Differential Analysis of the Metabolome and Microbiome in the Murine Colon

Cynthia Ramazani[1], Kyle Spencer[2], Haley Chatelaine[3], and Ewy Mathe[2].
[1]The Center for Genomic Advocacy at Indiana State University, Departments of [2]Biomedical Informatics, and [3]Food Science and Technology at Ohio State University

Introduction

- Recent studies have detected significant differences in microbiota profiles in the three different segments of the murine colon: the proximal colon (PC) , the medial colon (MC), and the distal colon (DC) (Xu et al., 2016) .
- Other studies have also highlighted the importance of focusing on sampling location instead of using fecal samples in animal studies on the microbiome because fecal samples are not representative of the gut microbe population and tumor location in colon cancer research.
- We examined the microbiome and metabolome composition across the three anatomical locations of the mouse colon to further examine differences of microbes and metabolites by colonic location.

Experimental Study Design

- Our analysis uses data collected by Xu et al. in their study examining the impact of various dietary interventions on the colonic microbiota of adult mice.

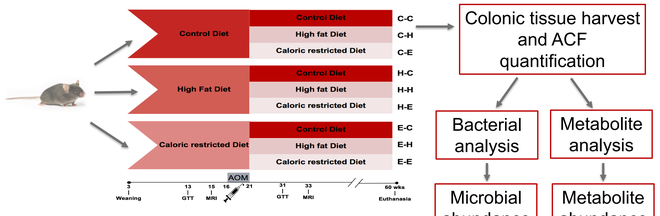

Figure 1: Study design summarizing feeding and data collection processes (Xu et a., 2016).

Objectives

1. Identification of differentially abundant metabolites across the colonic sites.
2. Identification of differentially abundant microbes across the colonic sites.
3. Functional analysis of the differential features to explain the relationship between metabolite and microbe composition and their implication in the incidence of colon cancer.

Metabolite Data Analysis

Step 1: Quality check and pre-processing

- Data deconvolution, grouping and alignment in Agilent Profinder software (B10.0).
- Manual re-integration and removal of non-peaks and peaks present in blanks.
- Filtering of metabolites based on >= 80% presence in QCs and >= 30% RSD in QCs.

Step 2: Normalization

- After undergoing a log2 transformation, the metabolite abundance data were normalized by Total Ion Current (TIC). This approach consists in multiplying metabolite abundance values for each sample by their respective pre-calculated scaling factors.

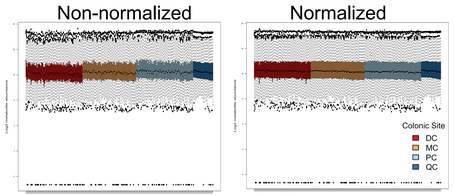

Figure 2: Distribution of the HILIC POSITIVE metabolite abundance data per sample. Colors represent colonic sites.

Step 3: Visualization

- Unscaled and non-centered principal coordinate analyses (PCA) were performed on the metabolite abundance data.
- Figures 3a and 3b show clusters by colonic location while figures 3c and 3d show clusters by order of run.

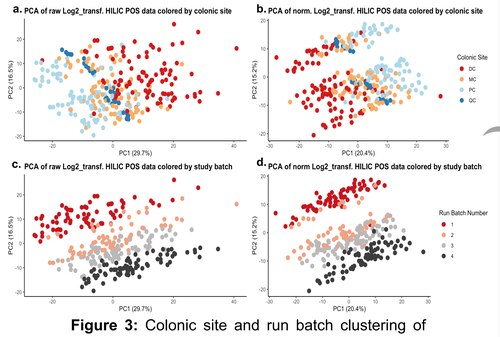

Figure 3: Colonic site and run batch clustering of metabolite abundance data showed in PCA.

Step 4: Differential abundance analysis

268 samples and 491 metabolites											
Colonic Sites			Diet Groups								
DC	MC	PC	CC	CE	CH	EC	EE	EH	HC	HE	HH
91	86	91	35	34	35	30	38	24	24	24	24
Study Batches			Run Batches								
Batch1	Batch2	Batch3	Batch1		Batch2		Batch3		Batch4		
52	107	109	60		70		68		70		

Table 1: Sample sizes for each group in HILIC POSITIVE dataset prior to differential abundance analysis.

- Since three features (DC, MC, and PC) were being examined, an analysis of variance (ANOVA) was performed. This analysis was adjusted to other covariates such as diet, study batch, and run batch.
- A subsequent Tukey post-hoc test was performed to yield pair-wise p-values.
- The results of this first analysis were also filtered by fold change.

Step 5: Results of the analysis

	MC_vs_DC	PC_vs_MC	PC_vs_DC
Before_filtering	327	282	380
After_FC_filtering	17	7	49

Table 2: This table presents the number of significantly abundant metabolites directly after the differential analysis test and after filtering by fold change.

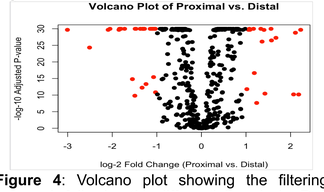

Figure 4: Volcano plot showing the filtering approach following the differential analysis. Red dots represent significantly abundant metabolites (**p-value<0.05** and **fold-change>=1**).

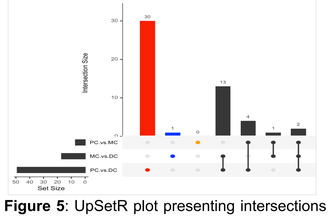

Figure 5: UpSetR plot presenting intersections in metabolites between all pairs of colonic locations statistically compared. The colored sections are the most relevant to our study.

Microbe Data Analysis

Step 1: Filtering

- The microbe data were proportional. Therefore, normalization was not performed. However, the abundance values were relatively low, hence a need for filtration.
- Genera completely absent were filtered out as well as those with 50% or more missing values in at least one of the examined colonic locations.

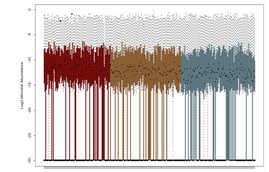

Figure 6: Distribution of the filtered microbe abundance data per sample. 62 genera passed the filtration step.

Step 2: Visualization

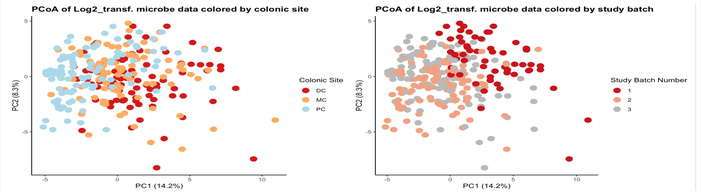

Figure 6: Colonic site and study batch clustering of microbe abundance data showed in PCoA.

Step 3: Differential abundance analysis

Step 4: Alpha diversity calculation

Conclusion & Future Directions

- We were successfully able to identify differentially abundant metabolites across the three distinct locations of the mouse colon. With the help of the NJS16 database and intLIM, we will gain more insight into the relationship between these metabolites and their associated microbes.
- When completed, the analysis of the microbe data will also provide an understanding of the specific microbiome features of each colonic site and their implication of cancer tumorigenesis.

Acknowledgements

- NIH R25-MD011712-01: Big Data for Indiana State University (BD4ISU).
- OSU collaborators: Dr. Rachel Kopec and Dr. Emmanouil Chatzakis.
- ISU mentors: Dr. Rusty Gonser, Dr. Jeff Kinne, and Dr. Jennifer Latimer.
- Members of the Mathe Lab.