important for colorectal cancer development. Many of the OTUs that were depleted in patients with lesions belonged to the Ruminococcoaceae and Lachnospiraceae families, which contain the predominant producers of butyrate, a short-chain fatty acid with anti-inflammatory and anti-

tumorigenic properties [38–41]. Likewise Zeller et al. observed a depletion of a potential butyrate-producing *Eubacterium spp.* in patients with CRC [18]. Loss of butyrate or other anti-inflammatory microbial metabolites may contribute to CRC development. These possibilities highlight the need for longitudinal studies to better understand how changes to an individual's microbiome or the metabolic profile of the gut might predispose them to CRC.

Like other groups, we noticed that the microbiota of CRC patients contained higher levels of bacterial taxa traditionally thought of as oral pathogens, including *Fusobacterium*, *Porphyromonas*, *Peptostreptococus*, *Gemella*, *Parvimonas*, and *Prevotella*. Periodontal pathogens have been shown to promote the progression of oral cancer [42]. Therefore it is possible that these taxa could influence the progression of CRC by a similar mechanism. These observation may warrant further investigation into a potential link between periodontal disease and CRC. Furthermore, since the structure of an individual's oral microbiome is correlated with that of the gut [43], alterations in the oral community could potentially be a proxy for ongoing or future changes to the gut community.

Although it is exciting that the addition of the microbiota can improve the sensitivity of FIT, further validation is needed prior to clinical adoption. This represents the largest cohort to date, but still only consists of 490 patients. In contrast, the cohort used to validate the Multitarget stool DNA test included 9,989 subjects. Development of a larger cohort will allow us to apply the MMT to a separate validation set. It is also unclear how sensitive the MMT is to variation in sample preparation and processing. Many of the samples included in the current study were collected 1-2 weeks after the subjects'

colonoscopy. A previous study showed that the microbiome quickly returns to normal following colonoscopy [20]. Likewise we found no difference in the microbiome between samples collected prior to or after colonoscopy (PERMANOVA: p=0.45). Regardless, we would have greater confidence in the predictive potential of the microbiota if all samples were collected prior to colonoscopy. Despite these shortcomings, the ability to improve the sensitivity of detecting adenomas suggests that further methods development and validation are warranted.

Conclusions

Our findings demonstrate the potential for combining the analysis of a patient's microbiota with conventional stool-based tests to improve CRC detection. Using the random forest algorithm it was possible to interpret FIT results in the context of the microbiota. The MMT had higher sensitivity for lesions, especially at early stages of tumorigenesis. Moreover the model detected the majority of cancers that FIT was unable to detect. The shortcoming of the MMT is its lower specificity. However, the potential value of the MMT is its higher sensitivity, which is the purpose of preventive screening – finding lesions earlier so that cancer would be avoided.

References

1. Siegel R, DeSantis C, Jemal A: **Colorectal cancer statistics, 2014**. *CA: a cancer journal for clinicians*. 2014, **64**:104–117.

2. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, Ahlquist DA, Berger BM: **Multitarget stool DNA testing for colorectal-cancer screening**. *New England Journal of Medicine*. 2014, **370**:1287–1297.

3. Jones RM, Devers KJ, Kuzel AJ, Woolf SH: **Patient-reported barriers to colorectal cancer screening: a mixed-methods analysis**. *American journal of preventive medicine*. 2010, **38**:508–516.

4. Hsia J, Kemper E, Kiefe C, Zapka J, Sofaer S, Pettinger M, Bowen D, Limacher M, Lillington L, Mason E, others: **The importance of health insurance as a determinant of cancer screening: evidence from the Women's Health Initiative**. *Preventive medicine*. 2000, **31**:261–270.

5. Centers for Disease Control and Prevention: **Vital signs: Colorectal cancer screening test use–United states, 2012.** *MMWR Morbidity and mortality weekly report*. 2013, **62**:881.

6. Hundt S, Haug U, Brenner H: **Comparative evaluation of immunochemical fecal occult blood tests for colorectal adenoma detection**. *Annals of Internal Medicine*. 2009, **150**:162–169.

7. Wang T, Cai G, Qiu Y, Fei N, Zhang M, Pang X, Jia W, Cai S, Zhao L: **Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers**. *The ISME journal*. 2012, **6**:320–329.

8. Chen H-M, Yu Y-N, Wang J-L, Lin Y-W, Kong X, Yang C-Q, Yang L, Liu Z-J, Yuan Y-Z, Liu F, others: **Decreased dietary fiber intake and structural alteration of gut microbiota in patients with advanced colorectal adenoma**. *The American journal of clinical nutrition*. 2013, **97**:1044–1052.

9. Chen W, Liu F, Ling Z, Tong X, Xiang C: **Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer**. *PloS one*. 2012, **7**:e39743.

10. Shen XJ, Rawls JF, Randall TA, Burcall L, Mpande C, Jenkins N, Jovov B, Abdo Z, Sandler RS, Keku TO: **Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas**. *Gut microbes*. 2010, **1**:138–147.

11. Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, Ojesina AI, Jung J, Bass AJ, Tabernero J, others: **Genomic analysis identifies association of Fusobacterium with colorectal carcinoma**. *Genome research*. 2012, **22**:292–298.

12. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, Zhang D, Xia H, Xu X, Jie Z, others: **Gut microbiome development along the colorectal adenoma–carcinoma sequence**. *Nature communications*. 2015, **6**.

13. Zackular JP, Baxter NT, Iverson KD, Sadler WD, Petrosino JF, Chen GY, Schloss PD: **The gut microbiome modulates colon tumorigenesis**. *MBio*. 2013, **4**:e00692–13.

14. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, Clancy TE, Chung DC, Lochhead P, Hold GL, others: **Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment**. *Cell host & microbe*. 2013, **14**:207–215.

15. Wu S, Rhee K-J, Albesiano E, Rabizadeh S, Wu X, Yen H-R, Huso DL, Brancati FL, Wick E, McAllister F, others: **A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses**. *Nature medicine*. 2009, **15**:1016–1022.

16. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan T-J, Campbell BJ, Abujamel T, Dogan B, Rogers AB, others: **Intestinal inflammation targets cancer-inducing activity of the microbiota**. *science*. 2012, **338**:120–123.

17. Zackular JP, Rogers MA, Ruffin MT, Schloss PD: **The human gut microbiome as a screening tool for colorectal cancer**. *Cancer Prevention Research*. 2014, **7**:1112–1121.

18. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Böhm J, Brunetti F, Habermann N, others: **Potential of fecal microbiota for early-stage detection of colorectal cancer**. *Molecular systems biology*. 2014, **10**:766.

19. Liaw A, Wiener M: **Classification and regression by randomForest**. *R news*. 2002, **2**:18–22.

20. O'Brien CL, Allison GE, Grimpen F, Pavli P: **Impact of Colonoscopy Bowel Preparation on Intestinal Microbiota**. *PLoS ONE*. 2013, **8**:e62815.

21. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD: **Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform**. *Applied and environmental microbiology*. 2013, **79**:5112–5120.

22. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, others: **Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities**. *Applied and environmental microbiology*. 2009, **75**:7537–7541.

23. Wang Q, Garrity GM, Tiedje JM, Cole JR: **Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy**. *Applied and environmental microbiology*. 2007, **73**:5261–5267.

24. Calle ML, Urrea V, Boulesteix A-L, Malats N: **AUC-RF: A new strategy for genomic profiling with random forest**. *Human heredity*. 2011, **72**:121–132.

25. DeLong ER, DeLong DM, Clarke-Pearson DL: **Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach**. *Biometrics*. 1988:837–845.

26. Youden WJ: **Index for rating diagnostic tests**. *Cancer*. 1950, **3**:32–35.

27. Pepe M, Longton G, Janes H: **Estimation and comparison of receiver operating characteristic curves**. *The Stata Journal*. 2009, **9**:1.

28. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M: **PROC: An open-source package for r and s+ to analyze and compare rOC curves**. *BMC bioinformatics*. 2011, **12**:1.

29. Dixon P, Palmer M: **VEGAN, a package of r functions for community ecology**. *Journal of Vegetation Science*. 2003, **14**:927–930.

30. Warren RL, Freeman DJ, Pleasance S, Watson P, Moore RA, Cochrane K, Allen-Vercoe E, Holt RA: **Co-occurrence of anaerobic bacteria in colorectal carcinomas**. *Microbiome*. 2013, **1**:16.

31. Yu J, Feng Q, Wong SH, Zhang D, Liang Q yi, Qin Y, Tang L, Zhao H, Stenvang J, Li Y, others: **Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer**. *Gut*. 2015:gutjnl–2015.

32. Rex DK, Johnson DA, Anderson JC, Schoenfeld PS, Burke CA, Inadomi JM: **American College of Gastroenterology guidelines for colorectal cancer screening 2008**. *The American journal of gastroenterology*. 2009, **104**:739–750.

33. Heitman SJ, Ronksley PE, Hilsden RJ, Manns BJ, Rostom A, Hemmelgarn BR: **Prevalence of adenomas and colorectal cancer in average risk individuals: a**

systematic review and meta-analysis. *Clinical Gastroenterology and Hepatology*. 2009, **7**:1272–1278.

34. Symonds EL, Osborne JM, Cole SR, Bampton PA, Fraser RJ, Young GP: **Factors affecting faecal immunochemical test positive rates: demographic, pathological, behavioural and environmental variables**. *Journal of Medical Screening*. 2015:0969141315584783.

35. Kapidzic A, Meulen MP van der, Hol L, Roon AH van, Looman CW, Lansdorp-Vogelaar I, Ballegooijen M van, Vuuren AJ van, Reijerink JC, Leerdam ME van, others: **Gender differences in fecal immunochemical test performance for early detection of colorectal neoplasia**. *Clinical Gastroenterology and Hepatology*. 2015.

36. Levi Z, Rozen P, Hazazi R, Vilkin A, Waked A, Maoz E, Birkenfeld S, Lieberman N, Klang S, Niv Y: **Sensitivity, but not specificity, of a quantitative immunochemical fecal occult blood test for neoplasia is slightly increased by the use of low-dose aspirin, NSAIDs, and anticoagulants**. *The American journal of gastroenterology*. 2009, **104**:933–938.

37. Schubert AM, Sinani H, Schloss PD: **Antibiotic-Induced Alterations of the Murine Gut Microbiota and Subsequent Effects on Colonization Resistance against Clostridium difficile**. *MBio*. 2015, **6**.

38. Pryde SE, Duncan SH, Hold GL, Stewart CS, Flint HJ: **The microbiology of butyrate formation in the human colon**. *FEMS microbiology letters*. 2002, **217**:133–139.

39. Segain J, De La Blétiere DR, Bourreille A, Leray V, Gervois N, Rosales C, Ferrier L, Bonnet C, Blottiere H, Galmiche J: **Butyrate inhibits inflammatory responses through nFκB inhibition: Implications for crohn's disease**. *Gut*. 2000, **47**:397–403.

40. D'Argenio G, Cosenza V, Delle Cave M, Iovino P, Delle Valle N, Lombardi G, Mazzacca G: **Butyrate enemas in experimental colitis and protection against large bowel cancer in a rat model**. *Gastroenterology*. 1996, **110**:1727–1734.

41. Hague A, Elder DJ, Hicks DJ, Paraskeva C: **Apoptosis in colorectal tumour cells: Induction by the short chain fatty acids butyrate, propionate and acetate and by the bile salt deoxycholate**. *International Journal of Cancer*. 1995, **60**:400–406.

42. Gallimidi AB, Fischman S, Revach B, Bulvik R, Maliutina A, Rubinstein AM, Nussbaum G, Elkin M: **Periodontal pathogens porphyromonas gingivalis and fusobacterium nucleatum promote tumor progression in an oral-specific chemical carcinogenesis model.** *Oncotarget*. 2015.

43. Ding T, Schloss PD: **Dynamics and associations of microbial community types across the human body**. *Nature*. 2014, **509**:357–360.

CHAPTER 4:

DNA from fecal immunochemical test can replace stool for microbiota-based colorectal cancer screening

The contents of this chapter have been submitted for publication as:

> Baxter NT, Koumpouras CC, Ruffin MT IV, Rogers MAM, Schloss PD. (2016). DNA from fecal immunochemical test can replace stool for microbiota-based colorectal cancer screening. *BMC Cancer* (under review).

