DNA from Fecal Immunochemical Test can replace stool for microbiota-based colorectal cancer screening

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

**Affiliations:**  
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.  
\*Corresponding author: [pschloss@umich.edu](mailto:pschloss@umich.edu)

### Abstract

(≤ 250 words, unstructured)

### Introduction

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. 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 structural exams, like colonoscopy and sigmoidoscopy, are barriers for many people (3,4). Unfortunately less expensive, non-invasive tests, such as the guaiac fecal occult blood test (gFOBT) and fecal immunochemical test (FIT), fail to reliably detect adenomas (5). 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 (6,7). Moreover, we 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 (8,9). One limitation of microbiota-based CRC screening is the need to collect and process separate stool samples for microbiota characterization. If FIT and microbiota characterization are to be included in a single test, it would be more efficient and less expensive to use a single stool sample for both tests. We hypothesize that the small amount of fecal material contained in FIT sampling cartridges is sufficient to perform both hemoglobin quantification and microbiota characterization. To test this hypothesis, we isolated bacterial DNA from the residual buffer of FIT cartriges that had already been used for quantifying hemoglobin concentrations from stool. We then compared the bacterial composition of the FIT cartridge to that of DNA isolated directly from the patient's stool sample and assessed the ability of FIT cartidge-derived DNA to be used for microbiota-based CRC screening.

### Materials and Methods

**Study Design / Diagnoses / Stool Collection.** Stool samples from 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. Patients were 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 FIT-CHEK 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 100µl of buffer was 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 (10). The 16S rRNA gene sequences were curated using the mothur software package, as described previously (10,11). 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) (12). 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.

**Statistical Methods.** All statistical analyses were performed using R (v.3.2.0). Random forest models were generated using the AUCRF algorithm for feature reduction and maximizing model performance (13). The most predictive OTUs were determined based on mean decrease in accuracy when removed from the model. The AUC of ROC curves were compared using the method described by DeLong et al. (**???**). The optimal cutoff for each model was determined using Youden's *J* statistic as implemented in the pROC package in R (14).

### Results

First we tested whether the bacterial community profiles from FIT cartridges recapiulated their stool counterparts. We compared the number of species shared between FIT/stool pairs from the same individual to the number of species shared between individuals. FIT cartridges and stool from the same individual had significantly more bacterial populations in common (Fig. 1A, p<0.001), indicating that community membership was conserved between stool and FIT cartridges. Next we calculated the similarity in community structure between samples using 1-thetaYC index. This metric compares the presence or absence of bacterial populations along with their relative abundance. The bacterial community structure of stool and FIT samples from the same individual were significantly more similar to each other than to stool or FIT from other individuals (Fig. 1B, p<0.001, two-sample Kolmogorov-Smirnov Test). Furthermore the inter-personal variation in community structure between the stool samples of individuals was conserved in samples from FIT cartridges (p<0.001, Mantel test). These findings suggest that that the overall composition of the microbiota in FIT cartridges matches that of stool.

Next we compared the abundance of each genus between FIT cartridges and stool (Fig. 2A). There is a strong correlation between the average abundance of each genus in stool and FIT cartridges (Spearman rho: 0.699, p<0.001), suggesting the abundance of bacterial genera is conserved FIT cartridges. This correlation especially strong for the 100 most abundant genera (Spearman rho: 0.886, p<0.001). However we observed some biases in the abudance of certain taxa. In particular, the genus Pantoea was detected in 399 out of 404 FIT cartridges with an average abundance of 2.4%, but was only detected in 34 stool sample. The genus Helicobacter was detected in 172 FIT cartrdges, 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.

Moreso than bacterial genera, the abundance individuals species used as biomarkers must be conserved for FIT cartridges to be a viable alternative to stool. Several bacterial species have been repeatedly associated with CRC, including *Fusobacterium nucleatum*, *Porphyromonas asaccharolytica*, *Peptostreptococcus stomatis*, and *Parvimonas micra* (**???**,6,7,9). We identified the OTUs associated with each of these species, then compared their abundances in stool and FIT cartridges (Fig. 2B). As expected, their abundance in stool was significantly correlated with their abundance in matched FIT cartridges (all p<0.001). Therefore, bacterial biomarkers of CRC are conserved between stool and FIT cartridges.

