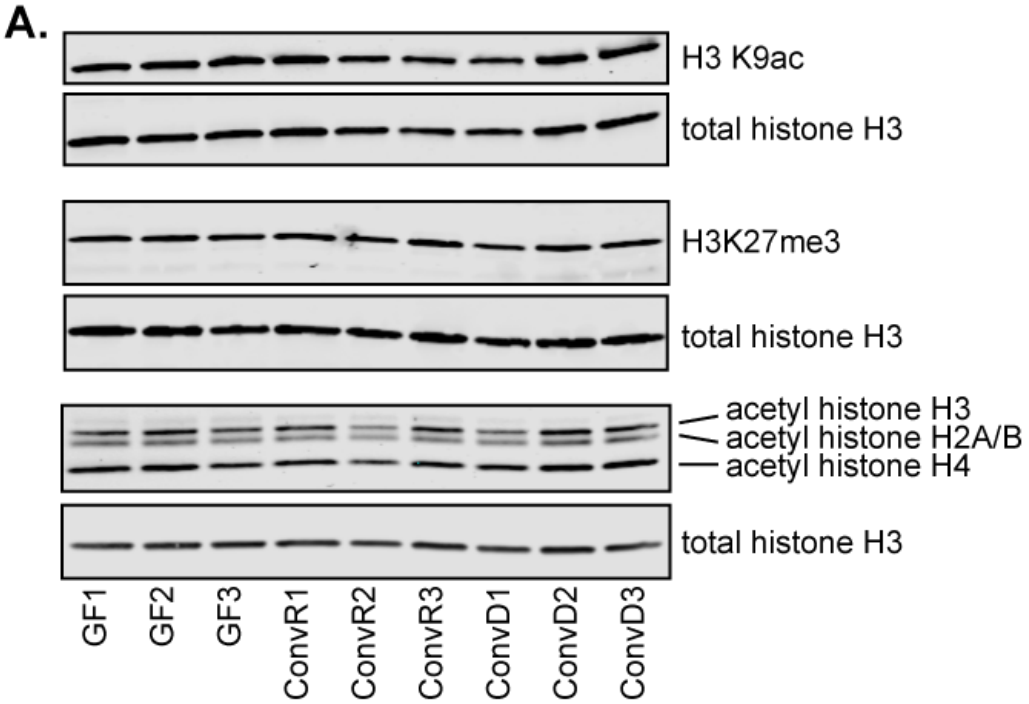
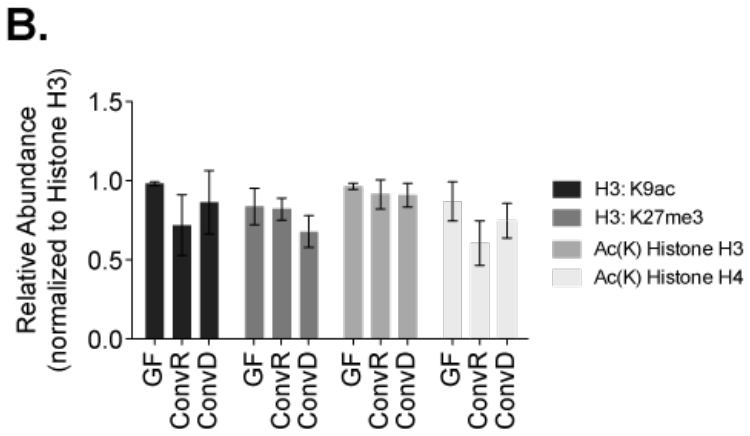
Figure S1, Related to Figure 1





