Cell Reports, Volume *22*

Supplemental Information

Diet, Genetics, and the Gut Microbiome Drive Dynamic Changes in Plasma Metabolites

Shiho Fujisaka, Julian Avila-Pacheco, Marion Soto, Aleksandar Kostic, Jonathan M. Dreyfuss, Hui Pan, Siegfried Ussar, Emrah Altindis, Ning Li, Lynn Bry, Clary B. Clish, and C. Ronald Kahn

Fujisaka Supplemental Methods

Mouse procedures

Male C57BL/6J mice (B6) and 129S1 mice (129J) were purchased from Jackson Laboratory (Bar Harbor, ME) and 129S6 mice (129T) were purchased from Taconic Farms (Germantown, NY). Mice were maintained on normal chow containing 22% calories from fat, 23% from protein and 55% from carbohydrates (Mouse Diet 9F 5020, PharmaServ, Framingham, MA) or a high fat diet (Open Source Diet, D12492, Research Diets, New Brunswick, NJ) containing 60% calories from fat, 20% from protein and 20% from carbohydrates. For antibiotic treatment, 6-week old mice were treated with either placebo, vancomycin (1g/L) or metronidazole (1g/L) (Sigma-Aldrich, St. Louis, MO) in drinking water then started on HFD from age 7 to 11 weeks. The mice were fasted for 2 hours and anesthetized with isoflurane before collecting cecum and plasma. Insulin resistance scores were based on the observation that B6J on HFD were the most insulin resistant (score = 4); this was improved by metronidazole (score = 3) and improved even more by vancomycin (score = 2); that chow-fed B6J mice were even more insulin sensitive (score = 1); and that 129 mice regardless of diet or antibiotics were the most insulin sensitive (score = 0).

16S rRNA sequence analysis

DNA was extracted from mouse cecum contents using a MoBio Fecal DNA extraction kit (MoBio Laboratories Inc., Carlsbad, CA). A multiplexed amplicon library covering the 16S rDNA gene V4 region was generated from DNA extracted samples. Reads were generated on the MiSeq instrument from the amplicon library and clustered into Operational Taxonomic Units (OTUs). A total of 1,353,060 sequence reads were generated, corresponding to an average of 27,331.5 (range 16,657 to 69,861) reads per sample. Differences in microbial community structure were visualized using phylogenetic methods. The number of OTUs per sample were then scaled so each sample had the same mean, filtered to only include OTUs that were present at 0.1% of the total counts in at least 3 samples, log-transformed (using $\log_2(\text{count}+0.5)$), and plotted in PCA space using the R software. 16S rRNA datasets have been deposited in Sequence Read Archive (SRA) database (accession number: WILL BE PROVIDED AT TIME OF PUBLICATION).

Untargeted metabolomic analysis

Metabolomic analyses of plasma samples. Four separate liquid chromatography tandem mass spectrometry (LC-MS) methods were used to measure polar metabolites and lipids in each sample. *Method 1*. Positive ion mode MS analyses of polar metabolites were conducted using a Nexera X2 U-

HPLC system (Shimadzu Scientific Instruments; Marlborough, MA) coupled to an Exactive Plus orbitrap mass spectrometer (Thermo Fisher Scientific; Waltham, MA). LC-MS samples were prepared from plasma (10 µL) via protein precipitation with the addition of nine volumes of 74.9:24.9:0.2 v/v/v acetonitrile/methanol/formic acid containing stable isotope-labeled internal standards (valine-d8, Isotec; and phenylalanine-d8, Cambridge Isotope Laboratories; Andover, MA). The samples are centrifuged (10 min, 9,000 x g, 4°C), and the supernatants were injected directly onto a 150 x 2 mm Atlantis HILIC column (Waters; Milford, MA). The column was eluted isocratically at a flow rate of 250 µL/min with 5% mobile phase A (10 mM ammonium formate and 0.1% formic acid in water) for 1 minute followed by a linear gradient to 40% mobile phase B (acetonitrile with 0.1% formic acid) over 10 minutes. MS analyses were carried out using electrospray ionization in the positive ion mode using full scan analysis over m/z 70-800 at 70,000 resolution and 3 Hz data acquisition rate. Additional MS settings were: ion spray voltage, 3.5 kV; capillary temperature, 350°C; probe heater temperature, 300 °C; sheath gas, 40; auxiliary gas, 15; and S-lens RF level 40.

