

RESEARCH

Open Access



DNA from fecal immunochemical test can replace stool for detection of colonic lesions using a microbiota-based model

Nielson T. Baxter^{1,2}, Charles C. Koumpouras¹, Mary A. M. Rogers², Mack T. Ruffin IV³ and Patrick D. Schloss^{1*}

Abstract

Background: 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.

Results: 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.

Conclusions: 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.

Keywords: Colorectal cancer, Gut microbiome, Microbiota, Fecal immunochemical test, Random forest

Background

Although colorectal cancer (CRC) mortality has declined in recent decades, it remains the second leading cause of death among cancers in the USA [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 Americans 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]. Some do not consider these invasive tests as a primary screening option, but they are in the USA [5]. Unfortunately, noninvasive tests, such as the guaiac fecal occult blood test (gFOBT), fecal immunochemical test (FIT), and the

multitarget DNA test, fail to reliably detect adenomas [6, 7] (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 noninvasive 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 [8–11]. 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 [8, 9, 11]. 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

* Correspondence: pschloss@umich.edu

¹Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI, USA

Full list of author information is available at the end of the article

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. The study was a multicenter, prospective, cross-sectional cohort validation study planned for 600 participants. The planned sample size was 200 participants with colorectal cancer, 200 participants with colorectal adenomas, and 200 participants with normal colonoscopies. The baseline visit was done either prior to a scheduled colonoscopy or after a colonoscopy was done. If a subject was suspected of having a colon adenocarcinoma or an adenoma, the baseline samples were collected before the colonoscopy or exposure to colonic preps so the lesions are present when stool collection is done. Normal controls were recruited at the time of their scheduled colonoscopy, prior to their colonoscopy, or within 4 months after completing a screening colonoscopy indicating no adenomas or colorectal cancer to provide a set of samples. The normal patients' stool samples were collected before colonic preps or at least 4 weeks after the colonoscopy was completed. 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 nonpolyposis 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 (MA, USA), Houston (TX, USA), and Ann Arbor (MI, USA). All participants collected a whole evacuated stool in a collection container with no preservatives. Stool samples were collected, 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.). FIT cartridges were stored at 4°C for up to 48 hours

prior to hemoglobin quantification using an OC-Auto Micro 80 automated system (Polymedco Inc.). FIT cartridges were then stored at -20°C until being thawed for DNA extraction. 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 the FIT cartridge buffers and the matching stool samples from which they were collected 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 ribosomal RNA (rRNA) gene was amplified using custom barcoded primers and sequenced as described previously using an Illumina MiSeq sequencer [12]. The 16S rRNA gene sequences were curated using the mothur software package, as described previously [12, 13]. 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) [14]. 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 area under the curve (AUC)-RF algorithm for feature reduction and maximizing model performance [15]. The most predictive OTUs were determined based on mean decrease in accuracy when removed from the model. The AUC of receiver operator characteristic (ROC) curves were compared using the method described by DeLong et al. [16] as implemented in the pROC R package [17]. There was sufficient power (80 %) to detect a 0.081 difference in AUC with alpha set at 0.05.

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. Second, we compared the number of OTUs shared within FIT/stool pairs from the same patient to the number of OTUs shared between patients (Fig. 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. Third, we calculated the similarity in community structure between samples using 1-thetaYC index [18]. 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. 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.

Next, we observed a significant correlation between the abundance of each genus in the paired FIT cartridge and stool samples (Fig. 2a, Spearman's rho 0.699, $p < 0.001$). This suggested that the abundance of bacterial genera was conserved. This correlation was especially strong when comparing the 100 most abundant genera from stool (Spearman's 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* [9–11, 19]. As expected, the abundance of these species in stool was significantly correlated with their abundance in matched FIT cartridges

(all $p < 0.001$, Spearman's rho ≥ 0.352) (Fig. 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 the overall bacterial community structure and the abundance of specific taxa in FIT cartridges and stool were similar.

