The Microbiota and Individual Community Members in Colorectal Cancer: Is There a Common Theme?

Marc A Sze[1] and Patrick D Schloss[1]†

† To whom correspondence should be addressed: pschloss@umich.edu

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

Co-author e-mails:

- marcsze@med.umich.edu

Abstract

Background.

Results.

Conclusions.

Keywords

microbiota; colorectal cancer; polyps; adenoma; meta-analysis.

Background

Results

Fecal Diversity is Lower in Those with Carcinoma and Increases Relative Risk: Using power transformed and Z-score normalized alpha diversity metrics both evenness and the Shannon diversity metrics in feces are lower in those with carcinoma then in controls but not for tissue samples [Figure 1]. Using linear mixed-effects to control for study and variable region there was a significant decrease from control to adenoma to carcinoma for both evenness (P-value = 0.025) and Shannon diversity (P-value = 0.043). This effect was not observed in tissue when additionally controlling for whether the sample came from the same individual (P-value > 0.05). For fecal samples a decrease in Shannon diversity and evenness resulted in a significantly increased relative risk for carcinoma (P-value = 0.01 and P-value = 0.0011, respectively) [Figure 2]. Although these values were significant the effect size was relatively small for both metrics (Shannon RR = 1.31 and evenness RR = 1.34) [Figure 2]. There was no increased relative risk for these metrics for adenoma or for tissue in general [Figure S1-3]

Genera Previously Associated with Carcinoma Increases Relative Risk More than Alpha Diversity: Both fecal and tissue samples had a significantly increased RR for carcinoma but not for adenoma [Figure 3] which was greater than either evenness or Shannon diversity [Figure 2 & 3]. The relative risk did not increase when considering the total abundance or increasing number of carcinoma associated genera [Figure 3]. The RR effect size was greater for stool (RR range = 1.78 - 2.64) then for tissue (RR range = 1.33 - 1.53) . This decrease may be explained by the fact that tissue samples include matched samples.

Section 3

Section 4

Discussion

Conclusion

Methods

Obtaining Data Sets: Studies used for this meta-analysis were identified through the review articles written by Keku, et al. and Vogtmann, et al. [1,2]. All studies were included that used tissue or feces as their sample source for 16S rRNA gene sequencing analysis. Studies using either 454 or Illumina sequencing technology were included. Only data sets that had the raw sequences available for analysis were included. Some studies did not have publically available raw sequences or did not have meta data in which the authors were able to share. After this filtering step the following studies remained: Ahn [3], Baxter [4], Brim [5], Burns [6], Chen [7], Dejea [8], Flemer [9], Geng [10], Hale [11], Kostic [12], Lu [13], Sanapareddy [14], Wang [15], Weir [16], and Zeller [17]. The Zackular [18] study was not included becasue the 90 individuals analyzed within the study are contained within the larger Baxter study. The Kostic study was not used since after sequence processing all the case samples did not have more than 100 sequences remaining. This left a total of 13 studies in which complete analysis could be completed.

Data Set Breakdown: In total there were 7 studies with only fecal samples (Ahn, Baxter, Brim, Hale, Wang, Weir, and Zeller), 5 studies with only tissue samples (Burns, Dejea, Geng, Lu, Sanapareddy), and 2 studies with both fecal and tissue samples (Chen and Flemer). The total number of individuals initially run through the sequence processing for the fecal samples was 1899 and for the tissue samples was 462.

Sequence Processing: For the majority of studies raw sequences were downloaded from the SRA (ftp://ftp-trace.ncbi.nih.gov/sra/sra-instant/reads/ByStudy/sra/SRP/) and metadata was obtained from the following website: http://www.ncbi.nlm.nih.gov/Traces/study/ by searching the respective accession number of the study. Of the studies that did not have sequences and meta data on the SRA one study had the data stored on DBGap [3] and four studies the data was obtained directly from the authors [9,11,14,16]. Each study was processed using the mothur (v1.39.3) software program [19]. Where possible quality filtering utilized the default methods used in mothur for either 454 or Illumina based sequencing. If it was not possible to use these defaults the author stated quality cut-offs were used instead. Chimeras were identifed and removed using the VSEARCH [20] program and *de novo* OTU clustering at 97% similarity using the OptiClust algorithm [21] was utilized.

Statistical Analysis: All statistical analysis after sequence processing utilized the R software package (v3.4.2). For the alpha diversity analysis values were power transformed using the rcompanion (v1.10.1) package and then Z-score normalized using the car (v2.1.5) package. Testing for alpha diversity differences utilized linear mixed-effect models created using the lme4 (v1.1.14) package to correct for both study and variable region effect in the diversity measures when analyzing colorectal cancer groups. Relative Risk was analyzed using both the epiR (v0.9.87) and metafor (v2.0.0) packages. Relative risk significance testing utilized the chi-squred test. Beta-diversity differences utilized a Bray-Curtis distance matrix and PERMANOVA executed with the vegan (v2.4.4) package. Random Forest models were built using both the caret (v6.0.77) and randomForest (v4.6.12) packages. Random Forest testing of the obtained AUC versus a random model AUC utilized T-tests. Power analysis and estimations were made using the pwr (v1.2.1) and statmod (v1.4.30) packages. All figures were created using both ggplot2 (v2.2.1) and gridExtra (v2.3) packages.