Method 2. Negative ion mode, targeted MS analyses of polar metabolites we conducted using an ACQUITY UPLC (Waters) coupled to a 5500 QTRAP triple quadrupole mass spectrometer (AB SCIEX). Plasma samples (30µL) were extracted using 120 µL of 80% methanol (VWR) containing 0.05 ng/µL inosine-15N4, 0.05 ng/µL thymine-d4, and 0.1 ng/µL glycocholate-d4 as internal standards (Cambridge Isotope Laboratories). The samples were centrifuged (10 min, 9,000 x g, 4ºC) and the supernatants (10 µL) were injected directly onto a 150 x 2.0 mm Luna NH2 column (Phenomenex) that was eluted at a flow rate of 400 µL/min with initial conditions of 10% mobile phase A (20 mM ammonium acetate and 20 mM ammonium hydroxide (Sigma-Aldrich) in water (VWR)) and 90% mobile phase B (10 mM ammonium hydroxide in 75:25 v/v acetonitrile/methanol (VWR)) followed by a 10 min linear gradient to 100% mobile phase A. The ion spray voltage was -4.5 kV and the source temperature was 500°C.

Method 3. Negative ion mode analysis of metabolites of intermediate polarity (e.g. bile acids and free fatty acids) were analyzed using a Nexera X2 U-HPLC system (Shimadzu Scientific Instruments; Marlborough, MA) coupled to a Q Exactive orbitrap mass spectrometer (Thermo Fisher Scientific; Waltham, MA). Plasma samples (30 µL) were extracted using 90 µL of methanol containing PGE2-d4 as an internal standard (Cayman Chemical Co.; Ann Arbor, MI) and centrifuged (10 min, 9,000 x g, 4°C). The supernatants (10 µL) were injected onto a 150 x 2.1 mm ACQUITY BEH C18 column (Waters; Milford, MA). The column was eluted isocratically at a flow rate of 450 µL/min with 20% mobile phase A (0.01% formic acid in water) for 3 minutes followed by a linear gradient to 100% mobile phase B (0.01% acetic acid in acetonitrile) over 12 minutes. MS analyses were carried out using electrospray ionization in the negative ion mode using full scan analysis over m/z 70-850 at 70,000 resolution and 3 Hz data acquisition rate. Additional MS settings were: ion spray voltage, -3.5 kV; capillary temperature,

320°C; probe heater temperature, 300 °C; sheath gas, 45; auxiliary gas, 10; and S-lens RF level 60. *Method 4*. Polar and nonpolar lipids were analyzed using a Nexera X2 U-HPLC system (Shimadzu Scientific Instruments; Marlborough, MA) coupled to an Exactive Plus orbitrap mass spectrometer (Thermo Fisher Scientific; Waltham, MA). Lipids were extracted from plasma (10 µL) using 190 µL of isopropanol containing 1,2-didodecanoyl-sn-glycero-3-phosphocholine as an internal standard (Avanti Polar Lipids; Alabaster, AL). After centrifugation (10 min, 9,000 x g, ambient temperature), supernatants (10 µL) were injected directly onto a 100 x 2.1 mm ACQUITY BEH C8 column (1.7 µm; Waters; Milford, MA). The column was eluted at a flow rate of 450 µL/min isocratically for 1 minute at 80% mobile phase A (95:5:0.1 vol/vol/vol 10 mM ammonium acetate/methanol/acetic acid), followed by a linear gradient to 80% mobile-phase B (99.9:0.1 vol/vol methanol/acetic acid) over 2 minutes, a linear gradient to 100% mobile phase B over 7 minutes, and then 3 minutes at 100% mobile-phase B. MS analyses were carried out using electrospray ionization in the positive ion mode using full scan analysis over m/z 200-1100 at 70,000 resolution and 3 Hz data acquisition rate. Additional MS settings were: ion spray voltage, 3.0 kV; capillary temperature, 300°C; probe heater temperature, 300 °C; sheath gas, 50; auxiliary gas, 15; and S-lens RF level 60.

