Microbiota analysis complements fecal immunochemical test for detection of colonic lesions

Nielson T. Baxter1, Mack T. Ruffin IV2, Mary A.M. Rogers3, and Patrick D. Schloss1\*

1Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan.  
2Department of Family Medicine, University of Michigan, Ann Arbor, Michigan.  
3Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan.  
\*Correspondence to: [pschloss@umich.edu](mailto:pschloss@umich.edu)

**Colorectal cancer is the second leading cause of death among cancers in the United States1. 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 sigmoidocsopy, are expensive and invasive1–4. Thus, there is a great need to improve the sensitivity of non-invasive tests to detect early stage cancers and adenomas. Numerous studies have demonstrated a causal link between the formation of colonic lesions and the activity of the gut microbiota in tissue culture and animal models5–8. These findings have been complemented by studies in human populations identifying shifts in the composition of the gut microbiota associated with the progression of colorectal cancer9–12. These results suggest 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). Using stool samples from 490 patients we developed a cross-validated 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 95% of cancers and 57.1% of adenomas while FIT alone only detected 75% and 15.7%, respectively. Of the colonic lesions missed by FIT, the model detected 80% of cancers and 49.1% of adenomas. These findings demonstrate the potential for microbiota analysis to complement existing screening methods to improve detection of colonic lesions. With a negative predictive value of 99.98%, our model could be used to accurately identify those patients for whom a colonoscopy is unnecessary, potentially reducing healthcare costs and complications due to invasive screening.**

CRC incidence and mortality have steadily declined in recent decades, due in large part to increased screening1. Further progress is possible by increasing access to and accuracy of diagnostic tests. Although structural exams including colonoscopy and sigmoidoscopy are able to detect and remove adenomas and carcinomas, the high cost and invasive nature are barriers for many people. For example, fear, discomfort, and embarrassment are among the most cited reasons patients choose not to undergo CRC screening4. Likewise the large disparity in screening rates between those with and without health insurance highlights the need for inexpensive screening methods1–3. Unfortunately cheaper, less invasive stool-based tests like guaic fecal occult blood test and FIT are unable to reliably detect adenomas13. 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 CRC9–12. Experiments in animal models have demonstrated that such alterations have the potential to accelerate tumorigenesis5. Furthermore several members of the gut microbiota have been shown to potentiate both the development and progression of CRC by a variety of mechanisms6–8. Although each of these organisms may play a role in certain cases of CRC, none of them is present in every case. Therefore no one organism is an effective biomarker on its own.

We have perviously shown that statistical models that take into account the abundances of multiple bacteria species can be used to distinguish healthy individuals from those with colonic lesions14. In the present study we expanded upon those findings by demonstrating the potential for microbiota analysis to complement FIT for improved detection of colonic lesions. We also improved upon previous studies by utilizing random forest (RF), a decision tree-based machine algorithm for classification, which includes an internal cross-validation to prevent overfitting15. By incorporating both FIT and bacterial abundances into a single RF model, we were able to improve the sensitivity for adenomas and most stages of cancer compared to FIT alone.

We characterized the bacterial communities of stool samples from 490 patients using 16S rRNA gene sequencing. Among these patients, 120 had CRC, 198 had colonic adenomas, and 172 had no colonic lesions. We also tested each sample for the concentraion of occult blood using FIT. With these data we developed a RF model that would differentiate normal individuals from those with any type of colonic lesion (i.e. adenoma or carcinoma). We determined the optimal model using the AUC-RF algorithm for maximizing the area under the curve (AUC) of a RF model16. The optimal model combining FIT and the microbiota used 23 bacterial populations, or operational taxonomic units (OTUs) (Extended Data Fig. 1). Of those OTUs, 16 were members of the Firmicutes phylum, including 3 from the Ruminococcaceae family and 10 from the Lachnospiraceae family, the predominant producers of butyrate in the gut17. Three OTUs were associated with the genus *Bacteroides*. The remaining OTUs were associated with *Porphyromonas*, *Parabacteroides*, *Collinsella*, and Enterobacteriaceae. Interestingly the majority of OTUs used in the model, especially the Lachnospiraceae, were enriched in normal patients, suggesting that a loss of beneficial organisms in addition to the emergence of pathogens may be indicative of CRC development.

To determine whether microbiota sequence data could be used to complement FIT, we compared the performance of the RF model to using FIT alone. The AUC for the RF model was significantly higher than FIT for distinguishing adenoma from normal (p=4.710^{-7}) or all lesions from normal (p=6.310^{-7}), but not cancer from normal (p=0.091) (Fig. 1). The ROC curves for the two tests [there's only one test - but three possible outcomes - need to fix] showed that the RF model does not outperform FIT until the specificity drops below approximately 0.9, at which point the sensitivity of the RF model greatly exceeded that of FIT.

