Supporting Information

Gut microbiome phenotypes driven by host genetics affect arsenic metabolism

Kun Lu[†,§,*], Ridwan Mahbub[§], Peter Hans Cable[#], Hongyu Ru[§], Nicola M.A. Parry[‖], Wanda M Bodnar[#], John S. Wishnok[†], Miroslav Styblo[#,¤], James A Swenberg#, James G. Fox[†,‖] and Steven R. Tannenbaum[†, Δ]

†. Department of Biological Engineering; ‖. Division of Comparative Medicine; Δ. Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, USA. §. Department of Environmental Health Science, University of Georgia, Athens, GA, 30602. #.Department of Environmental Sciences and Engineering; ¤. Department of Nutrition, University of North Carolina, Chapel Hill, NC, 27599, USA

Table of Contents

Supplemental Methods and Materials

Chemicals. Sodium arsenite and arsenobetaine (AsB) were obtained from Fisher Scientific (Pittsburgh, PA) and Sigma-Aldrich (Milwaukee, WI), respectively. Dimethylarsinic acid (DMAsV) and disodium monomethylarsonate (MMAsV) were obtained from Chem Service (West Chester, PA). MMAsIII (oxomethylarsine), DMAsIII (iododimethylarsine), dimethylmonothioarsinic acid (DMTAs) and trimethyl arsine oxide (TMAsO) were provided by Professor William Cullen, University of British Columbia (Canada). Other reagents used for the HPLC mobile phase and diluent were Puratonic® 99.999% purity grade ammonium carbonate (Alfa Aesar, Ward Hill, MA), ammonium sulfate (Mallinckrodt , Hazelwood, MO), and ammonium acetate (ICN Biochemicals, Aurora, OH).

Animals and Exposure. Specific pathogen free C57BL/6 female mice and $IL10^{-/-}$ mice (~6 weeks old) were purchased from Jackson Laboratories (Bar Harbor, ME). Mice were provided pelleted rodent diet (ProLab 3000; Purina Mills, MO) and filtered water ad libitum and were maintained in AAALAC accredited facilities in microisolator caging under standard environmental conditions. All experiments were approved by the MIT Committee on Animal Care. The animals were treated humanely and with regard for the alleviation of suffering. Each grouping was comprised of 10 mice unless otherwise stated. Inorganic arsenic (arsenic10 ppm) was administered to mice (~8 weeks old) through drinking water containing sodium arsenite for 4 weeks.

Animal monitoring and histological analysis. Throughout the experiments, mice were assessed for evidence of diarrhea, dehydration, and deteriorating body condition. Mice were euthanized with CO_2 and necropsied after 4 weeks of arsenic consumption. Formalin-fixed tissues were routinely processed, embedded in paraffin, sectioned at 4μm, stained with hematoxylin and eosin, and evaluated by a board-certified veterinary pathologist blinded to the sample identity. Inflammation, edema, epithelial defects, hyperplasia, and dysplasia of multiple regions of liver (left lateral lobe, media lobe, right lateral lobe and caudate lobe) and colon (distal, transverse and proximal colon) were scored on an ascending scale (0 to 4, with 0 being normal) of severity and invasiveness of the lesion if any.

16S rRNA gene sequencing. DNA was isolated from fecal pellets using a PowerSoil® DNA Isolation Kit as instructed by the manufacturer (MO BIO Laboratories, CA). The resultant DNA was quantified by UV spectroscopy and stored at -80^{o}C for further analysis. DNA was amplified using universal primers U515 (GTGCCAGCMGCCGCGGTAA) and E786 (GGACTACHVGGGTWTCTAAT) to target the V4 regions of 16S rRNA of bacteria. Individual samples were barcoded, pooled to construct the sequencing library, and sequenced (Illumina Miseq) to generate pair-ended 150 x 150 reads.