Abstract

There is a significant demand for colorectal cancer (CRC) screening methods that are noninvasive, inexpensive, and capable of accurately detecting early stage tumors. It has been shown that models based on the gut microbiota can complement the fecal occult blood test and fecal immunochemical test (FIT). However, a barrier to microbiota-based screening is the need to collect and store a patient's stool sample. Using stool samples collected from 404 patients we tested whether the residual buffer containing resuspended feces in FIT cartridges could be used in place of intact stool samples. We found that the bacterial DNA isolated from FIT cartridges largely recapitulated the community structure and membership of patients' stool microbiota and that the abundance of bacteria associated with CRC were conserved. We also found that

models for detecting CRC that were generated using bacterial abundances from FIT cartridges were equally predictive as models generated using bacterial abundances from stool. These findings demonstrate the potential for using residual buffer from FIT cartridges in place of stool for microbiota-based screening for CRC. This may reduce the need to collect and process separate stool samples and may facilitate combining FIT and microbiota-based biomarkers into a single test. Additionally, FIT cartridges could constitute a novel data source for studying the role of the microbiome in cancer and other diseases.

Background

Although colorectal cancer (CRC) mortality has declined in recent decades, it remains the second leading cause of death among cancers in the United States [1]. Early detection of CRC is critical since patients whose tumors are detected at an early stage have a greater than 90% chance of survival [1]. However more than a third of individuals for whom screening is recommended do not adhere to screening guidelines [2]. The high cost and invasive nature of procedures, such as colonoscopy and sigmoidoscopy are barriers for many people [3, 4]. Unfortunately non-invasive tests, such as the guaiac fecal occult blood test (gFOBT), fecal immunochemical test (FIT), and the multitarget DNA test fail to reliably detect adenomas [5, 6] (e.g., sensitivity for nonadvanced adenomas is 7.6% for FIT and 17.2% for the DNA test). Thus, there is a need for novel non-invasive screening methods with improved sensitivity for early stage colonic lesions.

Several studies have demonstrated the potential for the gut microbiota to be used to detect CRC [7–10]. Moreover, we and others have shown that combining microbiota-

analysis with conventional diagnostics, like gFOBT and FIT, can significantly improve the detection of colonic lesions over either method by itself [7, 8, 10]. One limitation of microbiota-based CRC screening is the need to collect and process separate stool samples for microbiota characterization. Given the widespread use of FIT to collect specimens for screening, the ability to use the same sample for microbiota characterization could make processing more efficient and less expensive. We hypothesized that the small amount of fecal material contained in FIT sampling cartridges was sufficient to perform both hemoglobin quantification and microbiota characterization. To test this hypothesis, we isolated bacterial DNA from the residual buffer of OC-Auto® FIT cartridges (Polymedco Inc.) that had already been used for quantifying fecal hemoglobin concentrations. We then compared the bacterial composition of the FIT cartridge to that of DNA isolated directly from a patient's stool sample and assessed the ability of FIT cartridge-derived DNA to be used for microbiota-based CRC screening.

Methods

Study Design / Diagnoses / Stool Collection. Stool samples were obtained through the Great Lakes-New England Early Detection Research Network. Patients were asymptomatic, at least 18 years old, willing to sign informed consent, able to tolerate removal of 58 mL of blood, and willing to collect a stool sample. Patient age at the time of enrollment ranged from 29 to 89 with a median of 60 years. Patients were excluded if they had undergone surgery, radiation, or chemotherapy for current CRC prior to baseline samples or had inflammatory bowel disease, known hereditary non-polyposis CRC, or familial adenomatous polyposis. Patient diagnoses were determined by

colonoscopic examination and histopathological review of any biopsies taken. Colonoscopies were performed and fecal samples were collected in four locations: Toronto (Ontario, Canada), Boston (Massachusetts, USA), Houston (Texas, USA), and Ann Arbor (Michigan, USA). Stool samples were packed in ice, shipped to a processing center via next day delivery and stored at -80˚C. Fecal material for FIT was collected from frozen stool aliquots using OC-Auto® FIT sampling bottles (Polymedco Inc.), processed using an OC-Auto Micro 80 automated system (Polymedco Inc.), and stored at -20C. The University of Michigan Institutional Review Board approved this study, and all subjects provided informed consent.

16S rRNA gene sequencing. Processed FIT samples were thawed, and 100 µl of buffer were withdrawn by pipette for DNA extraction. DNA was isolated from FIT samples or matching stool samples using the PowerSoil-htp 96 Well Soil DNA isolation kit (MO BIO Laboratories) and an epMotion 5075 automated pipetting system (Eppendorf). The V4 region of the bacterial 16S rRNA gene was amplified using custom barcoded primers and sequenced as described previously using an Illumina MiSeq sequencer [11]. The 16S rRNA gene sequences were curated using the mothur software package, as described previously [11, 12]. Curated sequences were clustered into operational taxonomic units (OTUs) using a 97% similarity cutoff with the average neighbor clustering algorithm. Sequences were classified using a naive Bayesian classifier trained against a 16S rRNA gene training set provided by the Ribosomal Database Project (RDP) [13]. Species-level classifications for OTUs of interest were determined by using blastn to compare the predominant sequence within each OTU to the NCBI 16S rRNA database. The putative species was only reported for OTUs with

greater than 99% sequence identity to a single species in the database; otherwise the consensus RDP classification was used.

Statistical Methods. All statistical analyses were performed using R (v.3.2.0). Random forest models were generated using the AUC-RF algorithm for feature reduction and maximizing model performance [14]. The most predictive OTUs were determined based on mean decrease in accuracy when removed from the model. The area under the curve (AUC) of receiver operator characteristic (ROC) curves were compared using the method described by DeLong et al. [15] as implemented in the pROC R package [16].

Results

DNA was isolated and 16S rRNA gene sequencing was performed on stool aliquots and the residual buffer of paired OC-Auto® FIT sampling cartridges from 404 patients. Among these patients, 101 had CRC, 162 had adenomas, and 141 had no colonic lesions. First, we tested whether the bacterial community profiles from FIT cartridges recapitulated their stool counterparts. First, we compared the number of OTUs shared within FIT/stool pairs from the same patient to the number of OTUs shared between patients (Fig. 4.1A). FIT cartridges and stool from the same patient (red line) had significantly more bacterial populations in common than those taken from different patients ($p<0.001$, two-sample Kolmogorov-Smirnov test), indicating that community membership was conserved within patients across stool and FIT cartridges. Second, we calculated the similarity in community structure between samples using 1-thetaYC index [17]. This metric compares the presence or absence of bacterial populations and their relative abundance. The bacterial community structure of stool and FIT samples from the same patient (red line) were significantly more similar to each other than to stool or

FIT from other patients (Fig. 4.1B, $p<0.001$). Finally, we used a Mantel test to determine whether the patient-to-patient thetaYC distances among stool samples were correlated with the patient-to-patient thetaYC distances among FIT cartridges. We found that there was a significant correlation (Mantel test $r=0.525$, $p<0.001$), suggesting that the inter-patient variation in community structure between the stool samples of patients was conserved in samples from FIT cartridges.