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 [11]. Using DNA from the FIT cartridge, the optimal model utilized 28 OTUs and had an AUC of 0.831 (Fig. 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 were correlated between the models generated using DNA isolated from the FIT cartridges and stool samples (Spearman's rho 0.633, $p < 0.001$, Fig. 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. 3c). Again, the probabilities of individuals having lesions according to the two models were significantly correlated (Spearman's rho 0.389, $p < 0.001$ Fig. 3d). These findings demonstrated that models based on bacterial DNA from FIT cartridges

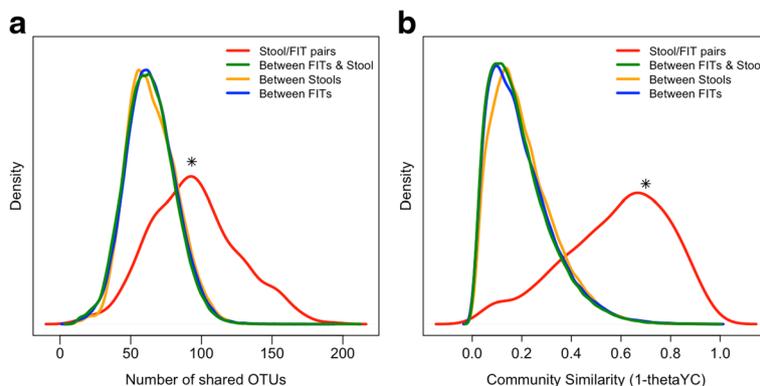
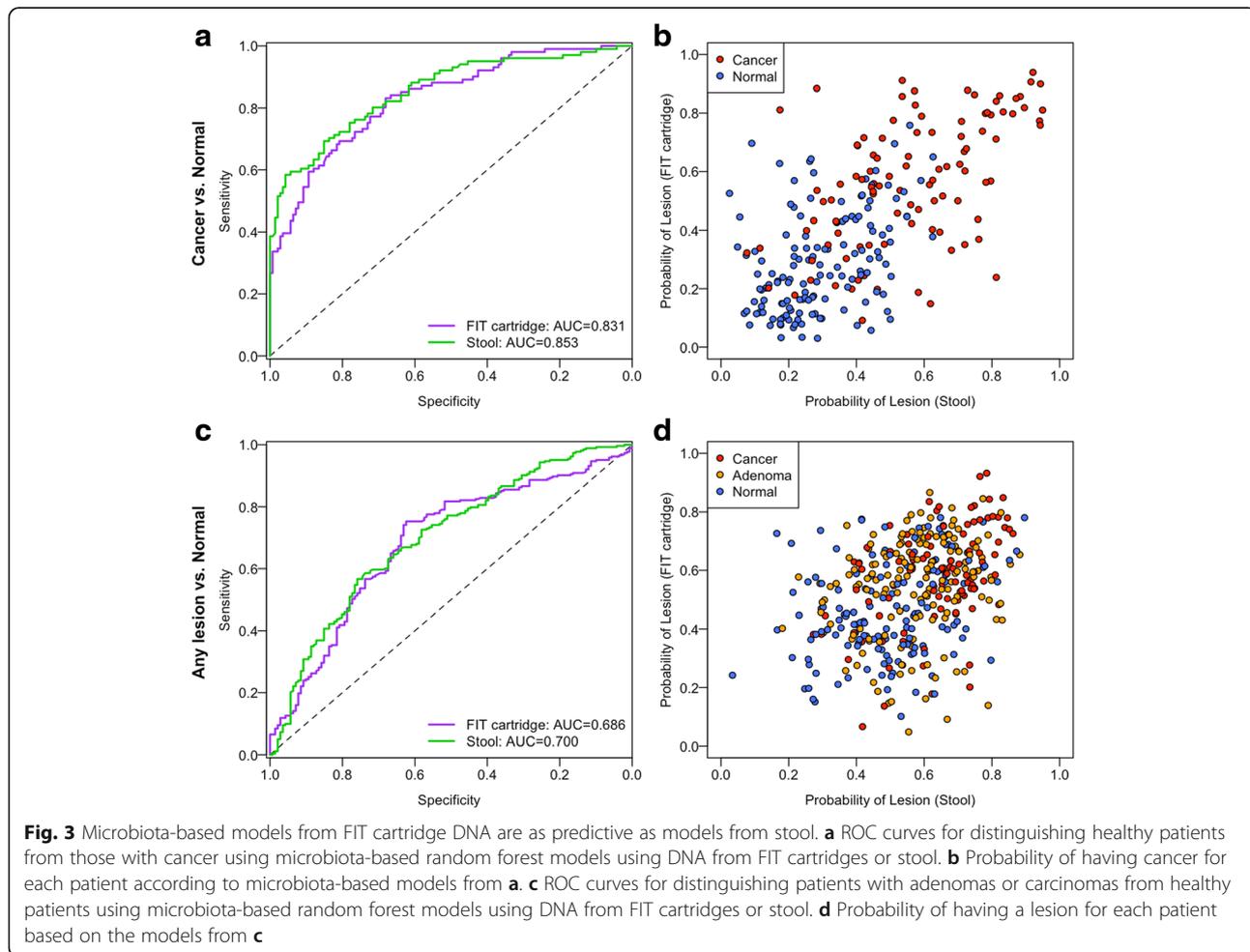
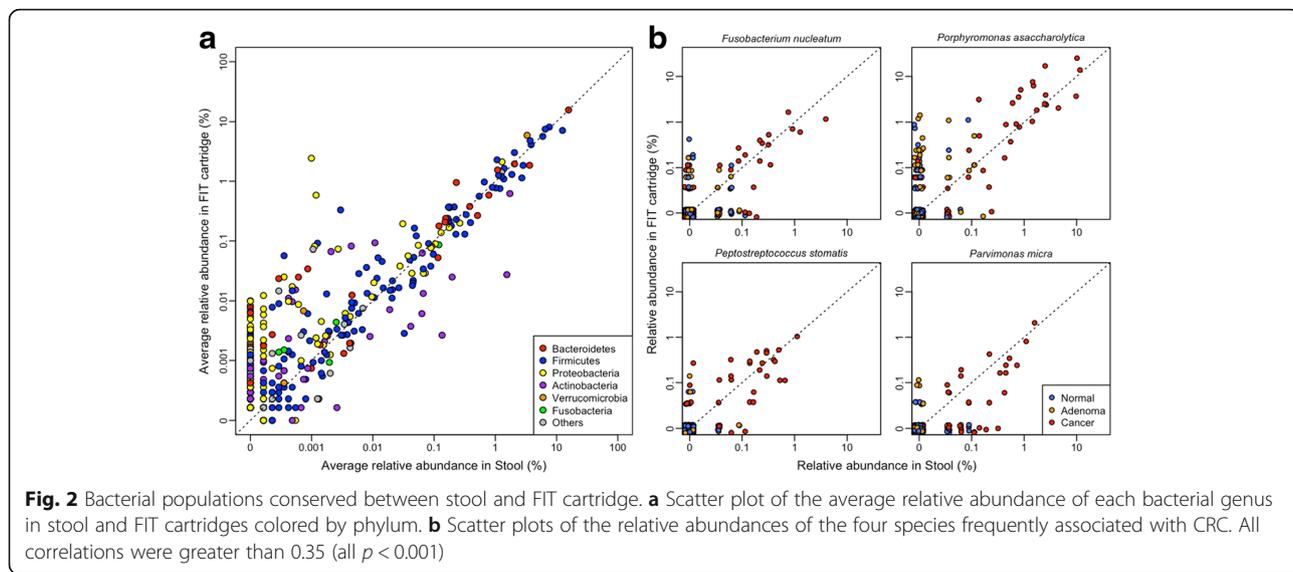