Using Youden's J statisitc18 we determined that the optimal cutoff for the RF model's probability lesion was 0.622. 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 RF model to those of FIT using the widely accepted cutoff of 100 ng/ml of hemoglobin. At these cutoffs the RF model detected 95% of cancers and 57.1% of adenomas compared to 75% and 15.7% for FIT (Table 1, Fig. 2A). When adenomas and cancers were pooled together, the RF model detected 71.4% of lesions, while FIT only detected 38.1%. The RF model had significantly improved sensitivity for both advanced and non-advanced adenomas as well as most stages of cancer (Fig. 2B). The increased sensitivity of the RF model was accompanied by a substantial decrease in specificity (83.7%) compared to FIT (97.1%).

To better understand the relationship between the RF model and FIT, we compared the results of the two tests for each sample (Fig. 3). All samples that tested positive by FIT also tested positive in the RF model, indicating that the RF model did not miss any of the lesions that FIT was able to detect. However the RF model was able to detect 80% of cancers and 49.1% of adenomas that FIT had failed to detect, while maintaining a specificity of 86.2% (Extended Data Fig. 3). This result demonstrated that incorporation of data from a subject's microbiota improved the sensitivity without affecting the specificity of the test.

As a final metric of our model's performance, we estimated the positive predictive value (PPV) and negative predictive value (NPV) by extrapolating its performance on an average-risk population using previously published values for CRC prevalence19 (Extended Data Table 2). Based on a prevalence of 0.3% for CRC, the model would have a relatively low PPV of 1.73%, but a high NPV of 99.98%. For advanced adenomas the model would have a PPV of 17.9% and NPV of 97.1% assuming a prevalence of 5.7%. With a prevalence of 17.7% for nonadvanced adenomas, the PPV for the model would be 42.1% and the NPV would be 89.6%. Although the PPV is quite low relative to FIT, its NPV was significantly higher than that of FIT.

Previous studies have identified differences in diagnostic test performance for certain demographic groups or for people taking certain medications20–22. Therefore we tested whether the RF model performance differed between patient populations. The model performed significantly better for females than males (p=0.016; Extended Data Fig. 4).

Questions... \* Does [FIT] differ significantly between M and F? \* Can you differentitate microbiota of M and F?

[remove from here...] We suspected that the difference could be due to having more adenoma samples from males (n=118) than females (n=80). To correct for this we normalized samples such that male and female groups contained equal numbers of adenomas, cancers, and normal samples. We resampled 1000 times and tested for a significant difference in AUC between males and females. With this method 61.8% of the iterations resulted in a p-value less than 0.05, making it unclear whether gender truly affects the performance of the model. [... to here]

[ Race? / Ethnicity? ]

We found no difference in model performance according to age, BMI, NSAID usage, diabetes, smoking, or previous history of polyps.

Similar to our results, it has recently been shown that when FIT was combined with host-associated DNA biomarkers the ability to detect adenomas and carcinomas was significantly improved over FIT alone23. 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 across the biomarkers. The accuracy of our model may be further improved by incorporating additional biomarkers such as the host-associated biomarkers or those targeting specific genes that encode for toxins7,8 [cite ETBF, PKS, and Bork metagenomics papers]. 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 [schubert]. Such models are likely to be useful as microbiota sequencing gains traction as a tool for characterizing health.

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 RF algorithm it was possible to interpret FIT results in the context of the microbiota. The RF model had significantly higher sensitivity for lesions at almost all stages of tumorigenesis. Moreover the model detected the majority of lesions that FIT was unable to detect. The shortcomings of the RF model were its lack of specificity and low PPV. In other systems, such flaws would result in potentially dangerous [potentially dangerous... what?] over treatment. However in the case of CRC, all people in the United States are encouraged to receive regular colonoscopies once they reach the age of 50 (?)[citation]. Therefore the potential value of the RF model is in its high sensitivity and NPV. With an NPV of 99.98%, the model could be used to identify people for whom a colonoscopy is unnecessary. This strategy could result in a decrease in the number of colonoscopies, thereby reducing both the financial costs and potential health risks of more invasive screening methods.

**Methods Summary.** Fecal samples were collected from 490 subjects in 4 locations: Toronto (Ontario, Canada), Boston (Massachusetts, USA), Houston (Texas, USA), and Ann Arbor (Michigan, USA). Patient diagnoses were determined by colonoscopy and subsequent histopathological examination. FIT was performed using OC FIT-CHEK sampling bottles and processed using an OC-Auto Micro 80 automated system (Polymedco Inc.). The V4 region of the bacterial 16S rRNA gene was amplified using custom barcoded primers and sequenced as described previously using an Illumina MiSeq sequencer24. A data analysis pipeline and all necessary scripts to generate this paper are available at github.com/SchlossLab/Baxter\_glne007Modeling\_2015. The sequence data are available in the Sequence Read Archive under Bioproject Accession XXXXXXXXXXXXX.