Raw data from orbitrap mass spectrometers were processed using Progenesis QI software (NonLinear Dynamics) for feature alignment, untargeted signal detection, and signal integration. Targeted processing of a subset of known metabolites was conducted using TraceFinder software (version 3.1, Thermo Fisher Scientific; Waltham, MA). Raw data from the 5500 QTRAP MS system were processed using MultiQuant 2.1 software (AB SCIEX). Compound identities were confirmed using reference standards and reference samples.

Metabolomic data analysis

Metabolomics data was analyzed in the R software. To preprocess the metabolomics abundances, missing values were imputed by half the minimum observed for that metabolite, the data was quantile normalized using the preprocess Core package and then $\log_2$-transformed, metabolites that were present in only two or fewer samples were filtered out, and one sample that had very low initial abundance and was deemed to be of poor quality by an unbiased quality weighting algorithm (Ritchie et al., 2006) was discarded. Between group comparisons were analyzed with the linear modeling package limma accounting for the data's mean-variance trend (Law et al., 2014; Ritchie et al., 2015). Cecum and plasma metabolite abundances were tested for association using Pearson correlation. Plasma metabolite abundances were then tested for association to 16S levels using Spearman rank correlation. Tests were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate (FDR). Heatmaps were plotted with the gplots package, and boxplots, PCA plots, and scatterplots were plotted with the ggplot2

package.

Statistical analysis

Statistical significance was evaluated using ANOVA and a post Tukey-Kramer test. A p value less than 0.05 was considered significant. The results were presented as the means ± SEM.

Study approval

All experiments complied with regulations and ethics guidelines and were approved by the IACUC of Joslin Diabetes Center (97-05) and Harvard Medical School (05131).

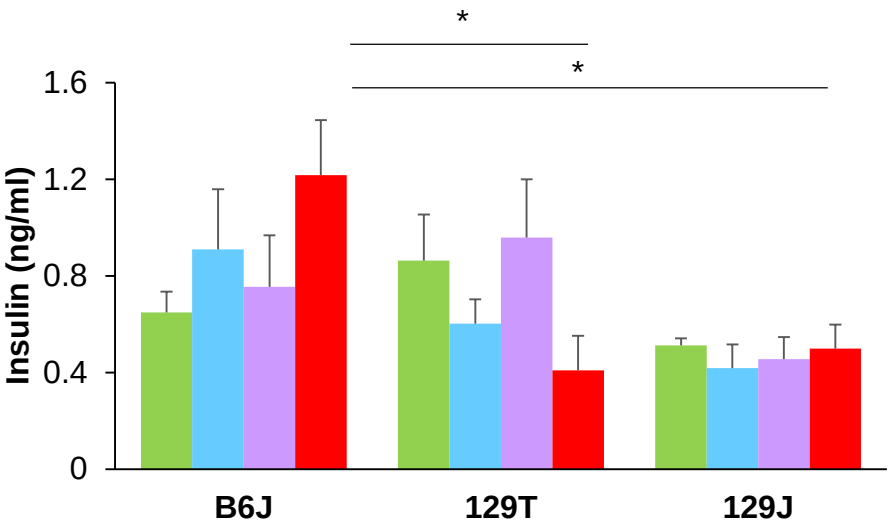
References

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biology *15*, R29-R29.

Ritchie, M.E., Diyagama, D., Neilson, J., van Laar, R., Dobrovic, A., Holloway, A., and Smyth, G.K. (2006). Empirical array quality weights in the analysis of microarray data. BMC Bioinformatics *7*, 261-261.