Fig. 1 Bacterial community structure from FIT cartridge recapitulates stool. Density plots showing distribution of the number of shared OTUs (a) and community similarity (b) between groups of samples ($*p < 0.001$ two-sample Kolmogorov-Smirnov test)



were as predictive as models based on DNA isolated directly from stool.

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 [20]. They found reproducible biases according to sampling method and time at ambient temperature. Likewise, we observed biases in the abundance of 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. It is likely 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 cartridges themselves [21]. This seems to be the case for genus *Pantoea*, since a sequence classified as *Pantoea* was detected in several uninoculated FIT cartridges used as negative controls. It is also possible that 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 while being analyzed for hemoglobin concentration. Therefore, it is possible that certain bacterial populations, especially aerobes, were able to grow. Alternatively, there could have been a difference in DNA extraction efficiency between the stool samples and FIT cartridge buffer. Actinobacteria, which were overrepresented in stool samples, require mechanical lysis for DNA extraction [22]. It is possible that the particulates in stool allowed for more efficient mechanical lysis of Actinobacteria cells compared to the FIT cartridge buffer. Regardless, these biases had no effect on the ability to detect CRC from FIT cartridge-derived DNA, as the random forest feature selection process did not select these populations.

Colonoscopy, rather than stool screening, remains the more effective CRC screening approach. However, noninvasive stool tests, including those based on microbiota, might have future utility in directing high risk subjects toward a colonoscopy, increasing the overall performance of

CRC screening and reducing morbidity and mortality. The ability to assay hemoglobin quantification and microbiota-based biomarkers with single sampling kit would reduce cost and improve efficiency. It would also facilitate combining results from the two methods into single model for predicting the presence of lesions, which has been shown to improve sensitivity over either method alone. Incorporating other host-associated stool biomarkers and clinical data could further improve the accuracy of such a model.

Conclusions

The findings from this study demonstrate the potential for using DNA from FIT cartridges for microbiota-based screening. This could reduce the need to collect and process separate stool samples, decreasing the cost of screening. It is also 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.