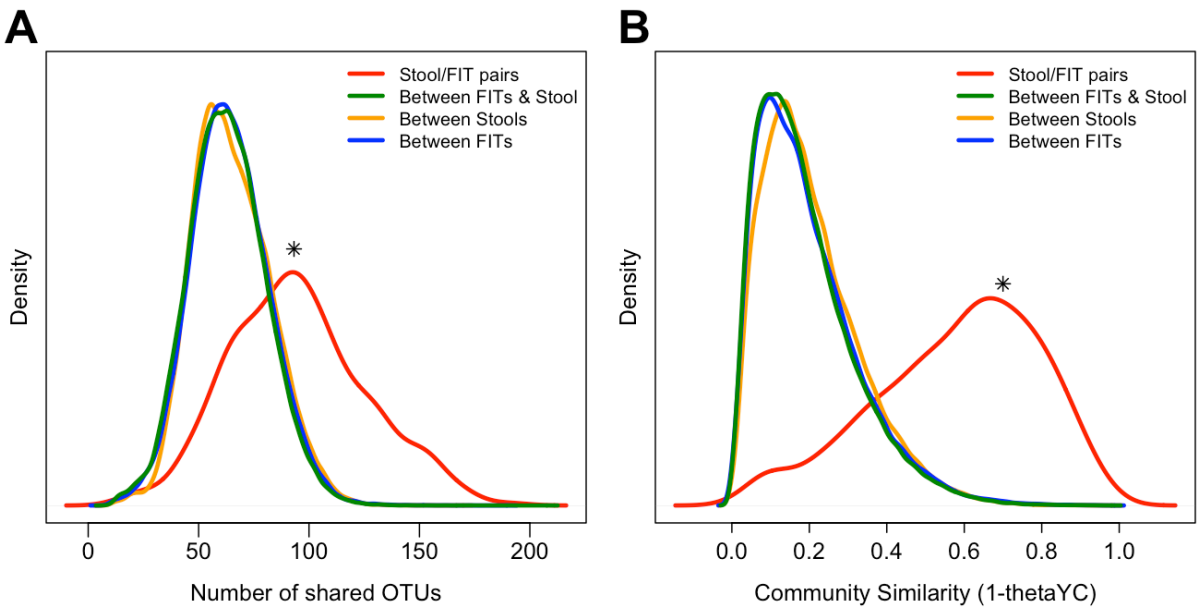

Figure 4.1. Bacterial community structure from FIT cartridge recapitulates stool. Density plots showing distribution of the number of shared OTUs (A) and community similarity (B) between groups of samples (* $p<0.001$ two-sample Kolmogorov-Smirnov Test).

Next, we observed a significant correlation between the abundance of each genus in the paired FIT cartridge and stool samples (Fig. 4.2A, Spearman rho: 0.699, $p<0.001$). This suggested that the abundance of bacterial genera was conserved. This correlation was especially strong when comparing only the 100 most abundant genera from stool (Spearman rho: 0.886, $p<0.001$). Several bacterial species have been repeatedly associated with CRC, including *Fusobacterium nucleatum*, *Porphyromonas asaccharolytica*, *Peptostreptococcus stomatis*, and *Parvimonas micra* [8–10, 18]. As expected, the abundance of these species in stool was significantly correlated with their abundance in matched FIT cartridges (all $p<0.001$, Spearman rho ≥ 0.352)(Fig. 4.2B). We observed some biases in the abundance of certain taxa. In particular, the genus *Pantoea* was detected in 399 of the 404 FIT cartridges with an average abundance of 2.4%, but was only detected in 1 stool sample. The genus *Helicobacter* was detected in 172 FIT cartridges, but only 10 stool samples. Likewise several genera of *Actinobacteria* were more abundant in stool samples compared to FIT. Notwithstanding these few exceptions, the abundance of the vast majority of genera were well conserved between stool and FIT cartridges. Overall, these findings suggested that that the overall bacterial community structure and the abundance of specific taxa in FIT cartridges and stool were similar.