Analysis of 16S rRNA sequencing data. The raw mate-paired fastq files were quality-filtered, demultiplexed and analyzed using Quantitative Insights into Microbial Ecology (QIIME).[1] For quality filtering, the default parameters of QIIME were maintained in which reads with a minimum Phred quality score less than 20, containing ambiguous base calls and containing-fewer than 113bp of consecutive high quality base calls, were discarded. Additionally, reads with three consecutive low quality bases were truncated. The samples sequenced were demultiplexed using 8bp barcodes, allowing 1.5 errors in the barcode. UCLUST was used to choose the Operational

Taxonomic Units (OTUs) with a threshold of 97% sequence similarity. A representative set of sequences from each OTU was selected for taxonomic identification of each OTU using the Ribosomal Database Project (RDP) classifier.[2] The Greengenes OTUs (4feb2011 build) reference sequences (97% sequence similarity) were used as the training sequences for RDP. A 0.80 confidence threshold was used for taxonomic assignment. Our analyses were typically conducted at the *family* level due to higher confidence in the assignment of taxa based on the sequencing reads, therefore, a significant change at the *family* level may reflect changes of multiple gut bacteria at *genus* and *species* levels.

Arsenic speciation. Urine was used for arsenic speciation because the majority of arsenic was excreted in urine and major arsenic metabolites were present. Urine samples were collected using a metabolic cage with dry ice placed around the urine collection vessel to prevent oxidation or degradation of metabolites during the collection period (~16 h). Arsenic species were measured using an Agilent 7500 ICP-MS (Santa Clara, CA), as described previously.[3] The ICP-MS was interfaced with an Agilent 1260 HPLC, and a Hamilton PRP-X100 column was used to separate arsenic species. The mobile phase A was 10 mM ammonium carbonate and 10 mM TRIS (pH=8.7), and the mobile phase B included 10 mM ammonium carbonate, 10 mM TRIS and 15 mM ammonium sulfate (pH 8.0). The following gradient was run: 0 – 5 min 0% B to 100% B, 5–11 min 100% B, 11-16 min 100% A. Sample preparation was conducted as follows: 10 μL of urine sample or arsenic reference standards were diluted with 0.1 M ammonium acetate, pH=5, to make a total volume of 40 μL in a polypropylene microcentrifuge tube. Microcentrifuge tubes containing samples were centrifuged at 12,000 rpm for 5 min at 4 ℃. Following centrifugation, 35 μL of supernatant was transferred to an HPLC autosampler vial, without disturbing the solid pellet at the bottom of the tube, followed by injection of 5-10 μL into the mass spectrometer.

Statistical analysis. Multivariate statistical methods were used to compare gut microbiome communities between groups. Principal coordinate analysis (PCoA) was performed to examine intrinsic clusters within the observations. In addition, Jackknifed βdiversity and hierarchical clustering analysis via Unweighted Pair Group Method with Arithmetic Mean (UPGMA) was used to differentiate the gut microbiome profiles of the controls and knockout mice. The difference in the individual gut microbiome composition between control and knockout groups was assessed using a non-parametric test with Metastats as described previously.[4] The abundance of arsenic species in the control and knockout groups were compared by Student's *t* test, and results were considered significant at $p < 0.05$. The correlation between different arsenic metabolites was generated using Pearson's correlation coefficient. All data handling and statistical analyses were performed using the statistical R package or SAS.

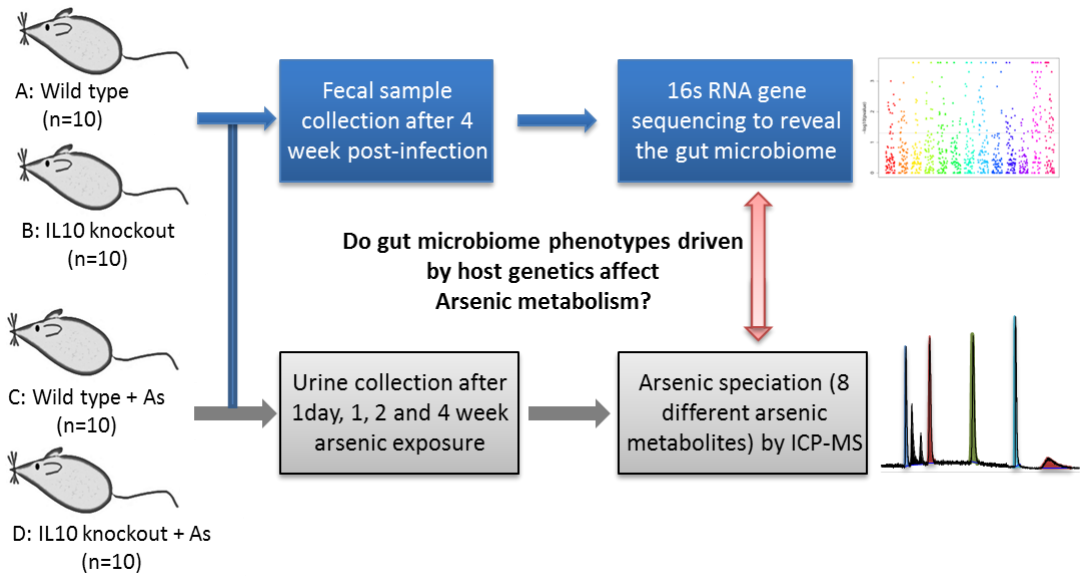

Figure S1. The integration approach combining a gene-knockout animal model, 16S rRNA gene sequencing and ICP-MS-based arsenic speciation to explore the impact of gut microbiome phenotypes driven by host genetics on the metabolism of inorganic arsenic.

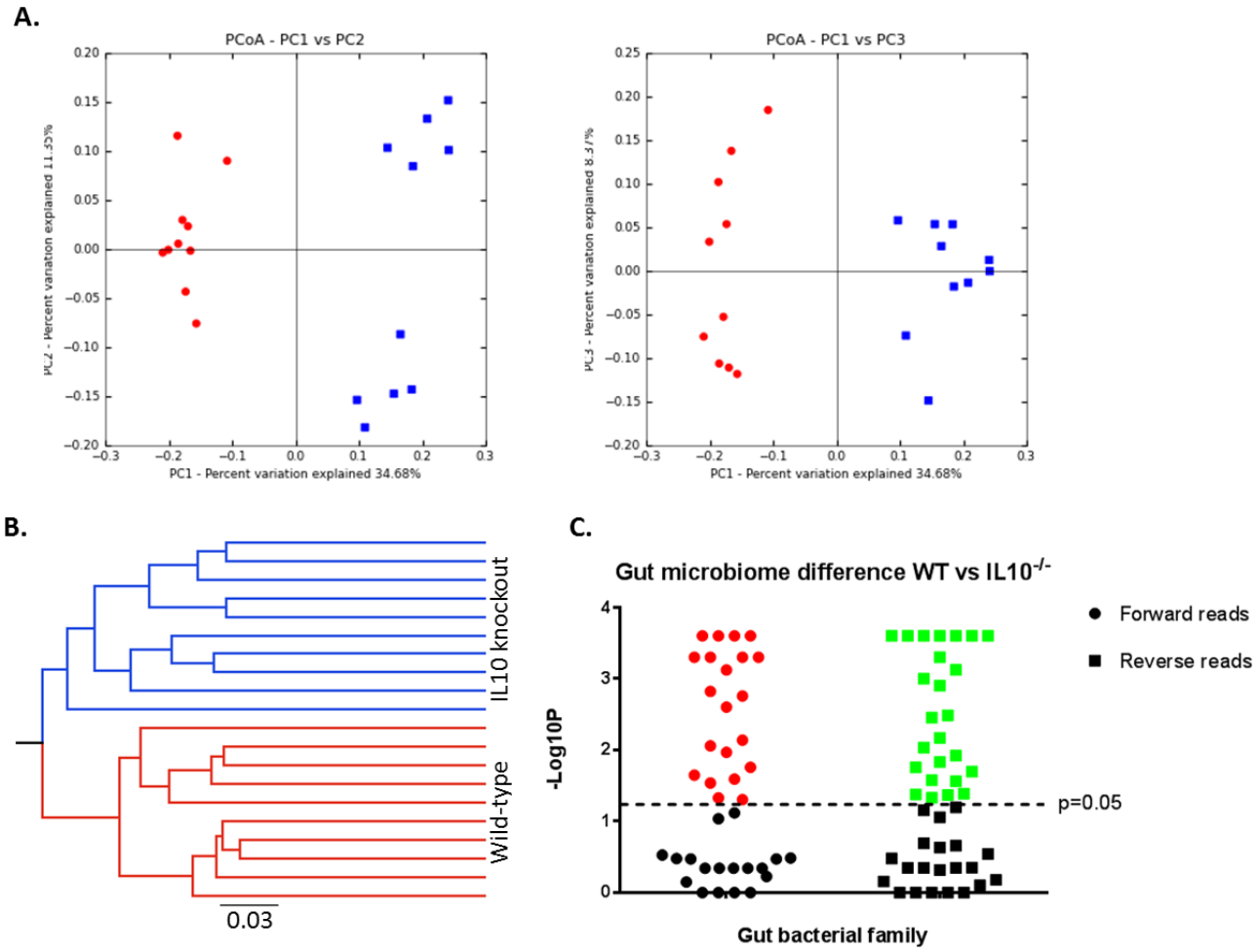

Figure S2. The gut microbiome patterns of control samples (Red) and knockout (Blue) mice are readily differentiated by Principal Coordinate Analysis (PCoA), with 34.7%, 11.4% and 8.4% variation explained by PC1, PC2 and PC3, respectively (A); Hierarchical Clustering analysis by the Unweighted Pair Group Method with Arithmetic Mean indicates that controls and arsenic-exposed mice cluster in their own groups, with the UPGMA distance tree constructed at distance of 0.03 (B); Statistically significant altered gut bacterial families compared to wild-type animals (those above the dotted line, $p<0.05$), as determined by 16S rRNA gene sequencing of forward (Red) and reverse reads before exposure to arsenic (Green) (C).

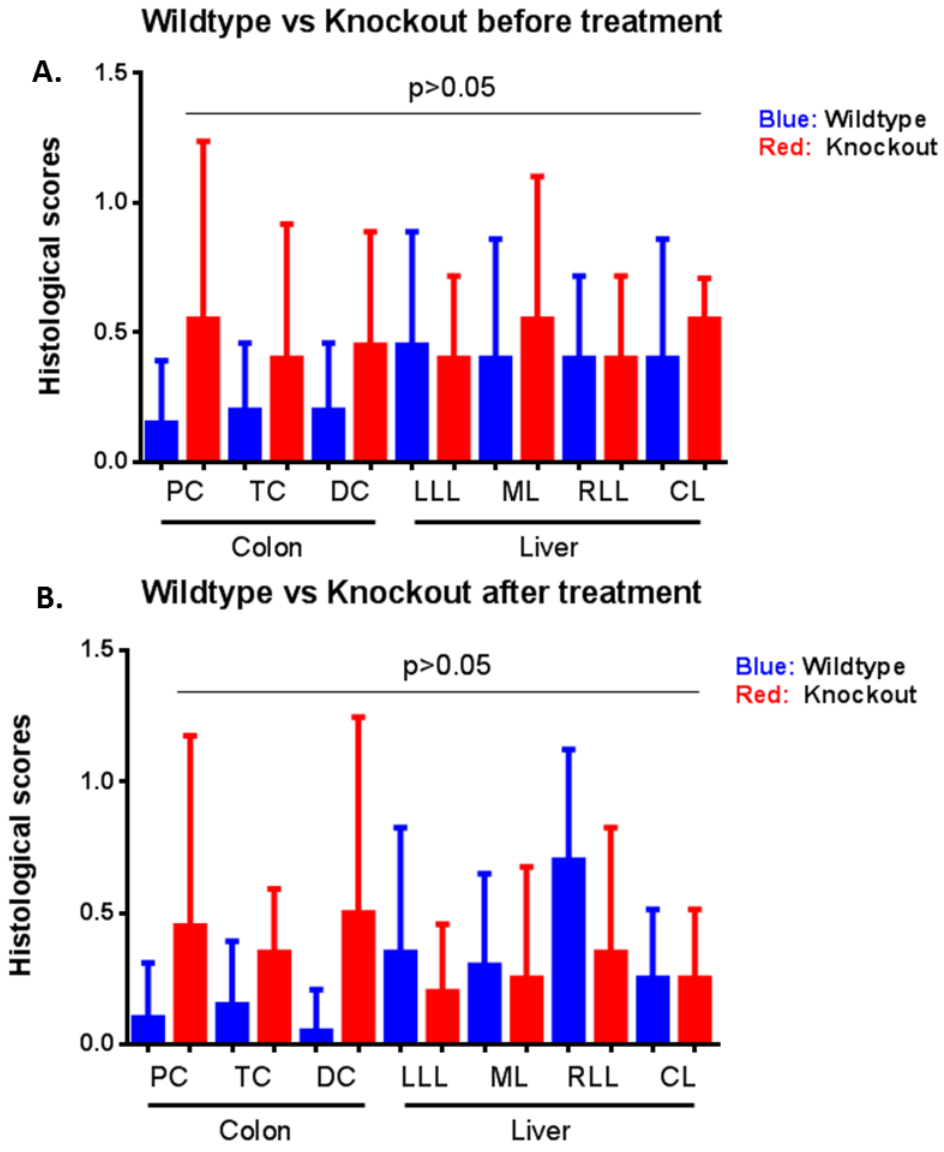

Figure S3. Histological analysis for the controls and knockout animals, with inflammation being scored for multiple regions of colon and liver before (A) and after arsenic treatment (B), with no statistically significant difference being identified between the wild-type and knockout mice. The scores for other endpoints, including edema, epithelial defects, crypt atrophy, hyperplasia and dysplasia, were also not different. (PC: Prox Colon; TC: Transverse Colon; DC: Distal Colon; LLL: Left Lateral Lobe; ML: Medial Lobe; RLL: Right Lateral Lobe; CL: Caudate Lobe)

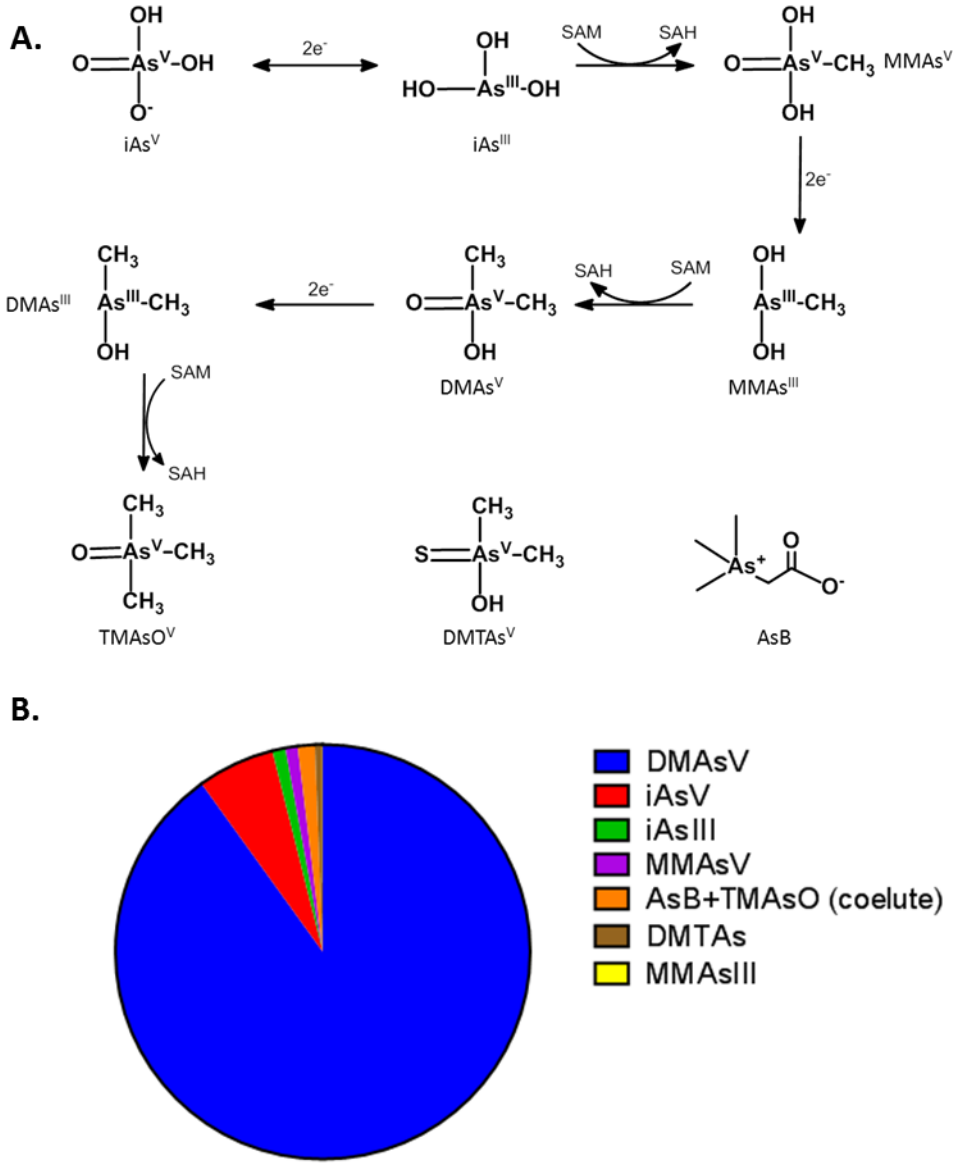

Figure S4. The metabolism pathway and major arsenic species (A), and the relative abundance of different arsenic species in urine of wild-type mice after normalization using the total counts of arsenic at *m/z* = 75, with DMAsV and iAsV being the top 2 arsenic metabolites (B), as determined by HPLC-ICP-MS speciation analysis.

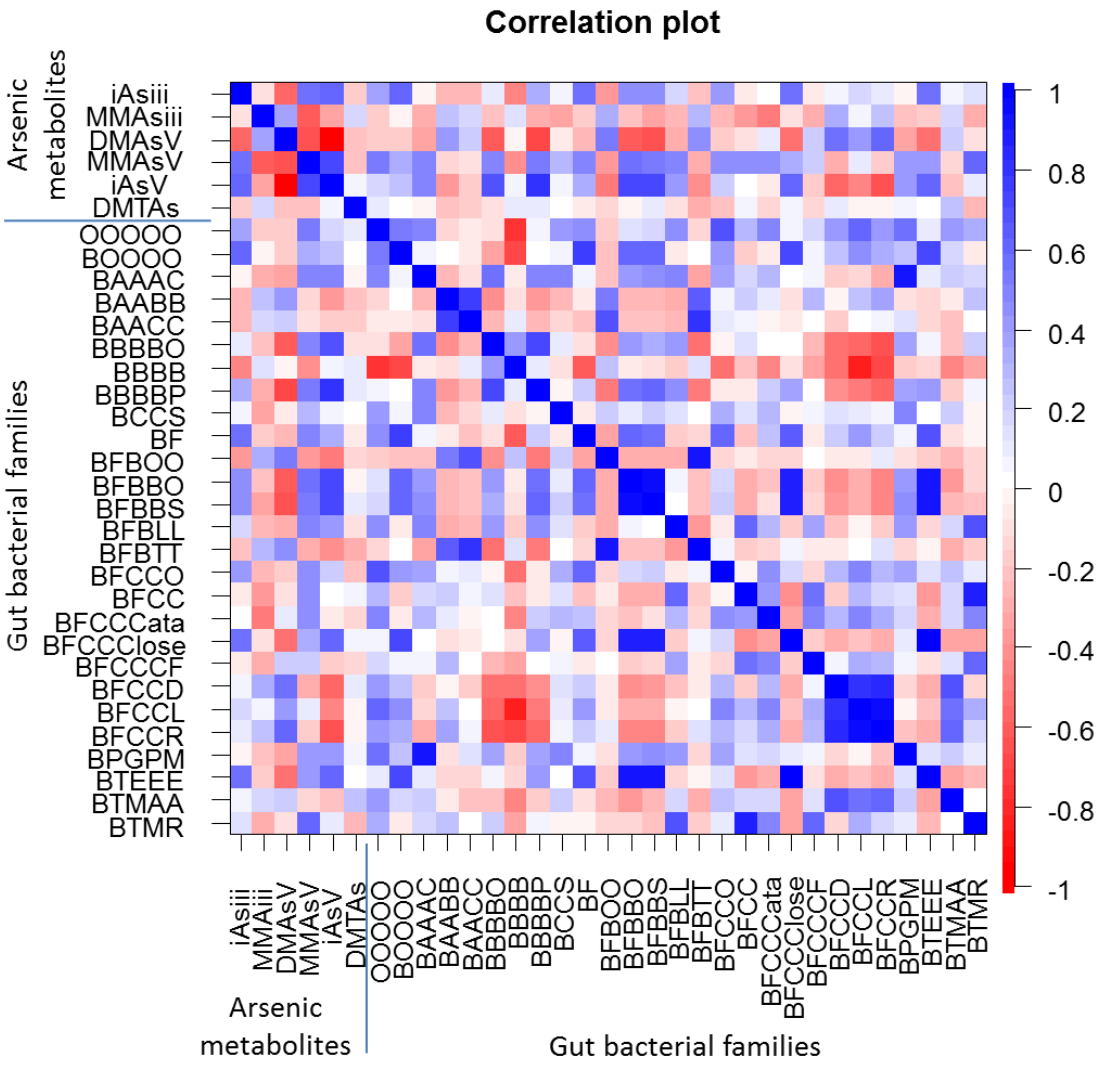

Figure S5. The correlation plot, calculated by Pearson's correlation coefficient using the relative abundance of gut bacterial families and arsenic species, demonstrates the functional correlation between perturbed gut bacteria families* and altered arsenic metabolites.

*: The abbreviations of bacterial families are listed as below:
Other;Other;Other;Other;Other: OOOOO;
k_Bacteria;Other;Other;Other;Other: BOOOO;
k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;Other: BAAAO;
k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Bifidobacteriales;Other: BAABO;
k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Bifidobacteriales;f_Bifidobacteriaceae: BAABB;
k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Coriobacteriales;f_Coriobacteriaceae:BAACC;
k_Bacteria;p_Bacteroidetes;Other;Other;Other:BBOOO;
k_Bacteria;p_Bacteroidetes; c_Bacteroidia;o_Bacteroidales;Other: BBBBO;
k_Bacteria;p_Bacteroidetes;c_Bacteroidia; o_Bacteroidales;f_unassigned: BBBB;
k_Bacteria;p_Bacteroidetes;c_Bacteroidia; o_Bacteroidales;f_Porphyromonadaceae: BBBBP;

k_Bacteria;p_Cyanobacteria;c_Chloroplast; o_Streptophyta;f_unassigned:BCCS;
k_Bacteria;p_Firmicutes;Other;Other;Other:BFOOO;
k_Bacteria;p_Firmicutes;c_Bacilli;Other;Other: BFBOO; k_Bacteria;p_Firmicutes;c_Bacilli;
o_Bacillales;Other: BFBBO; k_Bacteria;p_Firmicutes;c_Bacilli;o_Bacillales;
f_Staphylococcaceae: BFBBS;
k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;Other: BFBLO;
k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Lactobacillaceae: BFBLL;
k_Bacteria;p_Firmicutes;c_Bacilli;o_Turicibacterales;f_Turicibacteraceae: BFBTT;
k_Bacteria;p_Firmicutes;c_Clostridia;Other;Other: BFCOO;
k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;Other: BFCCO;
k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_unassigned:BFCC;
k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Catabacteriaceae:BFCCCata;
k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae: BFCClost;
k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiales Family XIII. Incertae Sedis:
BFCCF
k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Dehalobacteriaceae: BFCCD;
k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae: BFCCL;
k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Peptostreptococcaceae: BFCCP;
k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae:BFCCR;
k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Pseudomonadales;f_Moraxellaceae:
BPGPM;
k_Bacteria;p_Tenericutes;Other;Other;Other: BTOOO;
k_Bacteria;p_Tenericutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae:BTEEE;
k_Bacteria;p_Tenericutes;c_Mollicutes;Other;Other: BTMOO;
k_Bacteria;p_Tenericutes;c_Mollicutes;o_Anaeroplasmatales;f_Anaeroplasmataceae:BTMAA;
k_Bacteria;p_Tenericutes;c_Mollicutes;o_RF39;f_unassigned:BTMR