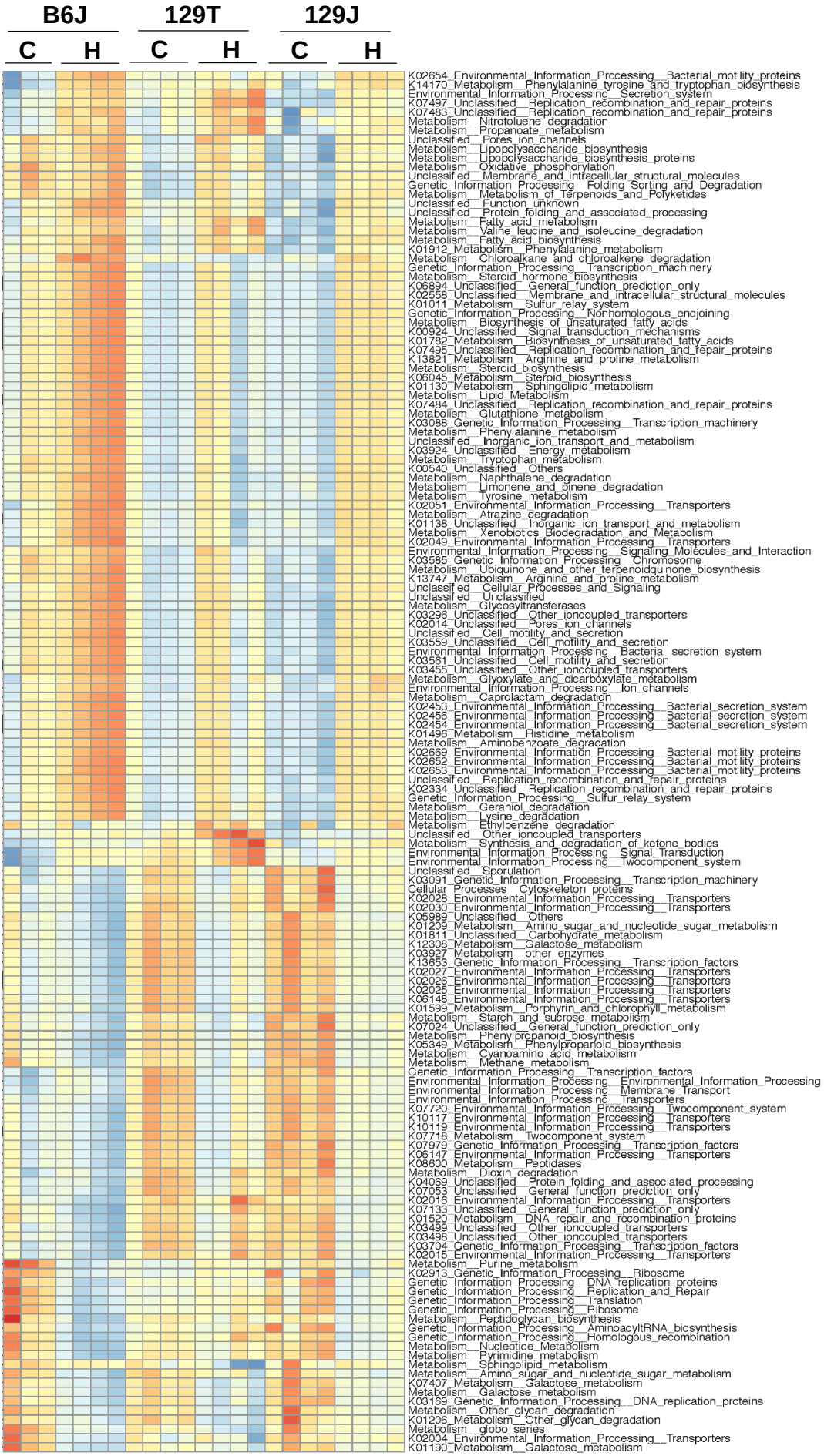
Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research *43*, e47-e47.