Study Analysis Overview: Alpha diversity was first assessed for differences between controls and adenoma versus cancer and controls versus adenoma. We analyzed the data using linear mixed-effect models, and relative risk. Beta-diversity was then assessed for each inidividual study. Next, four specific CRC-associated genera (*Fusobacterium*, *Parvimonas*, *Peptostreptococcus*, and *Porphyromonas*) were assessed for differences in relative risk. We then built Random Forest models based on all genera or the select

CRC-associated genera. The models were trained on one study then tested on the remaining studies for every study. The data was split between feces and tissue samples. Within the tissue groups the data was further divided between matched and unmatched tissue samples. Both prediction for adenoma and carcinoma were tested. This same approach was then applied at the OTU level with the exception that instead of testing on the other studies a 10-fold cross validation was utilized and 100 different models were created based on random 80/20 splitting of the data to generate a range of expected AUCs. The power of each study was assessed for and effect size ranging from 1% to 30%. An estimated sample n for these effect sizes was also generated based on 80% power.

Reproducible Methods: The code and analysis can be found here https://github.com/SchlossLab/Sze_CRCMetaAnalysis_Microbiome_2017. Unless mentioned otherwise the accession number for the raw sequences for the studies used in this analysis can be found directly in the respective batch file, on the GitHub repository or in the original manuscript.

Declarations

Ethics approval and consent to participate

Need to fill in.

Consent for publication

Not applicable.

Availability of data and material

Need to fill in.

Competing Interests

All authors declare that they do not have any relevant competing interests to report.

Funding

Need to fill in.

Authors’ contributions

Need to fill in.

Acknowledgements

Need to fill in.

References

1. Keku TO, Dulal S, Deveaux A, Jovov B, Han X. The gastrointestinal microbiota and colorectal cancer. American Journal of Physiology - Gastrointestinal and Liver Physiology [Internet]. 2015 [cited 2017 Oct 30];308:G351–63. Available from: http://ajpgi.physiology.org/lookup/doi/10.1152/ajpgi.00360.2012

2. Vogtmann E, Goedert JJ. Epidemiologic studies of the human microbiome and cancer. British Journal of Cancer [Internet]. 2016 [cited 2017 Oct 30];114:237–42. Available from: http://www.nature.com/doifinder/10.1038/bjc.2015.465

3. Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, et al. Human gut microbiome and risk for colorectal cancer. Journal of the National Cancer Institute. 2013;105:1907–11.

4. Baxter NT, Ruffin MT, Rogers MAM, Schloss PD. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Medicine. 2016;8:37.

5. Brim H, Yooseph S, Zoetendal EG, Lee E, Torralbo M, Laiyemo AO, et al. Microbiome analysis of stool samples from African Americans with colon polyps. PloS One. 2013;8:e81352.

6. Burns MB, Lynch J, Starr TK, Knights D, Blekhman R. Virulence genes are a signature of the microbiome in the colorectal tumor microenvironment. Genome Medicine. 2015;7:55.

7. 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.

8. Dejea CM, Wick EC, Hechenbleikner EM, White JR, Mark Welch JL, Rossetti BJ, et al. Microbiota organization is a distinct feature of proximal colorectal cancers. Proceedings of

the National Academy of Sciences of the United States of America. 2014;111:18321–6.

9. Flemer B, Lynch DB, Brown JMR, Jeffery IB, Ryan FJ, Claesson MJ, et al. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. Gut. 2017;66:633–43.

10. Geng J, Fan H, Tang X, Zhai H, Zhang Z. Diversified pattern of the human colorectal cancer microbiome. Gut Pathogens. 2013;5:2.

11. Hale VL, Chen J, Johnson S, Harrington SC, Yab TC, Smyrk TC, et al. Shifts in the Fecal Microbiota Associated with Adenomatous Polyps. Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology. 2017;26:85–94.

12. Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, et al. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Research. 2012;22:292–8.

13. Lu Y, Chen J, Zheng J, Hu G, Wang J, Huang C, et al. Mucosal adherent bacterial dysbiosis in patients with colorectal adenomas. Scientific Reports. 2016;6:26337.

14. Sanapareddy N, Legge RM, Jovov B, McCoy A, Burcal L, Araujo-Perez F, et al. Increased rectal microbial richness is associated with the presence of colorectal adenomas in humans. The ISME journal. 2012;6:1858–68.

15. Wang T, Cai G, Qiu Y, Fei N, Zhang M, Pang X, et al. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. The ISME journal. 2012;6:320–9.

16. Weir TL, Manter DK, Sheflin AM, Barnett BA, Heuberger AL, Ryan EP. Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. PloS

One. 2013;8:e70803.

17. 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. 2014;10:766.

18. Zackular JP, Rogers MAM, Ruffin MT, Schloss PD. The human gut microbiome as a screening tool for colorectal cancer. Cancer Prevention Research (Philadelphia, Pa.). 2014;7:1112–21.

19. 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. Appl.Environ.Microbiol. [Internet]. 2009 [cited 12AD Jan 1];75:7537–41. Available from: http://aem.asm.org/cgi/content/abstract/75/23/7537

20. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: A versatile open source tool for metagenomics. PeerJ. 2016;4:e2584.

21. Westcott SL, Schloss PD. OptiClust, an Improved Method for Assigning Amplicon-Based Sequence Data to Operational Taxonomic Units. mSphere. 2017;2.

Table 1:

Table 2:

Figure 1:

Figure 2:

Figure 3:

Figure 4:

Figure S1:

Figure S2:

Figure S3: