**Metabolomics analysis of colon adenoma in African Americans**

Metabolomic Analysis: NIH Eastern Regional Comprehensive Metabolomics Resource Core (RTI RCMRC)

PI, RTI Collaboration: Hassan Brim.

IRB Approval Number(s): IRB-06-MED-39

# **Abstract:**

Increasing evidence suggests a role in colorectal carcinogenesis to the gut microbiota. However, no specific bacteria have been unequivocally linked to either initiation or progression of colorectal cancer (CRC). A Microbiome analysis was conducted to analyze the microbiota composition and functional potential in African Americans with colorectal lesions with the goal of detecting markers of diagnostic value.

DNA from 10 CRC tumors and their matched normal tissues as well as stool samples from 10 colon adenomas and 10 healthy subjects were analyzed for their bacterial composition and genomic content. 16S rDNA amplicons were analyzed by HITChip 16S microarray and by sequencing in stool samples and colon tissues, respectively. The functional potential was determined by sequence-based metagenomics using Illumina at a depth of 15 million reads per tissue sample to compensate for the host’s DNA presence. For the stools, the metagenomic sequencing was performed at 3 million reads per sample. Metagenomic Linkage Groups (MLGs) were established and those with high discriminative power between healthy and neoplastic specimens were analyzed for their genetic content. Also, metagenomic reads from stool samples were mapped against bacterial genes from tissues and reads from tissues were mapped against stools assembled bacterial genes to identify common markers with discriminative power.

This NMR Metabolomics analysis was performed on feces samples derived from healthy and adenoma African American subjects with the goal of identifying perturbations in metabolomics profiles in colon cancer.

**Sample Description:**

Fecal samples (20 mg) were mixed with 1000 μL of D2O, mixed thoroughly by vortexing, and then centrifuged at 16000 x g for 20 min. The supernatants were removed and filtered through 0.22μ centrifuge filters at 16000 x g for 20 minutes. 540 μl of the filtrate was mixed with 60 μl of Chenomix Internal Standard mixture (containing DSS, Imidazole, and NaN3 in D2O). Aliquots of 550 μl were then transferred into 5mm NMR tubes. All 1H NMR spectra were recorded on a Bruker Avance III 950 MHz NMR spectrometer equipped with a cryoprobe (Bruker Biospin, Rheinstetten, Germany) located at the David H. Murdoch Research Institute at Kannapolis, NC, USA. Standard 1H NMR spectra were acquired at 27°C with a standard one dimensional pulse sequence of a NOESY scheme (1dnoesypr) with water suppression using a relaxation delay of 2 s and a mixing time of 100 ms. A total of 256 transients were collected into 32768 data points for each spectrum with a spectral width of 16 ppm. Free induction decays were zero filled and multiplied by an exponential function equivalent to a 0.5 Hz line-broadening factor prior to Fourier transformation. 1H NMR spectra were manually phased and baseline-corrected by using the software package Topspin 3.0 (Bruker Biospin, Rheinstetten, Germany). The 1H NMR spectra were referenced to the DSS at δ 0.0.

The data obtained for the NMR metabolomics analysis can be found in the accompanying files:

Procedures: 1. Colon Adenoma Metabolomics Procedure.docx

Study Design Tables: 2. Colon Adenoma Metabolomics Study Design Table.xlsx

Metadata: 3. Colon Adenoma Metabolomics METADATA.xlsm

Processed Data: 4. Colon Adenoma Metabolomics Normalized Binned Data.xlsx

Raw Data: 5. Colon Adenoma Metabolomics Raw NMR Data.zip

**Notes:**

Full sample preparation and analysis procedures are available in the accompanying document entitled **1. Colon Adenoma Metabolomics Procedure.docx**.

Descriptions of abbreviations for factors are available in the Variable Dictionary in the accompanying file no. **2. Colon Adenoma Metabolomics Study Design Table.xlsx**.

The phenotypic and normalized data are available in the accompanying files: **4. Colon Adenoma Metabolomics Normalized Binned Data.xlsx** for normalized binned NMR data. Sample ID and metadata can be found in the first 7 columns and other columns in the spreadsheet contain the normalized binned data. If the statistical program does not allow variable names to begin with a number then add a prefix to the column names, for example, bin\_8.98 instead of 8.98.

The Sample ID serves as the unique identifier (Graphical ID) of the individual samples and is used as the NMR folder name in the raw NMR data file **5. Colon Adenoma Metabolomics Raw NMR Data.zip**