Figure S1, Insulin levels in mice in response to HFD and antibiotic Treatment, Related to Figure 1



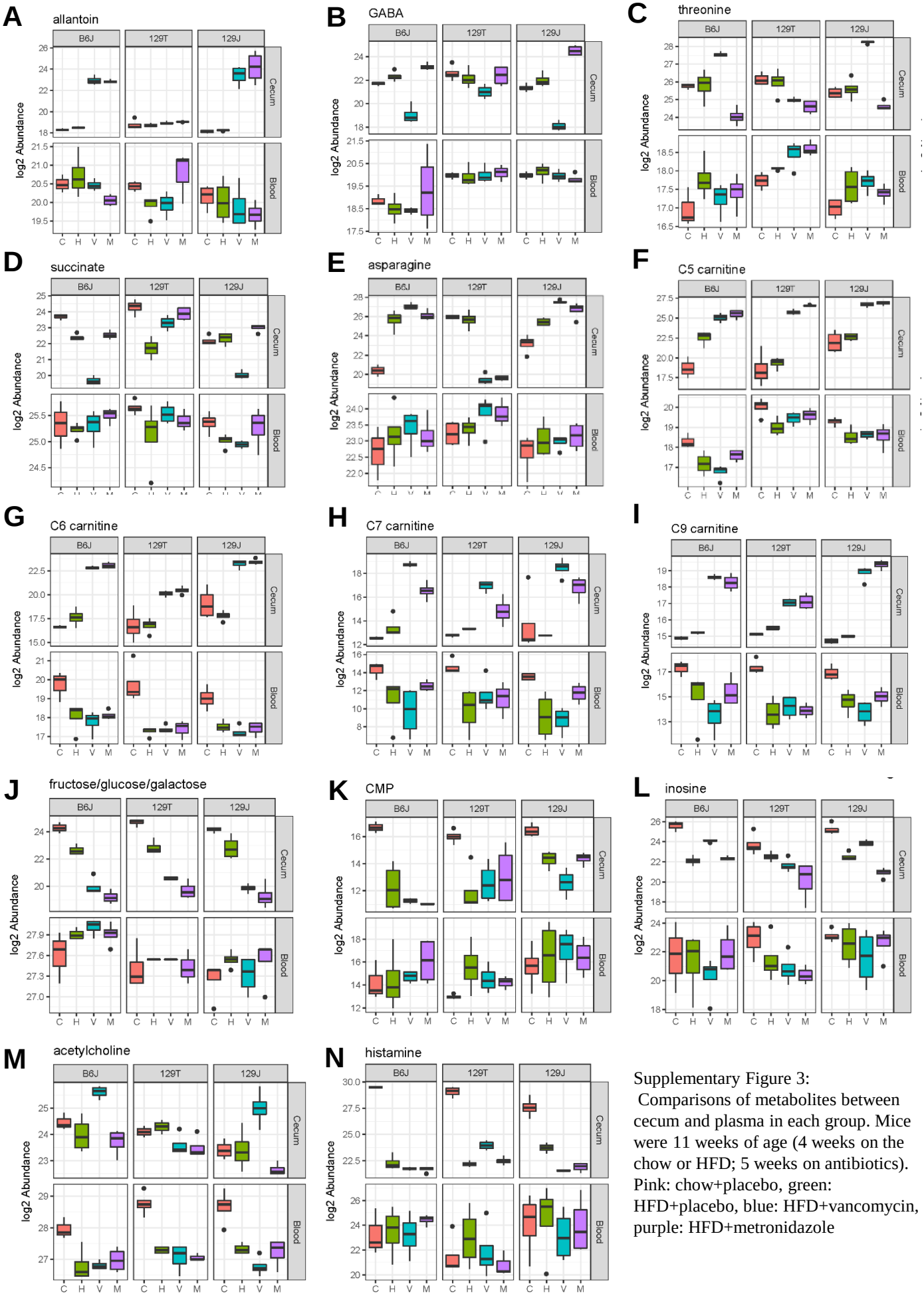
Plasma insulin levels of mice on chow (green), HFD (blue), HFD + vancomycin (purple) and HFD + metronidazole (red) in the random fed state at 11 weeks of age. (n = 3-4 /group). *P<0.05 by ANOVA, followed by Tukey-Kramer post-hoc. The results are shown as the mean ± SEM.

Figure S2, Heatmap of Metabolic Pathway Comparing Chow and HFD, Related to Figure 2



Heatmap of all significantly altered metabolic pathways obtained from PICRUSt analysis of 16S rRNA sequencing data from Chow (C) versus HFD+placebo (H).

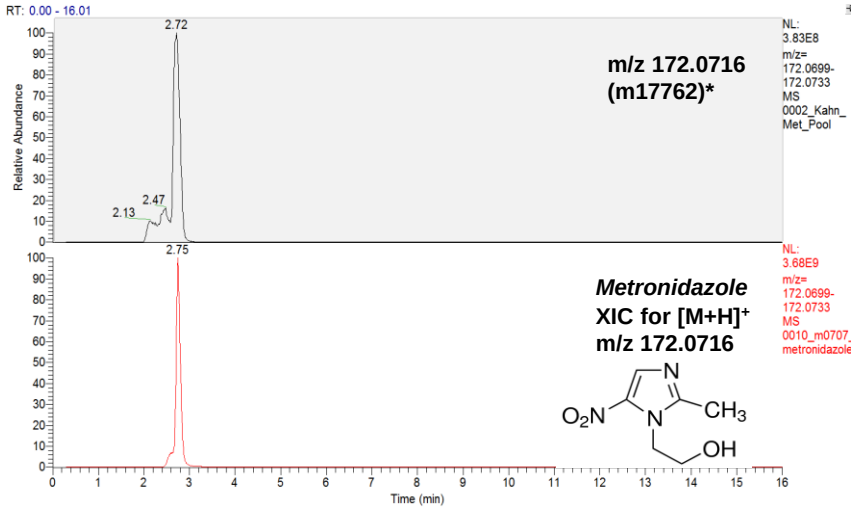
Figure S3, Comparison of Metabolites in Cecum and Plasma, Related to Figure 4

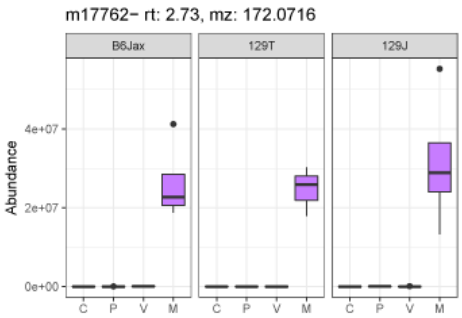


Supplementary Figure 3:
Comparisons of metabolites between cecum and plasma in each group. Mice were 11 weeks of age (4 weeks on the chow or HFD; 5 weeks on antibiotics). Pink: chow+placebo, green: HFD+placebo, blue: HFD+vancomycin, purple: HFD+metronidazole

Figure S4, Identifying Unknown Features in the Mass Spec Analysis of Metabolites, Related to Figure 7