Next we tested whether bacterial abundances from FIT cartridges could be used to differentiate healthy patients from those with CRC, as has been demonstrated with stool. First we generated random forest models for differentiating healthy individuals from those with carcinomas based on the abundance of bacterial populations from stool or from FIT cartridges. The AUC-RF algorithm was used to determine the optimal selection of OTUs for each model, based on maximizing the area under the curve (AUC) of a reciever operator characteristic (ROC) curve. The optmial model based on DNA from stool utilized 32 OTUs and had an AUC of 0.853. There was no significant difference in AUC for the model based on DNA from FIT cartridges, which used 28 OTUs and had an AUC of 0.831 (p=0.41). Therefore bacterial DNA from FIT cartridges can is just as predictive as DNA isolated directly from stool.

When screening for CRC it is important to detect adenomas in addition to carcinomas. Unfortunately the gut microbiota, like FIT, cannot reliably detect adenomas on its own. However, we have shown that combining bacterial abundances and hemoglobin concentration into a single model significantly imporoves detection of lesions, especially adenomas, over either method by itself. Therefore we generated random forest models that combined hemoglobin concentration as determined by FIT with bacterial abundance from either stool or FIT catrdiges. As with the cancer versus normal comparison above, there was no difference in AUC the model based on bacteria from stool and the model based on bacteria from FIT cartridge. Again, this shows that bacterial DNA isolated from FIT cartridges can be used in place of stool for detecting colonic lesions based on the microbiota.