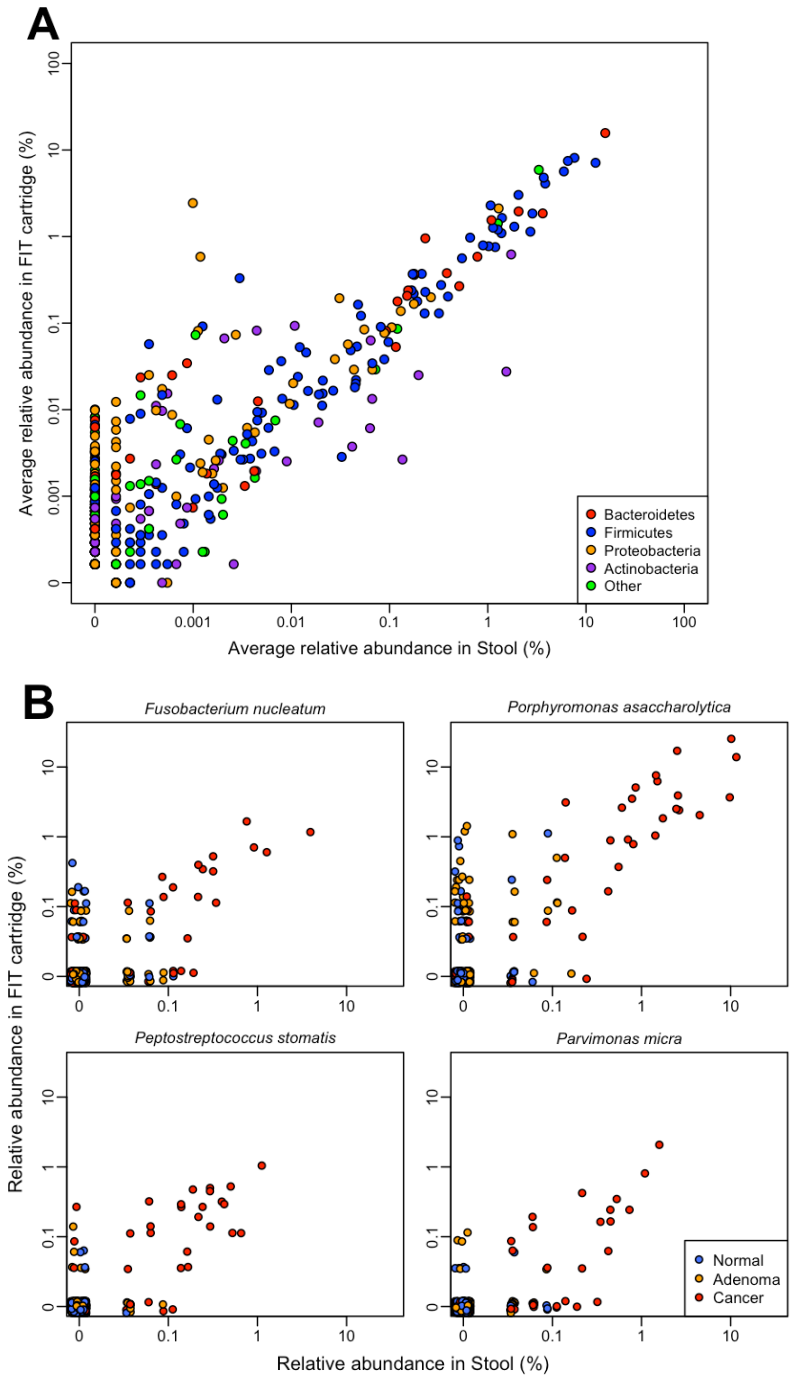

Figure 4.2. Bacterial populations conserved between stool and FIT cartridge. (A) Scatterplot of the average relative abundance of each bacterial genus in stool and FIT cartridges colored by phylum. (B) Scatterplots of the relative abundances of 4 species frequently associated with CRC. All correlations were greater than 0.35 (all $p<0.001$).

We tested whether the bacterial relative abundances we observed from FIT cartridges could be used to differentiate healthy patients from those with carcinomas using random forest models as we did previously using intact stool samples [10]. Using DNA from the FIT cartridge, the optimal model utilized 28 OTUs and had an AUC of 0.831 (Fig. 4.3A). There was not a significant difference in the AUC for this model and the model based on DNA isolated directly from stool, which used 32 OTUs and had an AUC of 0.853 ($p=0.41$). Furthermore, the probabilities of individuals having lesions was correlated between the models generated using DNA isolated from the FIT cartridges and stool samples (Spearman rho: 0.633, $p<0.001$, Fig. 4.3B). We also generated random forest models for differentiating healthy patients from those with any type of lesions (i.e. adenoma or carcinoma). There was not a significant difference in AUC between the stool-based model with 41 OTUs (AUC=0.700) and the FIT cartridge-based model with 41 OTUs (AUC=0.686, $p=0.65$, Fig. 4.3C). Again, the probabilities of individuals having lesions according to the two models were significantly correlated (Spearman rho: 0.389, $p<0.001$ Fig. 4.3D). These findings demonstrated that models based on bacterial DNA from FIT cartridges were as predictive as models based on DNA isolated directly from stool.

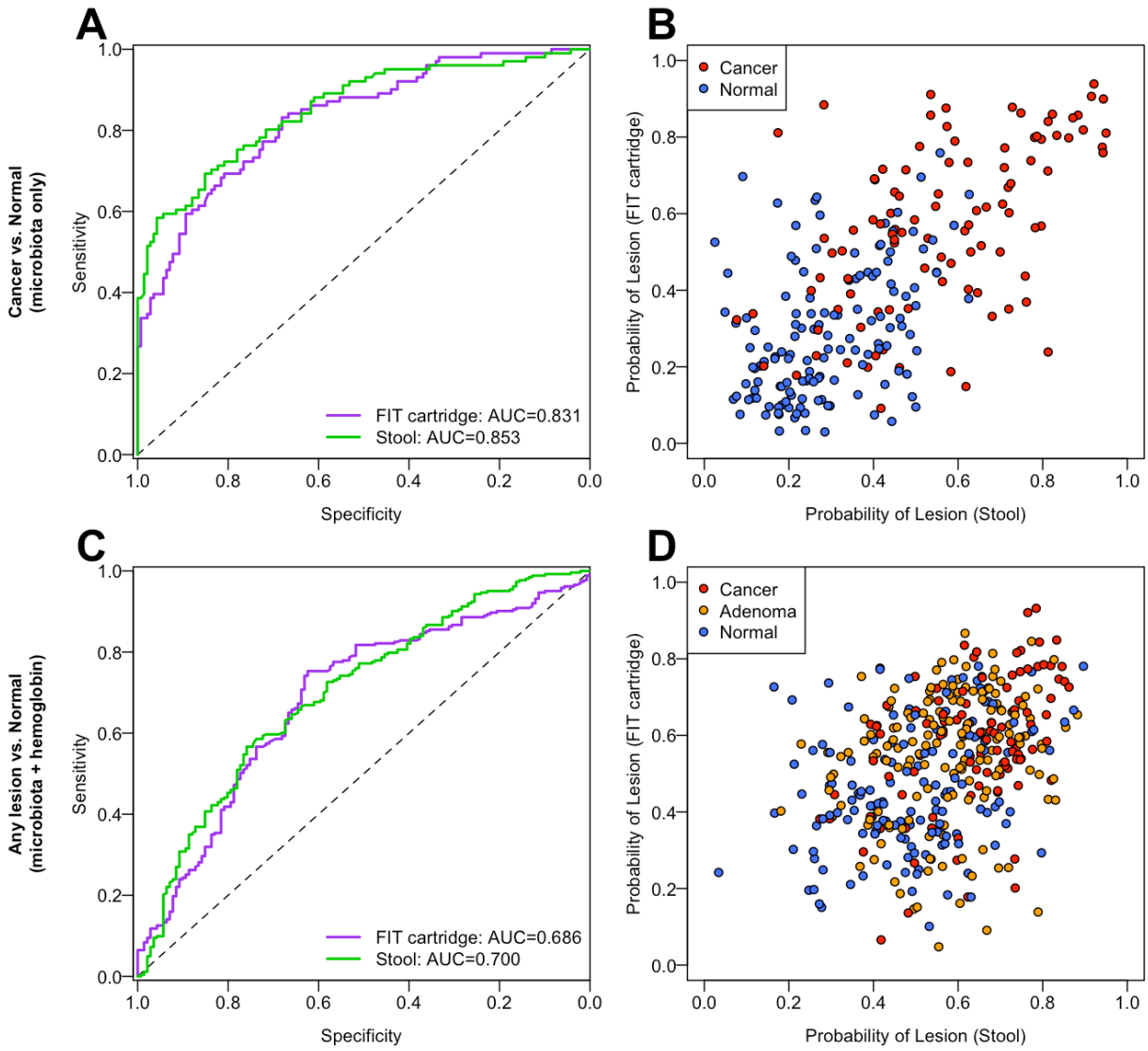

Figure 4.3. Microbiota-based models from FIT cartridge DNA are as predictive as models from stool. (A) ROC curves for distinguishing healthy patients from those with cancer using using microbiota-based random forest models using DNA from FIT cartridges or stool. (B) Probability of having cancer for each patient according to microbiota-based models from A. (C) ROC curves for distinguishing patients with adenomas or carcinomas from healthy patients using microbiota-based random forest models using DNA from FIT cartridges or stool. (D) Probability of having a lesion for each patient based on the models from C.