Abbreviations

AUC: Area under the curve; FIT: Fecal immunochemical test; gFOBT: Guaiac fecal occult blood test; OTU: Operational taxonomic unit; ROC curve: Receiver operating characteristic curve

Acknowledgements

The authors thank the Great Lakes-New England Early Detection Research Network for providing the fecal samples that were used in this study.

Funding

This study was supported by funding from the National Institutes of Health to P. Schloss (R01GM099514, P30DK034933) and to the Early Detection Research Network (U01CA86400).

Availability of data and materials

Raw sequencing data and patient metadata are available at the NCBI Sequence Read Archive (SRP062005 for stool samples and metadata, SRP073172 for FIT cartridges). The data processing steps for going from the raw sequence data to the final manuscript are available at http://www.github.com/SchlossLab/Baxter_FITs_Microbiome_2016.

Authors' contributions

PDS, MTR, MAMR, and NTB were involved in the conception and design of the study. NTB and CCK performed the DNA extractions and 16S rRNA gene sequencing. NTB analyzed the data. All authors interpreted the data. NTB and PDS wrote the manuscript. All authors reviewed and revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

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.

Author details

¹Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI, USA. ²Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA. ³Department of Family and Community Medicine, Penn State Hershey Medical Center, Hershey, PA, USA.

Received: 25 August 2016 Accepted: 31 October 2016

Published online: 14 November 2016

References

- Siegel R, DeSantis C, Jemal A. Colorectal cancer statistics, 2014. *CA Cancer J Clin.* 2014;64:104–17.
- Centers for Disease Control and Prevention. Vital signs: colorectal cancer screening test use—United States, 2012. *MMWR Morb Mortal Wkly Rep.* 2013;62:881.
- Hsia J, Kemper E, Kiefe C, Zapka J, Sofaer S, Pettinger M, Bowen D, Limacher M, Lillington L, Mason E, et al. The importance of health insurance as a determinant of cancer screening: evidence from the Women's Health Initiative. *Prev Med.* 2000;31:261–70.
- Jones RM, Devers KJ, Kuzel AJ, Woolf SH. Patient-reported barriers to colorectal cancer screening: a mixed-methods analysis. *Am J Prev Med.* 2010;38:508–16.
- Bibbins-Domingo K, Grossman DC, Curry SJ, Davidson KW, Epling JW, García FA, Gillman MW, Harper DM, Kemper AR, Krist AH, et al. Screening for colorectal cancer: US preventive services task force recommendation statement. *JAMA.* 2016;315:2564–75.
- Hundt S, Haug U, Brenner H. Comparative evaluation of immunochemical fecal occult blood tests for colorectal adenoma detection. *Ann Intern Med.* 2009;150:162–9.
- Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, Ahlquist DA, Berger BM. Multitarget stool DNA testing for colorectal-cancer screening. *N Engl J Med.* 2014;370:1287–97.
- Zackular JP, Rogers MA, Ruffin MT, Schloss PD. The human gut microbiome as a screening tool for colorectal cancer. *Cancer Prev Res.* 2014;7:1112–21.
- Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Böhm J, Brunetti F, Habermann N, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol.* 2014;10:766.
- Yu J, Feng Q, Wong SH, Zhang D, Liang Q yi, Qin Y, Tang L, Zhao H, Stenvang J, Li Y, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut.* 2015;0: 1–9.
- Baxter NT, Ruffin MT, Rogers MA, Schloss PD. Microbiota-based model improves the sensitivity for detecting colonic lesions. *Genome Medicine.* 2016;8:1–10.
- 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. *Appl Environ Microbiol.* 2013;79:5112–20.
- 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;75:7537–41.
- 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;73:5261–7.
- Calle ML, Urrea V, Boulesteix A-L, Malats N. AUC-RF: a new strategy for genomic profiling with random forest. *Hum Hered.* 2011;72:121–32.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44:837–45.
- 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.
- Yue JC, Clayton MK. A similarity measure based on species proportions. *Commun Stat Theory Methods.* 2005;34:2123–31.
- 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.
- Sinha R, Chen J, Amir A, Vogtmann E, Inman KS, Flores-Munguia R, Sampson JN, Knight R, Chia N, et al. Collecting fecal samples for microbiome analyses in epidemiology studies. *Cancer Epidemiology Biomarkers & Prevention.* 2015;25:407–416.
- 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 Biol.* 2014;12:87.
- Walker AW, Martin JC, Scott P, Parkhill J, Flint HJ, Scott KP. 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and pCR primer choice. *Microbiome.* 2015;3:1.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