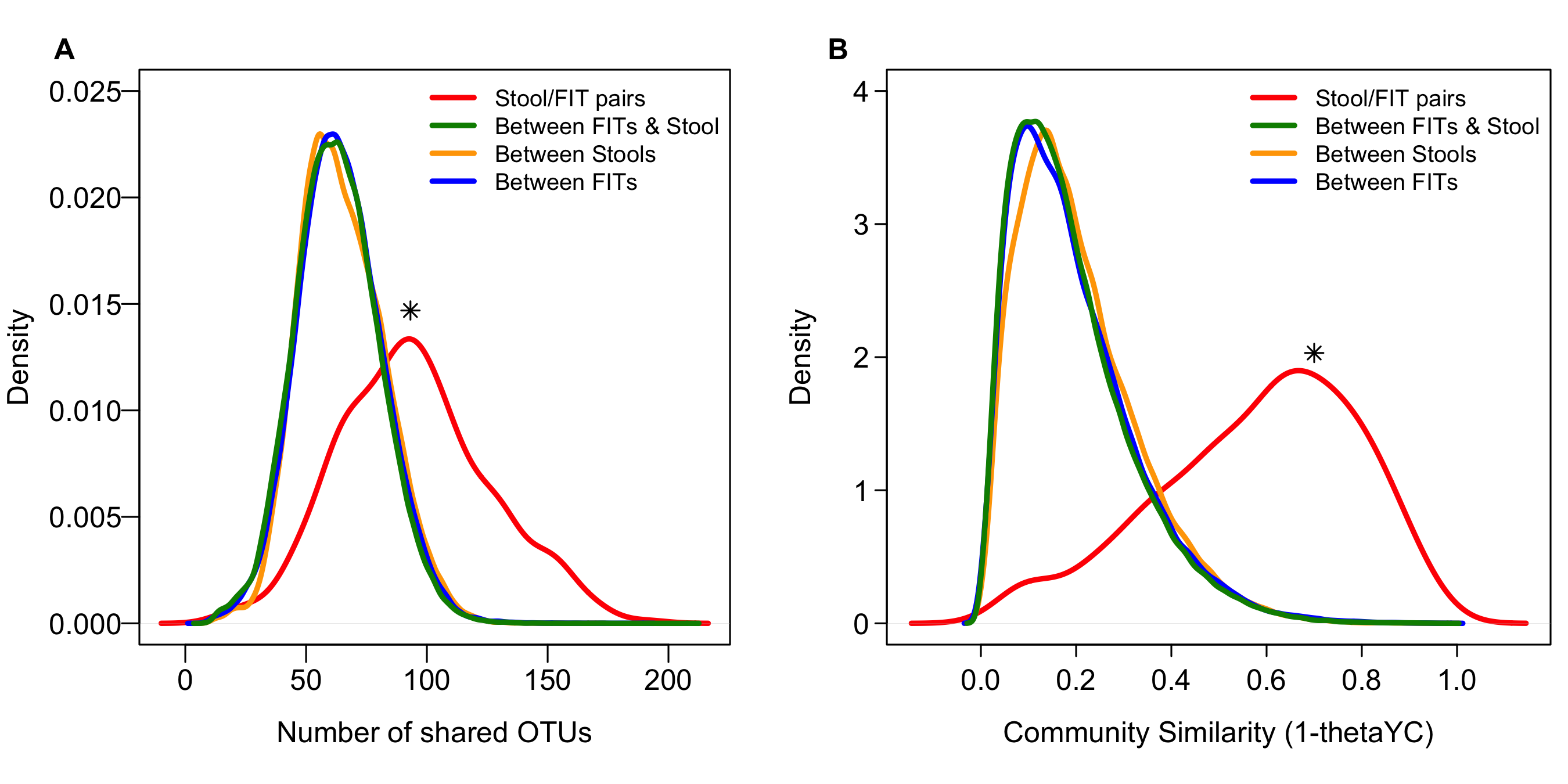
### Discussion

In this study we demonstrated that bacterial DNA isolated from FIT cartridges recapitulates the community structure and membership of patients' stool microbiota. FIT/stool pairs were significantly more similar within an individual than between individuals, and the interindividual differences in stool microbiota structure were conserved in FIT cartridge-derived microbiota. Furthermore, FIT cartridge-derived and stool-derived bacterial DNA were equally predictive for differentiating healthy individuals from those with CRC. Finally, hemoglobin concentrations determined by FIT could be combined with microbiota from FIT cartridges to improve the sensitivity for detecting cancer.