Discussion

Bacterial DNA isolated from the residual buffer of FIT cartridges recapitulated the community structure and membership of patients' stool microbiota. FIT/stool pairs collected from the same patient were significantly more similar to each other than samples from different patients and the inter-patient differences in stool microbiota structure were conserved in FIT cartridge-derived microbiota. More importantly, random forest models generated using bacterial abundances from FIT cartridge-derived and stool-derived DNA were equally predictive for differentiating healthy patients from those with adenomas and carcinomas.

Sinha et al. compared a variety of sampling and storage methods for fecal samples to be used for microbiome analyses [19]. They found reproducible biases according to sampling method and time at ambient temperature. Likewise, we observed biases in the abundance certain bacterial populations in FIT cartridges compared stool. For example, an OTU associated with *Pantoea* was found in 98.8% of FIT cartridge samples and only 0.2% of stool samples. There are several possible explanations for this result. It is possible that because the biomass contained in the FIT cartridges is considerably lower than that in stool, the analysis was more sensitive to contaminants in our reagents or the FIT cartridge [20]. Alternatively, storage conditions could have played a role in biasing the relative abundances of certain genera. The feces in the FIT cartridges spent more time exposed to ambient temperatures in order to be analyzed for hemoglobin concentration. Therefore it is possible that certain bacterial populations, especially aerobes, were able to grow. Considering *Pantoea* is rarely found in human feces and is more commonly found in soil, plant surfaces, and air we suspect that it was a

contaminant. Regardless of the source of this and the other suspicious populations, any biases were limited since the random forest feature selection process did not select these populations and did not affect the ability to detect CRC from FIT cartridge-derived DNA.

Conclusions

This could reduce the need to collect and process separate stool samples, decreasing the cost of screening. It may be possible to use FIT cartridges rather than separate stool samples for future studies on the role of the gut microbiota and cancer. Samples collected from patients who undergo annual FIT screening could be used to monitor temporal changes in a patient's microbiota, making it possible to detect shifts toward a disease-associated microbiota. Since FIT cartridges are currently used for CRC screening, our findings may facilitate large-scale validations of microbiota-based screening methods.

References

1. Siegel R, DeSantis C, Jemal A: **Colorectal cancer statistics, 2014**. *CA: a cancer journal for clinicians*. 2014, **64**:104–117.

2. Centers for Disease Control and Prevention: **Vital signs: Colorectal cancer screening test use–United states, 2012.** *MMWR Morbidity and mortality weekly report*. 2013, **62**:881.

3. Hsia J, Kemper E, Kiefe C, Zapka J, Sofaer S, Pettinger M, Bowen D, Limacher M, Lillington L, Mason E, others: **The importance of health insurance as a determinant of cancer screening: evidence from the Women's Health Initiative**. *Preventive medicine*. 2000, **31**:261–270.

4. Jones RM, Devers KJ, Kuzel AJ, Woolf SH: **Patient-reported barriers to colorectal cancer screening: a mixed-methods analysis**. *American journal of preventive medicine*. 2010, **38**:508–516.

5. Hundt S, Haug U, Brenner H: **Comparative evaluation of immunochemical fecal occult blood tests for colorectal adenoma detection**. *Annals of Internal Medicine*. 2009, **150**:162–169.

6. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, Ahlquist DA, Berger BM: **Multitarget stool DNA testing for colorectal-cancer screening**. *New England Journal of Medicine*. 2014, **370**:1287–1297.

7. Zackular JP, Rogers MA, Ruffin MT, Schloss PD: **The human gut microbiome as a screening tool for colorectal cancer**. *Cancer Prevention Research*. 2014, **7**:1112–1121.

8. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Böhm J, Brunetti F, Habermann N, others: **Potential of fecal microbiota for early-stage detection of colorectal cancer**. *Molecular systems biology*. 2014, **10**:766.

9. Yu J, Feng Q, Wong SH, Zhang D, Liang Q yi, Qin Y, Tang L, Zhao H, Stenvang J, Li Y, others: **Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer**. *Gut*. 2015:gutjnl–2015.

10. Baxter NT, Ruffin MT, Rogers MA, Schloss PD: **Microbiota-based model improves the sensitivity for detecting colonic lesions**. *Genome Medicine*. 2016, **8**:1–10.

11. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD: **Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform**. *Applied and environmental microbiology*. 2013, **79**:5112–5120.

12. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, others: **Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities**. *Applied and environmental microbiology*. 2009, **75**:7537–7541.

13. Wang Q, Garrity GM, Tiedje JM, Cole JR: **Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy**. *Applied and environmental microbiology*. 2007, **73**:5261–5267.

14. Calle ML, Urrea V, Boulesteix A-L, Malats N: **AUC-RF: A new strategy for genomic profiling with random forest**. *Human heredity*. 2011, **72**:121–132.

15. DeLong ER, DeLong DM, Clarke-Pearson DL: **Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach**. *Biometrics*. 1988:837–845.

16. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M: **pROC: an open-source package for R and S+ to analyze and compare ROC curves**. *BMC bioinformatics*. 2011, **12**:1.

17. Yue JC, Clayton MK: **A similarity measure based on species proportions**. *Communications in Statistics-Theory and Methods*. 2005, **34**:2123–2131.

18. Warren RL, Freeman DJ, Pleasance S, Watson P, Moore RA, Cochrane K, Allen-Vercoe E, Holt RA: **Co-occurrence of anaerobic bacteria in colorectal carcinomas**. *Microbiome*. 2013, **1**:16.

19. Sinha R, Chen J, Amir A, Vogtmann E, Inman KS, Flores-Munguia R, Sampson JN, Knight R, Chia N, others: **Collecting fecal samples for microbiome analyses in epidemiology studies**. *Cancer Epidemiology Biomarkers & Prevention*. 2015:cebp–0951.

20. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW: **Reagent and laboratory contamination can critically impact sequence-based microbiome analyses**. *BMC biology*. 2014, **12**:87.

CHAPTER 5: Discussion

The preceding chapters demonstrate the potential for using the gut microbiota for detecting CRC. In them we showed that the composition of the gut microbiota can potentially influence an individual's susceptibility to CRC, that shifts in microbiota composition can be used to predict the presence of lesions, and that microbiota-based screening can be combined with fecal immunochemical test into a single test with improved sensitivity for colonic lesions. This chapter includes a summary of the findings from the preceding chapters, discussion on the implications of those findings, and the next steps for microbiota-based CRC screening.

Summary and implications of Chapter 2

In Chapter 2, we used a germ-free mouse model to test the influence of microbiota structure on colon tumorigenesis. We initially hypothesized that mice inoculated with human cancer-associated microbiota would develop more tumors than mice who received microbiota from healthy human donors. There was evidence supporting this hypothesis. We had already shown that germ-free mice that received cancer-associated microbiota from other mice developed more tumors than mice inoculated with healthy mouse microbiota [1]. Other groups had reportedly transferred other disease phenotypes from humans to mice through microbiota transfer [2–4]. However, the same did not hold true in our experiment, as there was no difference in tumor burden between

mice who received a healthy or cancer-associated microbiota. There were several potential explanations for the lack of phenotype transfer. Chief among them is the inability to fully recapitulate the microbiota structure of the human donor. In an extreme example, one human donor's microbiota was composed of 70% Fusobacteria, however the mice who received that microbiota had no detectable Fusobacteria in their stool. Even if there had been completely transfer of the donor's microbiota structure, there was no guarantee that the species associated with CRC in humans would be able to accelerate tumorigenesis in mice.

Nonetheless, the study provided valuable insights into the microbiota's role in CRC and potential use as screening tool. The most striking observation from this study was that the amount of tumors mice developed was strongly associated with the structure of the microbiota prior to undergoing the chemically induced model. This was a new finding, as previously we had only associated the endpoint microbiota with the severity of disease [1]. This suggested that the structure of the microbiota may influence an individual's susceptibility to tumorigenesis, which opens the possibility of using the microbiota, not only to detect CRC, but to predict an individual's risk of developing tumors.

The predicted metagenomes from this study showed a negative correlation between butyrate producing genes and the number of tumors mice developed. This supports epidemiological data in which a high fiber diet and high butyrate levels are associated with lower risk of CRC [5, 6] and studies showing that butyrate protects against colon tumorigenesis [7–10]. On the other hand, genes potentially involved in mucin degradation were positively correlated with the number of tumors mice developed. These findings are consistent with a model in which the balance of pro- and anti-

inflammatory activities of the microbiota influence and individual's risk of developing CRC.

Summary and implications of Chapter 3

In Chapter 3, we confirmed previous studies, showing that the microbiota can be used to differentiate individual's with CRC from those with healthy colons. We expanded upon those findings by showing that the microbiota could be used to detect a subset of lesions that was distinct from those detected by FIT. Then, we combined microbial biomarkers and FIT into a single model that had improved sensitivity for colonic lesions, especially at the earliest stages of tumor development, where current noninvasive tests are least effective.

These findings demonstrated the advantage of combining multiple biomarkers into a single test. This concept is part of what makes the MT-sDNA test a relatively effective screening tool. Our MMT model and the MT-sDNA test have remarkably similar specificity and sensitivities for each type of lesion [11]. It is likely that combining the host-associated markers in the MT-sDNA with the microbial markers in the MMT would further improve screening accuracy. It is also likely that incorporating patient characteristics into the model would also improve their accuracy. That was true for our models in an earlier study [12], but not for the MMT model.

Summary and implications of Chapter 4

In Chapter 4, we took the findings from Chapter 3 one step further. Not only could the results from FIT and the microbiota analysis be combined into a single model, but the two tests could be physically linked by using the same sampling cartridge for both tests. Although the results of this study were not surprising, they could have a profound

impact on the way we study the microbiota and colorectal cancer. Thousands of FIT cartridges are analyzed in the U.S. every day [13]. If those could be repurposed for microbiota research, it could drastically increase the number of samples available for studying the microbiota's role in CRC and many other diseases. Furthermore, FIT cartridges are typically performed annually, which means an individual's microbiota can be monitored over time. This could allow for large-scale longitudinal study of the microbiota. It would even be possible to study changes in the microbiota that occur leading up to the development of tumors.

The next step for CRC screening

Monitoring changes to an individual's microbiota over time may be the most effective way to detect disease-associated alterations in the microbiota. The most predictive species for adenomas were not the potentially pathogenic species typically associated with carcinomas. Instead, patients with adenomas were identified by a lack of potentially beneficial organisms, especially the often butyrogenic Lachnospiraceae family. Based on the findings from Chapter 2 and other studies, a decrease in butyrate producers could lead to an increase in susceptibility to tumorigenesis [7, 8, 10]. However, many different bacterial populations were needed even to be weakly predictive for detecting adenomas. One potential explanation is that the loss of beneficial organisms is highly individual-specific. An underlying theme of human microbiota research is the high level of inter-individual variability. If no two individuals have the exactly the same "healthy" microbiota, then we cannot expect every individual to have the same deviation from that healthy state.

With that in mind, the best approach to microbiota-based screening may be to monitor a patient's microbiota over a long period of time, making it easier to detect individual-specific shifts in the structure of their microbiota. These idiosyncratic changes may be the best way to detect subtle, but important shifts that might be indicative of early tumor development. Only later in tumor development are there consistent blooms in potential pathogens in most individuals. Like many other fields of medicine, microbiota-related diseases may require a highly personalized approach to both screening and treatment.

Potential mechanism for microbiota-mediated tumorigenesis

Like several other groups, we found that bacteria typically associated with oral cavity were among the most enriched in the stool of patients with carcinomas [14–16]. *F. nucleatum* has received the most attention due to its ability to potentiate tumors in a mouse model, however *P. asaccharolytica* was even more enriched in patients with CRC in our study. Interestingly another *Porphyromonas* species, *P. gingivalis* is capable of synergizing with *F. nucleatum* to promote oral cancer [17]. It is possible that a similar mechanism can happen in the gut. Even if no causal link exists, *P. asaccharolytica* shows potential as an effective biomarker for CRC. Likewise *Parvimonas micra* and *Peptostreptoccus stomatis* have been repeatedly associated with carcinomas as well [14, 15, 18].