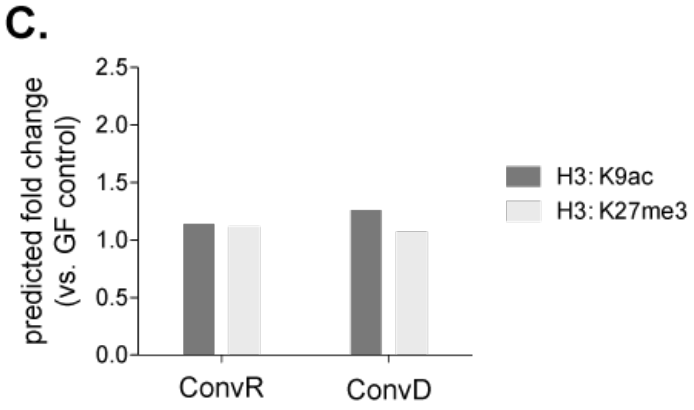


Figure S2, Related to Figure 2

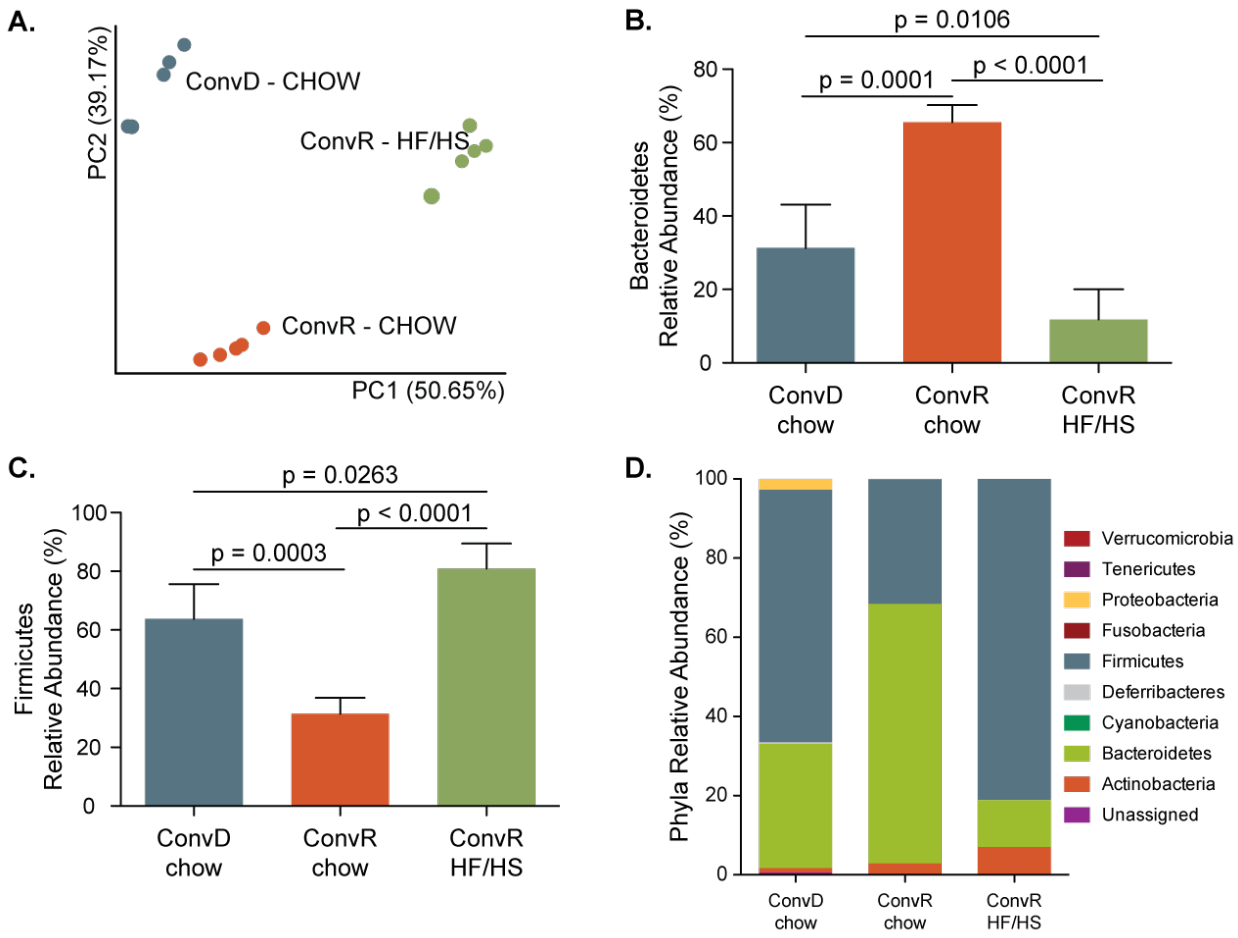
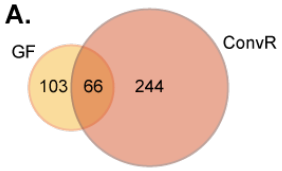
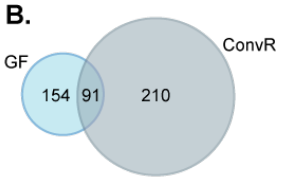


Figure S3, Related to Figure 3

A.



B.



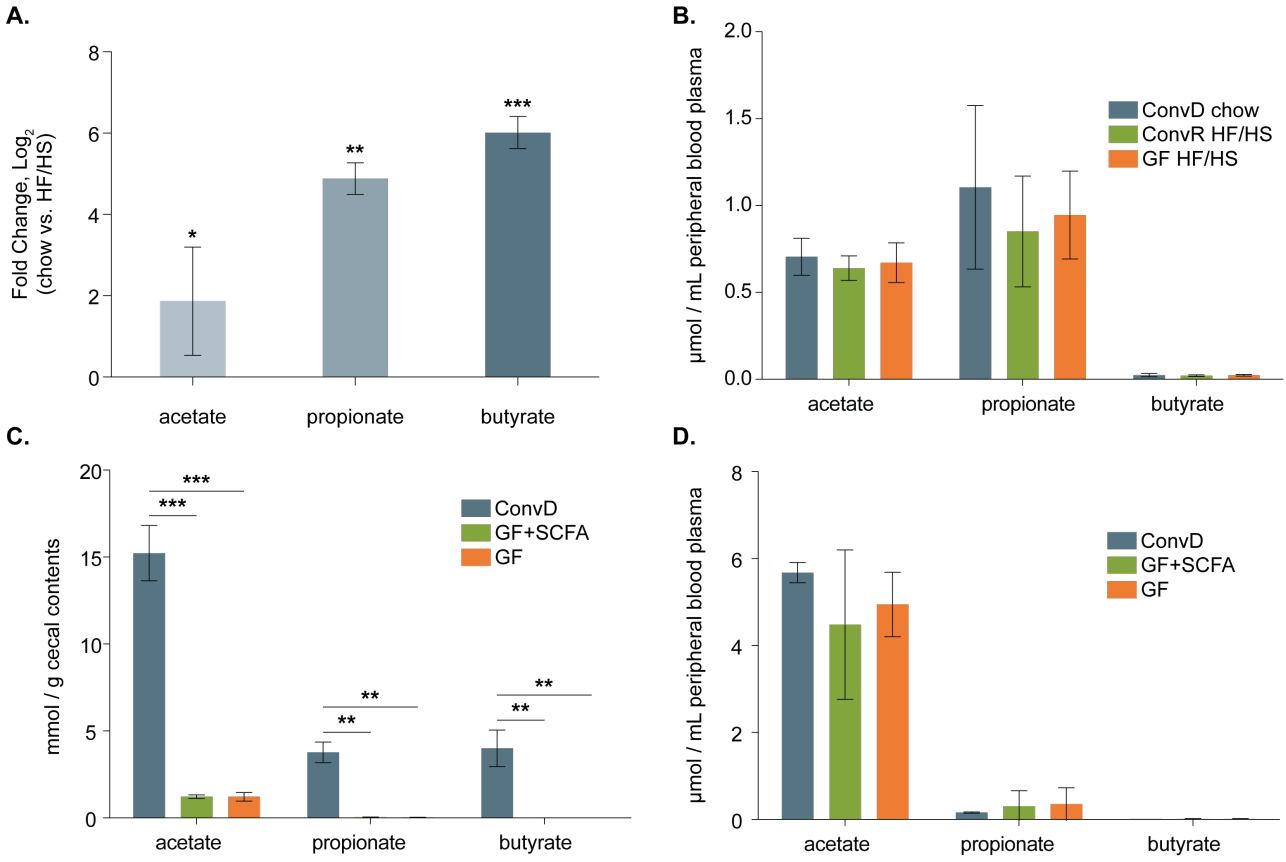
C.