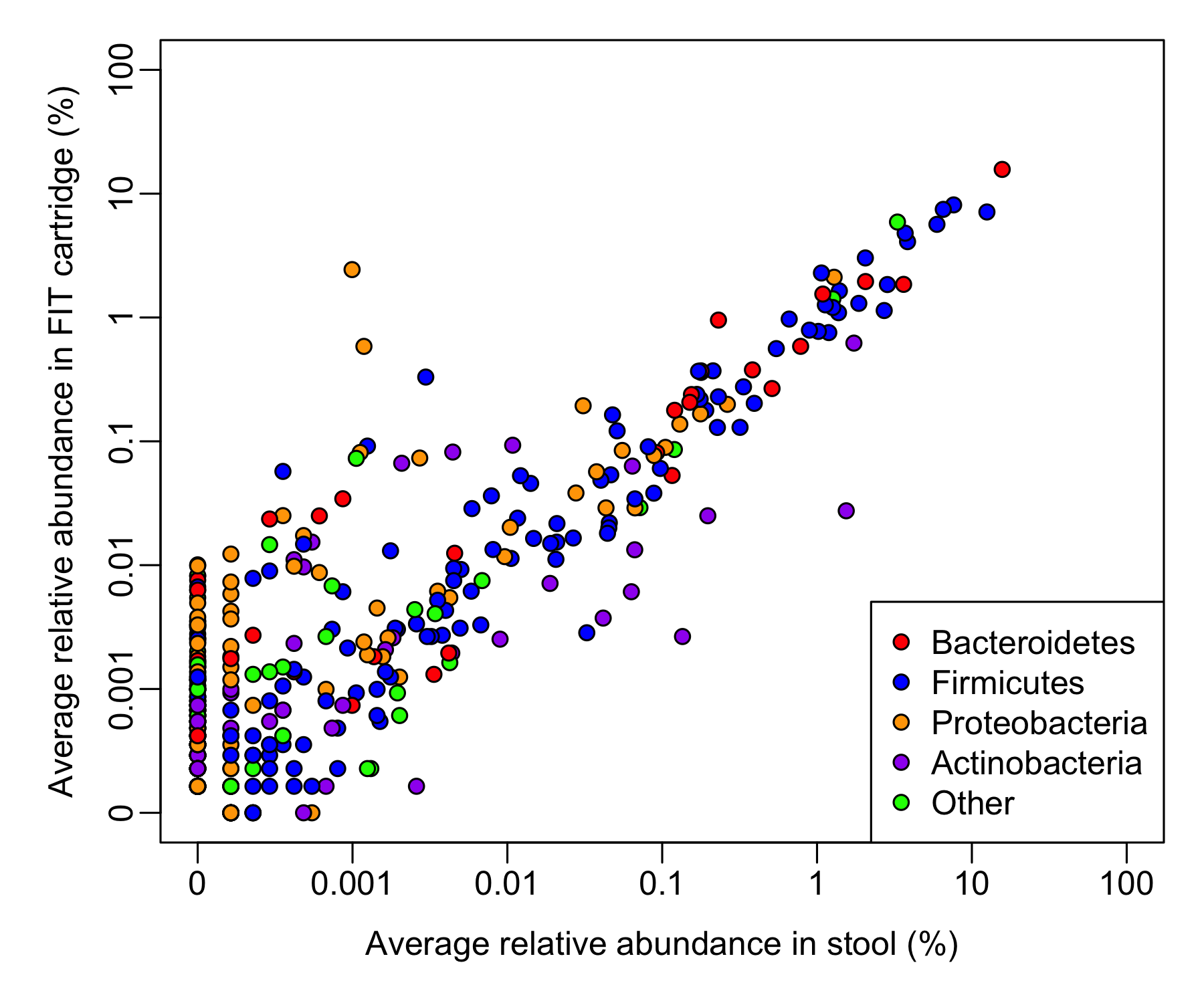
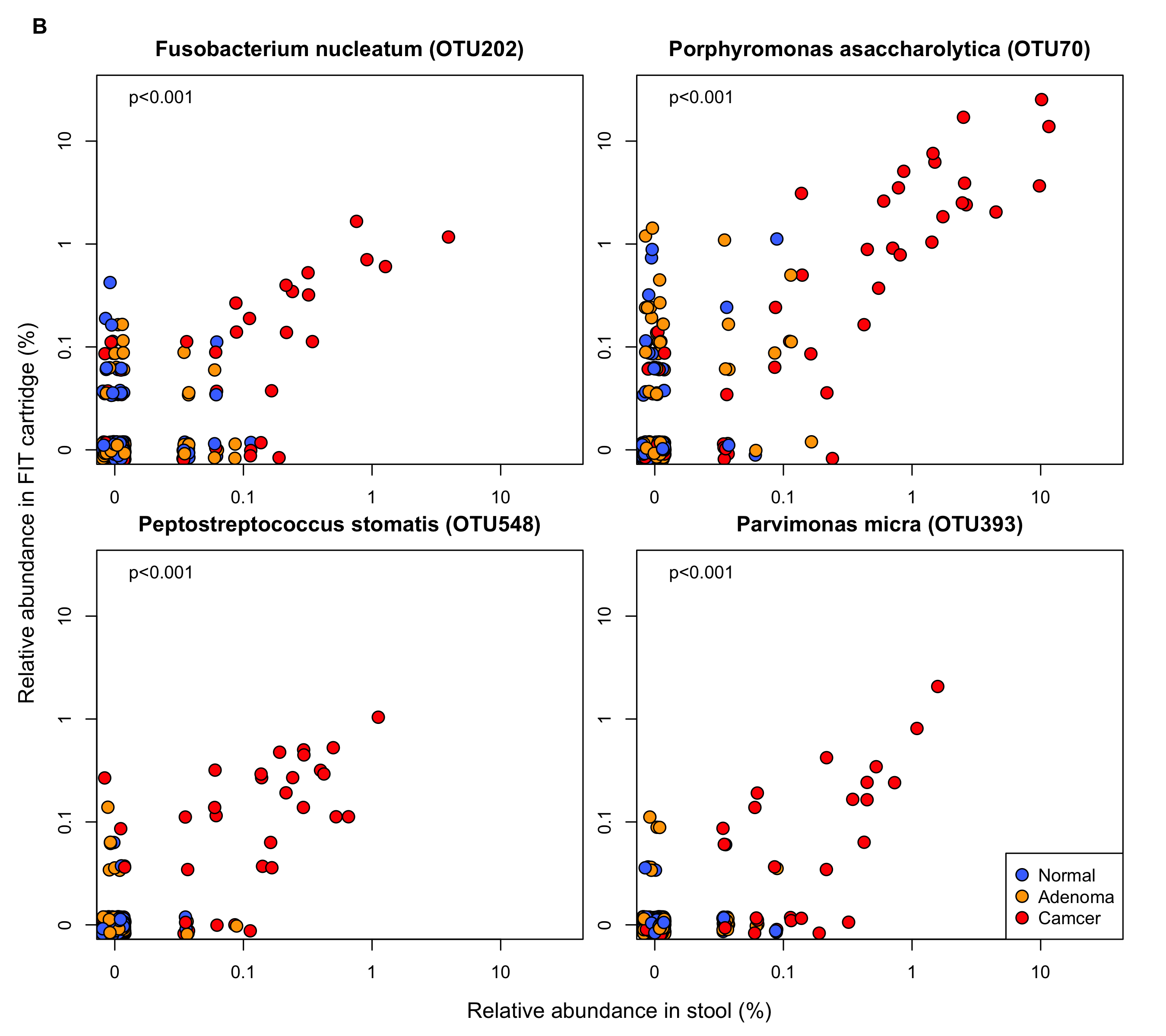
The lower biomass of FIT cartridge buffer compared to stool does increase the risk of results being influenced by contaminants. For example, an OTU associated with Pantoea was found in all FIT cartridge samples and only ## stool samples. This may have been due to reagent contamination that was at a low enough concentration to be undetectable in DNA from stool samples, but high enough to become apparent in the lower biomass FIT cartridges. Alternatively storage conditions could have played a role in biasing certain genera. FIT cartridges spent more time at or near room temperature so that they could be analyzed for hemoglobin concentration. Therefore its possible that certain species were able to grow, shifting their abundance relative to stool. Any biases were limitted to a small number of genera and had no apprent impact on the ability to detect CRC from FIT cartridged-derived DNA.

