Reviewer reports:

Reviewer #1:

1.    The study cohort from which the samples were derived is insufficiently described. As the prevalence of adenomas is around 20% and that of CRC <5% the investigators must have selected samples from a much larger stool collection. How many total paired samples were obtained and assuming that all CRC cases were included, how were adenoma cases and controls selected for inclusion into this study. Did they obtain samples from cases only AFTER colonoscopy detection (or other detection methods)? How about control samples?

The study was a multi-center, prospective, cross-sectional cohort validation study planned for 600 patients. The sample size planned for the study was 200 participants with colorectal cancer, 200 participants with colorectal adenomas, and 200 participants with normal colonoscopies. The study visit was done either after a colonoscopy was done or prior to a scheduled colonoscopy. If a subject was suspected of having a colon adenocarcinoma or an adenoma, the baseline samples was collected before the colonoscopy 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 four weeks after the colonoscopy was completed. We did not select study participants from a larger cohort. The recruitment strategy was not to consent all patient scheduled for a colonoscopy at a study center. Each site used local measures to target recruitment of patients likely to have colonic disease who were scheduled for colonoscopy. The study participants had to meet eligibility criteria described in the manuscript. We have added this information to the methods section. (Line 78)

2.    Where FIT/stool samples for each subject obtained from the same stool? If "YES" say so if "No" describe the difference between the two samples.

FIT/stool pairs were collected from the same stool sample. We have added clarification in the methods section (Line 90)

3.    The approach taken for establishing similarity between FIT and stool samples presented in Fig. 2 appears suboptimal. Rather than presenting average relative abundance, which might mask differences in sample pairs, the authors might want to consider presenting average difference within pair distances for each genus.

We appreciate the reviewer’s suggestion to improve Figure 2. Below we present the percentage point difference and percent difference between stool and FIT for the top 100 genera. The percentage point difference figure makes the data look “too good” by masking the variation near zero. In the original figure we used a log-scale to show the variation increases as the relative abundances decrease. This approach is not possible since the log of zero is undefined. When we plotted the percent difference between stool and FIT the differences at low relative abundances overwhelm the plot. For example, if a genus differs by 0.002 with an average relative abundance of 0.001, then that would suggest a much larger difference than if a genus differed by 0.002 with an average relative abundance of 0.01. We have decided to keep the original figure since it does the best job of demonstrating the fidelity of FIT to the stool result while also showing that at low relative abundances there is more variation.

4.    The authors allude in the discussion to potential contamination as a reason for different abundances seen between FIT/stool for some taxa. Inclusion of appropriate negative controls would have allowed the authors to avoid such unscientific speculation. While this is a very minor issue in the context of the presented work it questions the rigor of the experimental design.

Pantoea sequences were detected in uninoculated FIT cartridges that were included as controls for the experiment. This information has been added to the discussion. However, that does not explain taxa that were underrepresented in FIT cartridges. We have added the possibility that Actinobacteria were enriched in stool samples due to differences in extraction efficiency, as this is a well-documented source of bias for Actinobacteria.

5.    While the authors present strong data to indicate the utility of FIT derived microbiota analysis they limit their conclusion to "It may be possible to use FIT ...", which of course we knew beforehand (anything may be possible). Either provide a firmer conclusion, which appears appropriate based on the data, or state what specifically would be required to derive at a stronger conclusion.

We have removed the uncertainty from that statement by changing it to “It is also possible to use FIT cartridges…” (Line 195)

6.    Colonoscopy, rather than stool screening, is the more effective CRC screening approach. As this has important public health implications this issue should be discussed.  Non-invasive stool tests, including those based on microbiota, might have future utility in directing high risk subjects towards a colonoscopy, increasing the overall performance of CRC screening and reducing morbidity and mortality from this important disease.

We agree that this is an important consideration. However, these considerations were discussed in more detail in an earlier publication using the same dataset (Baxter 2016 *Genome Medicince*). The current paper is focused on the possibility of using FIT cartridges as source of material for a microbiota-based test, rather than how a microbiota-based test would be integrated with existing screening practices.

Reviewer #2:

The paper is entitled: "DNA from fecal immunochemical test can replace stool for microbiota-based colorectal cancer screening." This is over stating the data reported in this paper, as it implies that stool microbiota is currently used in standard practice for cancer screening which it is not. I think that the title needs to be changed to reflect more accurately the scientific findings.

We have modified the title to more accurately reflect the findings. “DNA from fecal immunochemical test can replace stool for detection of colonic lesions using a microbiota-based model”

The authors report in the introduction that the update from screening programs is low. Although it needs to be increased, there is significant global variation, and in Europe, and specifically in the UK the uptake in national screening is 55-60%. (Logan, R.F.A. et al. Gut doi:10.1136/gutjnl-2011-300843).

Yes, we agree with the comment. We have revised the introduction to clarify that the screening adherence statistic is specific to the United States.

The authors state that "The high cost and invasive nature of procedures, such as colonoscopy and sigmoidoscopy are barriers for many people".  These tests are performed on the basis of the FIT test, and they are not the primary screening tests and this needs to be reflected.