KEGG pathway	p-value
Cytokine-cytokine receptor interaction	1.52E-08
PPAR signaling pathway	2.96E-08
Biosynthesis of unsaturated fatty acids	4.56E-06
Pyruvate metabolism	1.60E-05
Lysosome	3.13E-04
Endocytosis	1.81E-03
Retinol metabolism	3.22E-03
Viral myocarditis	3.33E-03
Adipocytokine signaling pathway	3.45E-03
Pathways in cancer	6.91E-03
Fatty acid biosynthesis	7.33E-03
Chemokine signaling pathway	2.02E-02
Focal adhesion	2.22E-02
Citrate cycle (TCA cycle)	3.47E-02
Propanoate metabolism	3.61E-02
Aldosterone-regulated sodium reabsorption	4.18E-02
Steroid hormone biosynthesis	4.60E-02
PPAR signaling pathway	2.17E-07
Glycolysis / Gluconeogenesis	7.57E-06
Adipocytokine signaling pathway	1.84E-05
Starch and sucrose metabolism	3.09E-04
Type II diabetes mellitus	4.97E-04
Neurotrophin signaling pathway	3.73E-03
Insulin signaling pathway	4.15E-03
Maturity onset diabetes of the young	1.29E-02
Pentose phosphate pathway	1.54E-02
Biosynthesis of unsaturated fatty acids	1.88E-02
Pyruvate metabolism	2.81E-02
Fructose and mannose metabolism	2.89E-02
Amino sugar and nucleotide sugar metabolism	3.39E-02
Retinol metabolism	4.80E-02
MAPK signaling pathway	2.20E-09
Arachidonic acid metabolism	3.96E-06
Fatty acid metabolism	2.47E-05
Pathways in cancer	7.57E-05
PPAR signaling pathway	1.37E-04
Vascular smooth muscle contraction	4.24E-04
Lysosome	4.39E-04
Tight junction	6.91E-04
Insulin signaling pathway	8.35E-04
Purine metabolism	1.53E-03
Sphingolipid metabolism	1.54E-03
Glycolysis / Gluconeogenesis	2.42E-03
Long-term depression	3.36E-03
p53 signaling pathway	3.62E-03
VEGF signaling pathway	4.31E-03
ErbB signaling pathway	6.52E-03
Fc gamma R-mediated phagocytosis	7.78E-03
Ubiquitin mediated proteolysis	1.58E-02
Biosynthesis of unsaturated fatty acids	3.57E-02
Butanoate metabolism	3.88E-02
Starch and sucrose metabolism	4.83E-02

D.

KEGG pathway	p-value
Glutathione metabolism	1.59E-09
Metabolism of xenobiotics by cytochrome P450	6.10E-09
Drug metabolism	1.06E-08
Pentose and glucuronate interconversions	5.07E-07
Tryptophan metabolism	5.05E-04
Arachidonic acid metabolism	2.14E-03
Huntington's disease	1.06E-02
Ascorbate and aldarate metabolism	1.48E-02
Porphyrin and chlorophyll metabolism	2.63E-02
Starch and sucrose metabolism	3.15E-02
Ribosome	2.37E-20
Steroid biosynthesis	2.98E-07
Metabolism of xenobiotics by cytochrome P450	3.68E-07
Drug metabolism	5.71E-07
Retinol metabolism	2.33E-05
Linoleic acid metabolism	4.97E-04
Systemic lupus erythematosus	1.70E-03
Arachidonic acid metabolism	1.76E-03
Complement and coagulation cascades	1.98E-03
Terpenoid backbone biosynthesis	1.15E-02
Glutathione metabolism	4.43E-02
Ribosome	1.40E-22
Jak-STAT signaling pathway	4.41E-08
Cytokine-cytokine receptor interaction	5.77E-08
Complement and coagulation cascades	6.80E-06
MAPK signaling pathway	5.44E-04
TGF-beta signaling pathway	8.13E-03
Hematopoietic cell lineage	8.55E-03
Pyrimidine metabolism	1.05E-02
Steroid biosynthesis	2.46E-02
Purine metabolism	2.48E-02
Chemokine signaling pathway	2.86E-02
Endocytosis	3.41E-02
Regulation of actin cytoskeleton	3.61E-02

Figure S4, Related to Figure 4



Supplemental Figure Legends:

Figure S1, related to Figure 1: Histone PTM abundance via western blot. (A) Western blot images for histone H3 K9ac, H3 K27me3, and pan-acetyl(K) of histone extracts from GF, ConvR, and ConvD mouse livers. (B) Relative quantitation of western blots. (C) Predicted fold change (*vs.* GF) based on mass spectrometry data.