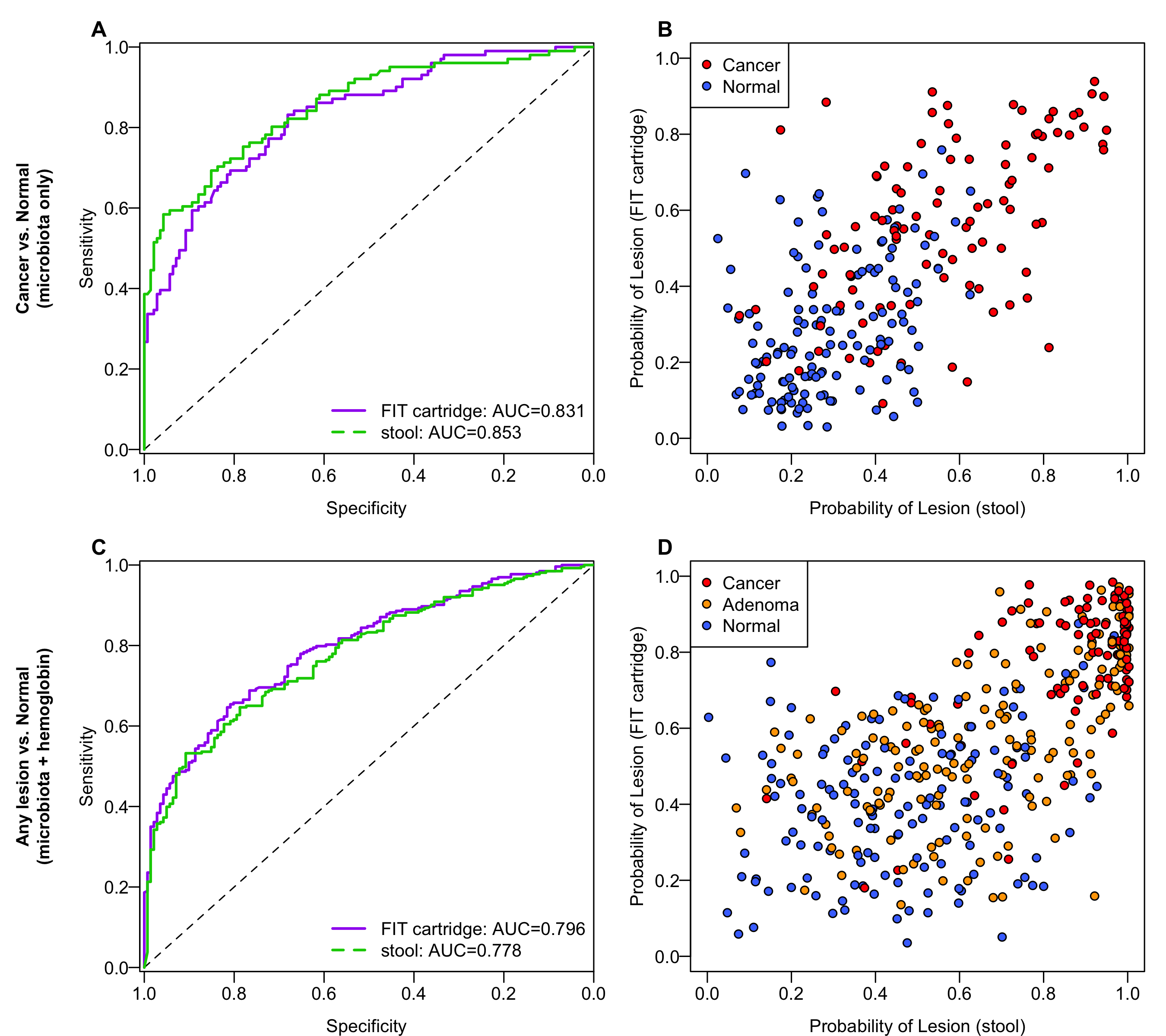
These findings demonstrate the potential for using DNA from FIT cartridges for microbiota-based screening. This could eliminate the need to collect and process separate stool samples, reducing the cost of screening. Furthermore, annual FIT cartridges could be used to moniter changes to a patients microbiota over time, making it possible to detect any shifts toward a disease-associated microbiome. It would also be possible to use FIT cartridges, rather than separate stool samples, for future studies on the role of the gut microbiota and CRC. Since FIT cartridges are already collected for CRC screening, it may be easier to recruit patients for large-scale validations of microbiota-based screening methods.