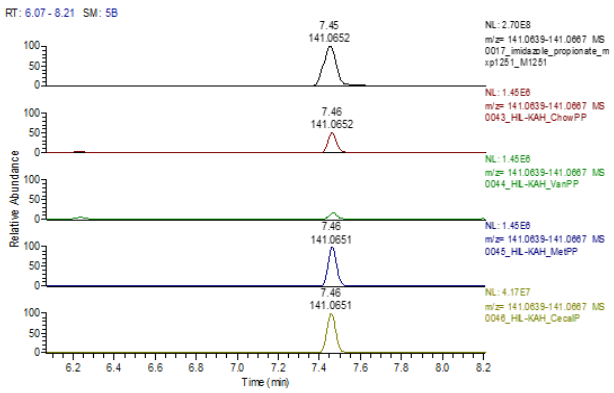
A



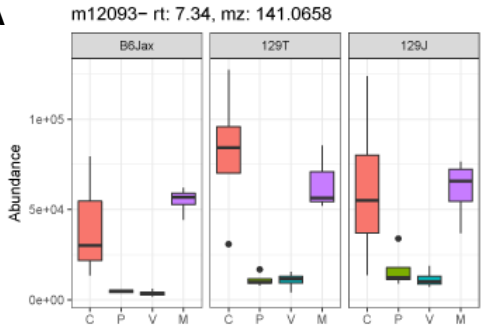


B

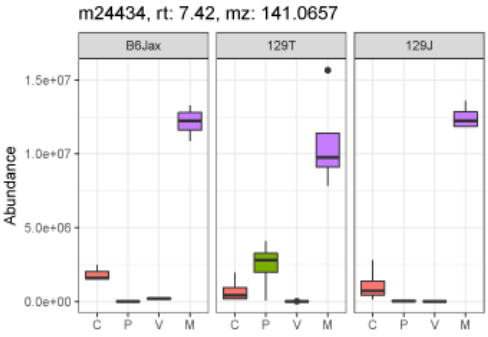
1) Imidazole propionate standard
2) Chow Pool Plasma
3) Vancomycin Pool Plasma
4) Metronidazole Pool Plasma
5) Cecal Pool Plasma



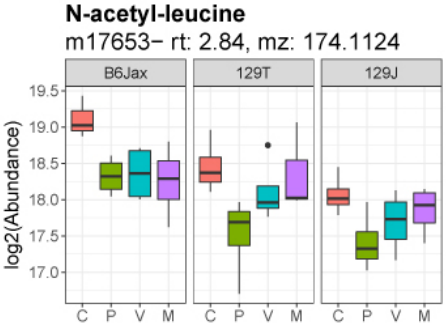
PLASMA

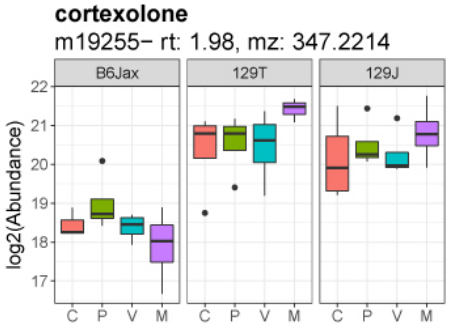


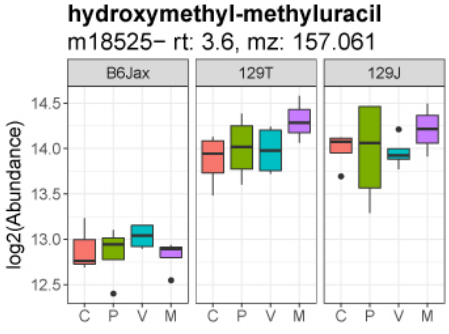
CECAL



C







Supplementary Figure 4:
(A) Representative chromatograms of metronidazole with the raw abundance in plasma samples. (B) Representative chromatograms of imidazole propionate with the raw abundance in plasma and cecum. (C) Plasma levels of newly identified metabolites.

Table S1, Overview of Mass Spec Analysis of Metabolites, Related to Figure 3

PLASMA	Features	Annotated Compounds	Unannotated Compounds
C18-neg	41	41 (41 Features)	none
C8-pos*	6,244	157 (779 Features)	3,679 (5,465)
HILIC-neg	80	80 (80 Features)	none
HILIC-pos*	13,262	96 (446 Features)	7,364 (12,816)
TOTAL	19,627	374 (1,225 Features)	11,043 (18,281 Features)

CECAL	Features	Annotated Compounds	Unannotated Compounds
C18-neg	10,716	39 (283 Features)	6,206 (10,433)
C8-pos*	11,387	254 (679 Features)	5,520 (10,708)
HILIC-neg	71	0 (Features)	none
HILIC-pos*	27,538	118 (629 Feaures)	12,870 (26,909)
TOTAL	49,712	491 (1,591 Features)	24,596 (48,050 Features)

Supplementary Table 1:
The number of compounds detected and the features representing them were estimated as follows: For each individual untargeted method (*), the Pearson correlation of features co-eluting within a 0.03 retention time window was calculated and clusters trimmed based on a minimum correlation coefficient of 0.8. Ions with the highest mean abundance were selected as representative within each cluster.