**Arsenic exposure and the microbiome**

De-identified feces samples were shipped to the NIH RTI-RCMRC on dry ice and immediately stored at -80 °C after being logged in for metabolomics analysis.

The samples of thawed feces from each participant were transferred to massed MagNA tubes, frozen, and lyophilized. Lyophilized samples were massed and phosphate buffer was added depending on the mass of each sample (500 µL for up to 200 mg, and 1.0 mL for over 200 mg of feces) and homogenized in 30 second pulses at 2000 rpm. Homogenized samples were centrifuged and the supernatant was filtered. A volume of the filtered supernatant necessary to analyze 25 mg per sample was mixed with 70 µL of Chenomx ISTD solution, and a calculated volume of 0.5 mM phosphate buffer (pH 7.5) needed to bring the total sample volume to 630 µL. The tubes were vortexed for 2 min on a multi-tube vortexer and centrifuged at 16,000 rcf for 5 min. A 600 µl aliquot of the supernatant was transferred into pre-labeled 5mm NMR tubes for data acquisition on a 700 MHz spectrometer. Additionally, ten random QC pools were created from the samples with adequate mass to spare. An aliquot containing 5 mg of fecal water from each individual sample was transferred into the pools, and three replicates were created from each pool for a total of 30 pooled samples. Each pool was run once a day over the 3 day experimental run time.

1H NMR spectra of feces samples were acquired on a Bruker 700 MHz NMR spectrometer (located at David H. Murdock Research Institute, NC, USA) using a 5 mm cryogenically cooled ATMA inverse probe and ambient temperature of 25 ℃. A 1D NOESY presaturation pulse sequence (noesypr1d, [recycle delay (RD)-90°-t1-90°-tm-90°-acquire free induction decay (FID) was used for data acquisition. For each sample 64 transients were collected into 65k data points using a spectral width of 12.02 ppm, 2 s relaxation delay, 100 ms mixing time, and an acquisition time of 2.726 s per FID. The water resonance was suppressed using resonance irradiation during the relaxation delay and mixing time. NMR spectra were processed using TopSpin 3.2 software (Bruker-Biospin, Germany). Spectra were zero filled, and Fourier transformed after exponential multiplication with line broadening factor of 0.5. Phase and baseline of the spectra were manually corrected for each spectrum. Spectra were referenced internally to the DSS-d6 signal. The quality of each NMR spectrum was assessed for the level of noise and alignment of identified markers. Spectra were assessed for missing data and underwent quality checks. NMR bins (0.5-9.0 ppm) were created excluding water (4.77-4.81 ppm) using intelligent bucket integration of 0.04 ppm bucket width with 50% looseness using ACD NMR Processor (ACD Labs Inc, Toronto, Canada). Integrals of each of the bins were normalized to total integral of each of the spectrum.