Figure S2, related to Figure 2: Differences in the fecal microbial communities among experimental groups. (A) Principal Coordinates Analysis of weighted UniFrac distances between the fecal bacterial communities from conventionally raised (ConvR) animals fed chow or high fat/high sugar diet (HF/HS) and Conventionalized animals (ConvD) consuming chow. Each circle represents an individual mouse. Percent variation explained by each principal coordinate (PC) is shown in parentheses. (B-D) Relative abundance of microbial phyla identified in experimental groups. (B) Relative abundance of Bacteroidetes, (C) Firmicutes, and (D) detected dominant phyla. n=5 mice/experimental group.

Figure S3, related to Figure 3: Regulation of hepatic transcription as a function of diet in GF and ConvR mice. Overlap of hepatic DE genes that significantly (A) increase and (B) decrease in HF/HS fed mice *vs.* chow-fed controls. Numbers in each portion of the venn diagram represent the number of corresponding DE genes in each category.(C-D) KEGG pathway enrichment in (C) DE up genes, and (D) DE down genes. Orange = DE up in GF only, orange-red = DE up in both GF and ConvR, red = DE up in ConvR only. Blue = DE down in GF only, blue-grey = DE down in both GF and ConvR, grey = DE down in ConvR only. n=3 mice per group, significance was determined by an FDR cutoff = 0.05.

Figure S4, related to Figure 4: SCFA measurement in peripheral blood and cecal contents. (A) Fold change (log_2) in SCFA levels of cecal contents in ConvR mice on HF/HS vs. chow diets. n = 4 mice per dietary condition. (B) Peripheral venous blood (plasma) levels of SCFAs in 19 wk old ConvD mice fed a chow diet, and ConvR and GF mice fed a HF/HS diet. n = 5 mice per group. (C-D) Cecal content and plasma levels of SCFAs in 14wk old ConvD, GF+SCFA, or GF mice. n = 4 mice per group. Statistical significance was determined by Welch's T-test. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table S1, related to Figures 1-2: Fold change in PTM states
Table S2, related to Figure 3: RNA-seq analyses on livers of colonized (both ConvR and ConvD) and GF mice on two diets with accompanying K-means clustering
Table S3, related to Figure 3: Effects of diet on GF and colonized mice gene expression
Table S4, related to Figure 4: RNA-seq analyses with clustered data comparing the effects of SCFAs and colonization, and GO-terms analysis

Additional acknowledgments
We thank the members of the Epigenetics Theme at the Wisconsin Institute for Discovery for their support and expertise, in particular the Rupa Sridharan group (Rupa Sridharan and Khoa Tran) for their guidance with RNAseq work and input throughout the duration of the project. We thank Marie Adams and the staff at the University of Wisconsin Biotechnology Center DNA Sequencing Facility for their assistance in acquiring RNAseq data. We also thank the University of Wisconsin - Madison Molecular Archaeology Group for providing access to their GC/MS instrument and expertise in performing GC/MS data acquisition. Finally, we thank the Center for High Throughput

Computing (CHTC) staff, with particular gratitude to Lauren Michael, Christina Koch, and Neil Van Lysel, for their invaluable guidance while using HTCondor. We apologize to authors whose work was able to be cited given the strict length constrains of a short article.

Supplemental Methods and Information

Tissue fractionation, histone extraction, and label-free chemical derivatization – Nuclear fractions were prepared from flash-frozen whole post-mortem tissues. In brief, 50-100mg of tissue was dounce homogenized on ice in hypotonic lysis buffer (10mM Tris-HCl, 10mM NaCl, 3mM $MgCl_2$, and 0.1% NP-40 alternative for adipose tissue only) with histone deacetylase and protease inhibitors (1mM sodium butyrate, 4 M trichostatin A, 100 M phenylmethylsolfonyl fluoride, 10 g/mL leupeptin, and 10 g/mL aprotinin). Crude sub-cellular fractions were separated by differential centrifugation at 800 x g for 10 minutes at 4C. The crude nuclear pellet was then washed twice with ice-cold PBS prior to acid extraction. Histones were acid extracted as described previously (Krautkramer et al., 2015). Histone extract (5 g) was then subjected to a hybrid chemical derivatization method (Maile et al., 2015) using propionic anhydride to label protein N-termini and all unmodified and monomethylated lysine residues, followed by trypsinization for 6 hours at 37C, and derivatization of newly generated peptide N-termini with phenylisocyanate (PIC). This chemical derivatization method protects all lysine residues from cleavage by trypsin, which cleaves C-terminally to unmodified lysine and arginine residues, enabling consistent generation of histone peptides amenable to analysis by mass spectrometry. Derivatized histone extracts were then desalted using in-house C18 stage tips.