In some countries the primary screening modality is non-invasive test like FIT. However, the primary screening modality options in some countries like the United States includes invasive procedures like flexible sigmoidoscopy and colonoscopy. (*JAMA.*2016;315(23):2564-2575) However, the performance of FIT alone is limited with opportunity for improvement with the addition of other stool markers. We have revised the introduction to reflect the US centric view that includes invasive tests are possible primary screening modalities.

More information on patient recruitment is needed. Stools samples were obtained through the Great Lakes-New England Early Detection Research Network. It is important as it is not clear if other GI conditions that may warrant a colonoscopy may be confounding this data set. The assumption is that the control group (negative for adenoma or cancer) were 'healthy' may not be true in this scenario.

It is not clear whether these patients were part of a standard screening program or whether they were recruited specifically for the purpose of this study.

See response above. With respect to the issue that the healthy or normal colons were not healthy, the exclusion criteria was colorectal cancer patients with prior treatment of any type, history of inflammatory bowel disease, known HIV or hepatitis, or other cancers within three years. Study participants were without any symptoms. All of these exclusion criteria were reported in the original paper.

Samples were collected at four different centers. What steps were taken to standardized sample collection?

Yes, at all four sites whole evacuated stool was collected in a hat with no preservatives, then shipped on ice to a single processing center. (line ##)

There is no calculation provided to demonstrate the numbers required to demonstrate statistical power.

An a priori power analysis was not performed as the number of samples used was limited to the number patient samples in the dataset. However, the statistical tests used for comparing beta diversity in Figure 1 and for correlating taxa abundances in Figure 2 were statistical significantly, clearly indicating that there was sufficient power for those comparisons. The only tests in which there was not statistical significance were the comparisons of ROC curves in Figures 3A and 3C. In those cases we hypothesized that there would NOT be a difference in AUC. Nonetheless, we have added a power calculation to methods (line ##) per the reviewer’s request. We had sufficient power to detect a difference in AUC of 0.081 (9.5% for cancer vs. normal comparison). The differences we observed between FIT and stool were 0.022 (2.5%) for cancer vs normal and 0.0014 (2.0%) lesion vs normal. We contend that a 2.5% difference in AUC is negligible.

Not enough information is provided here on the methods of faecal sample collection. Were these collected on site or in the community? Were both the stool and the FIT samples collected simultaneously?

FIT/stool pairs were collected from the same stool sample. We have added clarification in the methods section (Line 90)

How long were they stored at ambient conditions before they were put on ice and was this controlled? This is critical, as in the UK, samples are sent via the post. Therefore, it may be at room temperature for days before analysis. The authors quote the paper by Sinah et al. (ref 19), but this paper significantly lacks a similar systematic approach to the sample analysis to provide analytical insight into sample stability, which is critical for determining whether the findings reported here are translatable into clinical practice.

FIT cartridges were store 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.

Absolutely no clinical data is provided on the patients enrolled into this study. Even basic demographic data is missing. Adenomas are a heterogenous pathology, and it is completely unclear what the anatomical location of these cancers were and their stage. This is important for understanding possible bias in the data set.

Clinical data, including demographics, tumor staging, antibiotic use, FIT results, etc. are available with the raw sequencing data through the NCBI SRA database. We have added this information to the “Availability of data and materials” section.

Possible biases related to patient demographics, medications, BMI, smoking, etc. have been tested with samples from this dataset in two previous publications (Zackular 2014 *Canc. Prev. Res.*, Baxter 2016 *Genome Medicine*).

The bacterial community structure of stool and FIT samples from the same patient were significantly more similar to each other than to stool or FIT from other patients and that inter community structure is conserved. The authors use a 1-thetaYC index and mantel test. suggesting that the interpatient variation in community structure between the stool samples of patients was conserved. The main finding from Sinah et al. confirmed the findings from numerous other studies that inter-individual variation is the most important cause of community variance. It would have therefore been interesting to see if this finding is reproduced in this data set, and to see if these data could be repeated with multiple sampling time points.

Figure 1 shows that the intra personal variation in community structure and membership was less than the variation between sampling methods and the interpersonal variation. Additionally, the there was no difference in variation for comparisons within a sampling method or between sampling method, suggesting that the interpersonal variation is much greater than the variation between sampling methods. The results demonstrate (albeit indirectly) that interpersonal variation is the greatest source of variation in this dataset.

The discussion section is short, and I think that it could be expanded to explore in more detail the findings reported here.

We are grateful that the reviewer is enthusiastic about this research. We agree that the discussion is relatively short; however, the study was relatively simple and follows up on our other work published in Genome Medicine. The goal of this manuscript was relatively simple, but important - we wanted to know whether the bacterial DNA extracted from the reagents in the FIT cartridge could serve as a surrogate for whole stool. Our discussion summarizes the results (1st paragraph), discusses caveats dealing with sampling and contamination (2nd paragraph), and points a way forward (3rd paragraph). We think that most readers will appreciate a compact Discussion and Conclusion section that does not reach beyond the data.