### Figures



**Figure 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).

  **Figure 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. (B) Scatterplots of relative abundances of 4 OTUs frequently associated with CRC.

 **Figure 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 from FIT cartridge DNA or stool. (B) Probability of having cancer for each patient according to microbiota-based models from stool or FIT cartridges. (C) ROC curves for distinguishing patients with adenomas or carcinomas from healthy patients using both hemoglobin concentration and microbiota from either stool or FIT cartrdges. (D) Probability of having a lesion for each patient based on the models from C.

### References (≤25 references)

1. Siegel R, DeSantis C, Jemal A. Colorectal cancer statistics, 2014. CA: a cancer journal for clinicians. Wiley Online Library; 2014;64:104–17.

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, et al. The importance of health insurance as a determinant of cancer screening: evidence from the Women’s Health Initiative. Preventive medicine. Elsevier; 2000;31:261–70.

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. Elsevier; 2010;38:508–16.

5. Hundt S, Haug U, Brenner H. Comparative evaluation of immunochemical fecal occult blood tests for colorectal adenoma detection. Annals of Internal Medicine. Am Coll Physicians; 2009;150:162–9.

6. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. Molecular systems biology. EMBO Press; 2014;10:766.

7. Yu J, Feng Q, Wong SH, Zhang D, Liang Q yi, Qin Y, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut. BMJ Publishing Group Ltd and British Society of Gastroenterology; 2015;gutjnl–l2015.

8. Zackular JP, Rogers MA, Ruffin MT, Schloss PD. The human gut microbiome as a screening tool for colorectal cancer. Cancer Prevention Research. AACR; 2014;7:1112–21.

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

10. 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. Am Soc Microbiol; 2013;79:5112–20.

11. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and environmental microbiology. Am Soc Microbiol; 2009;75:7537–41.

12. 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. Am Soc Microbiol; 2007;73:5261–7.

13. Calle ML, Urrea V, Boulesteix A-L, Malats N. AUC-RF: A new strategy for genomic profiling with random forest. Human heredity. Karger Publishers; 2011;72:121–32.

14. Youden WJ. Index for rating diagnostic tests. Cancer. Wiley Online Library; 1950;3:32–5.