We propose the following as one potential mechanism for the microbiota's role in CRC. First there are perturbations to an individual's gut microbiota, involving the loss of mutualistic species, including a decrease in butyrate production and/or other beneficial metabolic activity. These changes could predispose the colon to adenoma formation. The altered environment of the adenomatous tissue could allow for *F. nucleatum* to

colonize the distal gastrointestinal tract, binding epithelial cells via its adhesion FadA. Then, as it does in the periodontal cavity, *F. nucleatum* can mediate the binding of other oral pathogens, forming a multispecies biofilm [19, 20]. These oral pathogens could promote inflammation and the secretion of peptides to sustain their asaccharolytic metabolism, as they do in periodontitis [21]. This would lead to cycle in which the oral pathogens promote and benefit from an inflamed microenvironment that accelerates the progression of tumorigenesis. This mechanisms, though highly speculative, is consistent with the changes in microbiota structure that occur over the course of tumor development and the known pro-inflammatory and tumorigeneic activity of these species in the mouth [22].

References

1. Zackular JP, Baxter NT, Iverson KD, Sadler WD, Petrosino JF, Chen GY, Schloss PD: **The gut microbiome modulates colon tumorigenesis**. *MBio*. 2013, **4**:e00692–13.

2. Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V, Henrissat B, Bain JR, others: **Gut microbiota from twins discordant for obesity modulate metabolism in mice**. *Science*. 2013, **341**:1241214.

3. Koren O, Goodrich JK, Cullender TC, Spor A, Laitinen K, Bäckhed HK, Gonzalez A, Werner JJ, Angenent LT, Knight R, others: **Host remodeling of the gut microbiome and metabolic changes during pregnancy**. *Cell*. 2012, **150**:470–480.

4. Smith MI, Yatsunenko T, Manary MJ, Trehan I, Mkakosya R, Cheng J, Kau AL, Rich SS, Concannon P, Mychaleckyj JC, others: **Gut microbiomes of malawian twin pairs discordant for kwashiorkor**. *Science*. 2013, **339**:548–554.

5. Chen H-M, Yu Y-N, Wang J-L, Lin Y-W, Kong X, Yang C-Q, Yang L, Liu Z-J, Yuan Y-Z, Liu F, Wu J-X, Zhong L, Fang D-C, Zou W, Fang J-Y: **Decreased dietary fiber intake and structural alteration of gut microbiota in patients with advanced colorectal adenoma.** *Am J Clin Nutr*. 2013, **97**:1044–1052.

6. Huxley RR, Ansary-Moghaddam A, Clifton P, Czernichow S, Parr CL, Woodward M: **The impact of dietary and lifestyle risk factors on risk of colorectal cancer: A quantitative overview of the epidemiological evidence.** *Int J Cancer*. 2009, **125**:171–180.

7. D'Argenio G, Cosenza V, Delle Cave M, Iovino P, Delle Valle N, Lombardi G, Mazzacca G: **Butyrate enemas in experimental colitis and protection against large bowel cancer in a rat model**. *Gastroenterology*. 1996, **110**:1727–1734.

8. Hu Y, Le Leu RK, Christophersen CT, Somashekar R, Conlon MA, Meng XQ, Winter JM, Woodman RJ, McKinnon R, Young GP: **Manipulation of the gut microbiota using resistant starch is associated with protection against colitis-associated colorectal cancer in rats**. *Carcinogenesis*. 2016:bgw019.

9. Hague A, Elder DJ, Hicks DJ, Paraskeva C: **Apoptosis in colorectal tumour cells: Induction by the short chain fatty acids butyrate, propionate and acetate and by the bile salt deoxycholate**. *International Journal of Cancer*. 1995, **60**:400–406.

10. Segain J, De La Blétiere DR, Bourreille A, Leray V, Gervois N, Rosales C, Ferrier L, Bonnet C, Blottiere H, Galmiche J: **Butyrate inhibits inflammatory responses through nFκB inhibition: Implications for crohn's disease**. *Gut*. 2000, **47**:397–403.

11. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, Ahlquist DA, Berger BM: **Multitarget stool DNA testing for colorectal-cancer screening**. *New England Journal of Medicine*. 2014, **370**:1287–1297.

12. Zackular JP, Rogers MA, Ruffin MT, Schloss PD: **The human gut microbiome as a screening tool for colorectal cancer**. *Cancer Prevention Research*. 2014, **7**:1112–1121.

13. Centers for Disease Control and Prevention: **Vital signs: colorectal cancer screening test use–United States, 2012.** *MMWR Morbidity and mortality weekly report*. 2013, **62**:881.

14. Warren RL, Freeman DJ, Pleasance S, Watson P, Moore RA, Cochrane K, Allen-Vercoe E, Holt RA: **Co-occurrence of anaerobic bacteria in colorectal carcinomas**. *Microbiome*. 2013, **1**:16.

15. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Böhm J, Brunetti F, Habermann N, others: **Potential of fecal microbiota for early-stage detection of colorectal cancer**. *Molecular systems biology*. 2014, **10**:766.

16. Baxter NT, Ruffin MT, Rogers MA, Schloss PD: **Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions**. *Genome Medicine*. 2016, **8**:1.

17. Fischman S, Revach B, Bulvik R, Maliutina A, Rubinstein A, Nussbaum G, Elkin M, others: **Periodontal pathogens porphyromonas gingivalis and fusobacterium nucleatum promote tumor progression in an oral-specific chemical carcinogenesis model.** *Oncotarget*. 2015, **6**:22613–22623.

18. Yu J, Feng Q, Wong SH, Zhang D, Liang Q yi, Qin Y, Tang L, Zhao H, Stenvang J, Li Y, others: **Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer**. *Gut*. 2015:gutjnl–2015.

19. Kuboniwa M, Lamont RJ: **Subgingival biofilm formation**. *Periodontology 2000*. 2010, **52**:38–52.

20. Kolenbrander PE, Palmer RJ, Periasamy S, Jakubovics NS: **Oral multispecies biofilm development and the key role of cell–cell distance**. *Nature Reviews Microbiology*. 2010, **8**:471–480.

21. Takahashi N: **Microbial ecosystem in the oral cavity: Metabolic diversity in an ecological niche and its relationship with oral diseases**. In *International congress series*. *Volume 1284*. Elsevier; 2005:103–112.

22. Flynn KJ, Baxter NT, Schloss PD: **Oral bacteria in colorectal cancer: Metabolic and community synergy**. *mSphere*. 2016:1–1.