Table S1. Significantly different gut microbiome families between the wild-type and knockout mice ($p < 0.05$) and their taxonomic assignments based on 16s rRNA sequencing.

Taxa	Mean of wild-type	Mean of knockout	Fold change	p-value
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__unassigned	0.397214	0.665640	1.68	0.0092
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Lachnospiraceae	0.242294	0.101760	0.42	0.0032
k__Bacteria;Other;Other;Other;Other	0.079377	0.049484	0.62	0.0420
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;Other	0.071839	0.046130	0.64	0.0275
k__Bacteria;p__Verrucomicrobia;c__Verrucomicrobiae;o__Verrucomicrobiales;f__Verrucomicrobiaceae	0.000000	0.031896	-	0.0002
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Porphyromonadaceae	0.000011	0.030330	2812	0.0200
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae	0.039426	0.015592	0.40	0.0005
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;Other	0.000398	0.008307	20.87	0.0002
k__Bacteria;p__Tenericutes;c__Mollicutes;o__Anaeroplasmatales;f__Anaeroplasmataceae	0.020037	0.004790	0.24	0.0007
k__Bacteria;p__Firmicutes;c__Clostridia;Other;Other	0.015939	0.003188	0.20	0.0002
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiaceae	0.021130	0.003174	0.15	0.0010
k__Bacteria;p__Firmicutes;c__Bacilli;o__Turicibacterales;f__Turicibacteraceae	0.032489	0.002100	0.06	0.0002
k__Bacteria;p__Tenericutes;c__Erysipelotrichi;o__Erysipelotrichales;f__Erysipelotrichaceae	0.006661	0.001823	0.27	0.0465
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Catabacteriaceae	0.005962	0.001222	0.21	0.0120
k__Bacteria;p__Tenericutes;c__Mollicutes;o__RF39;f__unassigned	0.002427	0.001160	0.48	0.0067
k__Bacteria;p__Bacteroidetes;Other;Other;Other	0.000270	0.000707	2.62	0.0035
k__Bacteria;p__Verrucomicrobia;Other;Other;Other	0.000000	0.000256	-	0.0002
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Coriobacteriales;Other	0.000000	0.000056	-	0.0002
k__Bacteria;p__Tenericutes;Other;Other;Other	0.000198	0.000042	0.21	0.0175
k__Bacteria;p__Firmicutes;c__Bacilli;Other;Other	0.000160	0.000041	0.25	0.0427
k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta;f__unassigned	0.000217	0.000026	0.12	0.0012
Other;Other;Other;Other;Other	0.000085	0.000022	0.26	0.0147
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Peptostreptococcaceae	0.000039	0.000004	0.11	0.0265
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales;f__Bifidobacteriaceae	0.000754	0.000000	0.00	0.0002
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Clostridiales Family XIII. Incertae Sedis	0.000026	0.000000	0.00	0.0411

Reference

(1) Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight, R. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods 7*, 335-336.

(2) Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol. 73*, 5261-5267.

(3) Verdon, C. P., Caldwell, K. L., Fresquez, M. R., and Jones, R. L. (2009) Determination of seven arsenic compounds in urine by HPLC-ICP-DRC-MS: a CDC population biomonitoring method. *Anal. Bioanal. Chem. 393*, 939-947.

(4) White, J. R., Nagarajan, N., and Pop, M. (2009) Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS. Comput. Biol. 5*, e1000352.