Nano-liquid chromatography and electrospray ionization tandem mass spectrometry – For each sample, 0.3 g of derivatized histone peptides was injected onto a Dionex Ultimate3000 nanoflow HPLC with a Waters nanoAcquity UPLC C18 column (100 m x 150 mm, 3m) coupled to a Thermo Fisher Q-Exactive mass spectrometer at 700 nL/min. Mobile phase consisted of water + 0.1% formic acid (A) and acetonitrile + 0.1% formic acid (B). Histone peptides were resolved with a 2-step linear gradient of 2% to 25% mobile phase B over 60 minutes followed by 25% to 40% mobile phase B over 15 minutes. Data was acquired using data-independent acquisition (DIA) mode. The mass spectrometer was operated with a MS1 scan at resolution = 35,000, automatic gain control target = 1×10^6, and scan range = 390-910 m/z, followed by a DIA scan with a loop count of 10. DIA settings were as follows: window size = 10 m/z, resolution = 17,500, automatic gain control target = 1×10^6, DIA maximum fill time = AUTO, and normalized collision energy = 30. For each cycle, one full MS1 scan was followed by 10 MS2 scans using an isolation window size of 10 m/z.

Histone PTM quantification – The Thermo .raw files were imported into Skyline for quantitative analysis (MacLean et al., 2010). All MS1 and MS2 peaks were matched to a spectral library upon import based on retention times and the presence of appropriate transitions. All imported histone peptide MS1 peaks and their integration bounds were also manually verified using XCalibur Qual Browser (v2.2). Peak areas for all selected transitions were exported and precursor ion peak areas were combined for quantitation. Three precursor ions were used for quantitation and a minimum of 4 MS2 ions were used to validate peak picking. This ensured unbiased, accurate measurement. The spectral library, which can be found in the supplemental information, is a curated in-house library prepared from Mascot search results from numerous data-dependent acquisition (DDA) injections of histone peptides that were aggregated using the redundant library feature of Skyline. Briefly, peak lists were generated by converting Thermo .raw files from DDA experiments to .mgf files for Mascot database search using MSConvert, a tool in the ProteoWizard library (Kessner et al., 2008). Mascot parameters were as follows: fixed modifications = N-terminal PIC and lysine propionylation; variable modifications = lysine acetylation, lysine butyrylation (an equivalent mass shift to monomethylation + propionylation), lysine dimethylation, and lysine trimethylation; peptide mass tolerance = 10 ppm; fragment mass tolerance = 0.01 Da. The enzyme was set to ArgC, since trypsin will only cleave after arginine due to chemical derivatization, and a maximum of one missed cleavage was allowed. A probability based Mowse score of 26 or higher indicated identify with $p < 0.05$.

To quantify the percent of total for each peptide species, all peptide areas belonging to a peptide "family" were summed to obtain the total area for that "family." A peptide "family" is defined as a group of peptides spanning the same residues within a histone protein, e.g. H3 residues 18-26 (KQLATKAAR), which contains lysines 18 and 23 is a family of peptides with 6 members: K18ac+K23un, K18un+K23ac, K18ac+K23ac, K18me1+K23un, K18un +K23me1, and K18un+K23un. The proportion of the total for each peptide member of the family was then obtained by dividing the individual peptide area by the summed family area. Isobaric and co-eluting histone H3 and H4

peptides were not deconvoluted, and are denoted as such (eg: K18ac+K23un and K18un+K23ac are isobaric and co-eluting and are denoated as a single value for K18ac/K23ac since the MS1 ions for these two species are identical and thus are representative of both peptide species). Western blot-like estimates of histone H3 K9ac and K27me3 abundance were calculated from mass spectrometry data using the following formulas: K9ac PTM abundance = %K9acK14un + %K9acK14ac; K27me3 PTM abundance = %K27me3K36un + %K27me3K36me1 + %K27me3K36me2. Fold changes were then calculated as the PTM abundance in ConvR and ConvD mice vs. GF.