### Methods (online only)

***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. 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 nonpolyposis 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. Lesions were biopsied and diagnosed as adenoma or cancer based on subsequent histopathological examination. Whole evacuated stool was collected from each patient packed in ice, shipped to a processing center via next day delivery and stored at -80˚C. [ need to borrow language from joe's paper about when the stool was collected wrt the colonoscopy]

***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.). Raw FIT results were used for generating ROC curves and for building RF models. Sensitivities and specificities reported for FIT are based on a cutoff of 100 ng/ml.

***16S rRNA Sequencing.*** DNA was extracted from roughly 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 sequencer24. The 490 samples were divided into three sequencing runs to increase the per sample sequencing depth. Samples were randomly assigned to the sequencing runs to balance the representation of diangosis, sex, ..., within each run.

***Sequence Curation.*** The 16S rRNA gene sequences were curated using the mothur software package, as described previously24. Briefly, paired-end reads were merged into contigs, screened for quality, aligned to SILVA 16S rRNA sequence database, and screened for chimeras. Curated sequences were clustered in to operational taxonomic units (OTUs) using a 97% similarity cutoff. The number of sequences in each sample was rarefied to 10,000 per sample to minimize the effects of uneven sampling.

***Statistical Methods.*** All statistical analyses were performed using the R software package (citation). Random Forest models were generated using the AUCRF package16. The AUC of ROC curves was compared using the method described by DeLong et al.25. The optimal cutoff for the RF model was determined using Youden's J statistic as implemented in the pROC package in R18. The sensitivities of FIT and the RF model were compared using McNemar's chi-squared test.

***Data Availability.*** Raw fastq files and MIMARKS file are available through the NCBI Sequence Read Archive [insert accession number]. A data analysis pipeline and all necessary scripts are available at github.com/SchlossLab/Baxter\_glne007Modeling\_2015.

### Literature cited

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

2. Centers for Disease Control and Prevention. Vital signs: Colorectal cancer screening among adults aged 50-75 years-united states, 2008. *MMWR. Morbidity and mortality weekly report* **59,** 808 (2010).

3. Hsia, J. *et al.* The importance of health insurance as a determinant of cancer screening: Evidence from the women’s health initiative. *Preventive medicine* **31,** 261–270 (2000).

4. Jones, R. M., Devers, K. J., Kuzel, A. J. & Woolf, S. H. Patient-reported barriers to colorectal cancer screening: A mixed-methods analysis. *American journal of preventive medicine* **38,** 508–516 (2010).

5. Zackular, J. P. *et al.* The gut microbiome modulates colon tumorigenesis. *MBio* **4,** e00692–13 (2013).

6. Kostic, A. D. *et al.* Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell host & microbe* **14,** 207–215 (2013).

7. Wu, S. *et al.* A human colonic commensal promotes colon tumorigenesis via activation of t helper type 17 t cell responses. *Nature medicine* **15,** 1016–1022 (2009).

8. Arthur, J. C. *et al.* Intestinal inflammation targets cancer-inducing activity of the microbiota. *science* **338,** 120–123 (2012).

9. Wang, T. *et al.* Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. *The ISME journal* **6,** 320–329 (2012).

10. Chen, H.-M. *et al.* Decreased dietary fiber intake and structural alteration of gut microbiota in patients with advanced colorectal adenoma. *The American journal of clinical nutrition* **97,** 1044–1052 (2013).

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

12. Shen, X. J. *et al.* Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas. *Gut microbes* **1,** 138–147 (2010).

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

14. Zackular, J. P., Rogers, M. A., Ruffin, M. T. & Schloss, P. D. The human gut microbiome as a screening tool for colorectal cancer. *Cancer Prevention Research* **7,** 1112–1121 (2014).

15. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R news* **2,** 18–22 (2002).

16. Calle, M. L., Urrea, V., Boulesteix, A.-L. & Malats, N. AUC-rF: A new strategy for genomic profiling with random forest. *Human heredity* **72,** 121–132 (2011).

17. Pryde, S. E., Duncan, S. H., Hold, G. L., Stewart, C. S. & Flint, H. J. The microbiology of butyrate formation in the human colon. *FEMS microbiology letters* **217,** 133–139 (2002).

18. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3,** 32–35 (1950).

19. Heitman, S. J. *et al.* Prevalence of adenomas and colorectal cancer in average risk individuals: A systematic review and meta-analysis. *Clinical Gastroenterology and Hepatology* **7,** 1272–1278 (2009).

20. Symonds, E. L. *et al.* Factors affecting faecal immunochemical test positive rates: Demographic, pathological, behavioural and environmental variables. *Journal of Medical Screening* 0969141315584783 (2015).

21. Kapidzic, A. *et al.* Gender differences in fecal immunochemical test performance for early detection of colorectal neoplasia. *Clinical Gastroenterology and Hepatology* (2015).

22. Levi, Z. *et al.* 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* **104,** 933–938 (2009).

23. Imperiale, T. F. *et al.* Multitarget stool dNA testing for colorectal-cancer screening. *New England Journal of Medicine* **370,** 1287–1297 (2014).

24. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. 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* **79,** 5112–5120 (2013).

25. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 837–845 (1988).