Histone PTM data normalization and statistics – Exported peak area values from Skyline were normalized within peptide families to the total area prior to calculation of fold changes and statistics. All p-values were generated using a Welch's t-test (biological replicates, n = 4). Statistical significance was determined by $p < 0.05$.

Gene expression profiling – RNA was extracted from flash-frozen post-mortem livers using the Qiagen RNeasy Mini Kit (#74104) and quantified using a Nanodrop 2000 spectrophotometer. A single-end cDNA library was prepared using the Illumina TruSeq stranded mRNA Sample Preparation kit (RS-122-2101) according to the manufacturer's specifications. Samples were multiplexed, normalized, and pooled prior to sequencing on a HiSeq2500 instrument at the University of Wisconsin Biotechnology Center DNA Sequencing Facility. Experiments were performed in biological triplicate.

RNAseq processing and analysis – Greater than 21 million 100bp reads were sequenced per sample. Reads were processed and filtered using the FastX Toolkit. The first 10 bp of each read were removed to eliminate GC bias at the beginning of each read using the FastX Trimmer. The 6-bp Illumina adapter sequences were then clipped using the FastX Clipper, and a quality filter was applied to all reads to eliminate any reads with a quality score <30 as determined by the FastX Quality Filter. Reads were then aligned to the mm9 mouse reference genome constructed to include only annotated genes (NM_RefSeqs) by Bowtie2 alignment. Seed length was set to 28 and mismatches per seed was set to 2. Gene expression was then calculated using RSEM (Li and Dewey, 2011) with a forward probability of 0.0 (for stranded libraries). Differential expression (DE) was calculated using EBSeq (Leng et al., 2013) with false discovery rate (FDR) = 0.05. EBSeq uses the empirical Bayes method and expectation maximization to compute normalized fold changes and determine differential expression. This model assigns a posterior probability of the differential expression status for each individual gene and uses an FDR threshold to reduce false discoveries. Data have been deposited in the Gene Expression Omnibus (GEO) database (Barrett et al., 2013), accession codes GSE81115 and GSE81117.

K-means clustering was performed using the MatLab function kmeans and silhouette indices were used to determine the optimal K. Cluster enrichment for functional annotation was determined using DAVID Bioinformatics Resources v6.7 (Dennis et al., 2003). Interaction networks were constructed using stringApp (http://www.cgl.ucsf.edu/cytoscape/stringApp/index.shtml), which queries string-db v10.0 (Jensen et al., 2009) and returns a network, within Cytoscape v3.3.0 (Shannon et al., 2003).

Gas chromatography – mass spectrometry measurement of SCFAs – Flash-frozen cecal contents (100mg) or plasma (50 L) were mixed with 20 L internal standards (acetic-d4 acid, Sigma-Aldrich #233315; propionic-3,3,3-d3 acid, CDN isotopes #D-80; and butyric-d7 acid, CDN isotopes #D-171) and acidified with 20 L 33% HCl. Two rounds of extraction using 1mL diethyl ether were carried out by mixing for 10 minutes at room temperature following by centrifugation at 1932 x g for 10 minutes at 4C. Extracts (60 L) were then incubated at room temperature for 2 hours with 2 L N-tert-Butyldimethylsilyl-N-methyltrifluoroacetamide (MTBSTFA, Sigma-Aldrich #394882). Derivatized samples (1 L) were injected onto an Agilent 7890B/5977A GC/MSD instrument with an Agilent DB-1 0.25mm x 60m column with 0.25µm bonded phase. A discontinuous oven program was used starting at 40°C for 2.25 min, then ramping at 20°C/min to 200°C, then ramping at 100°C/min to 300°C and holding for 7 min. The total run time was 18.25 minutes. Linear column flow was maintained at 1.26mL/min. The inlet temperature was set to 250C with an injection split ratio of 15:1. Quantitation was performed using selected ion monitoring (SIM) acquisition mode and metabolites were compared to relevant labeled internal standards using Agilent Mass Hunter v. Acquisition B. 07.02.1938. The m/z of monitored ions are as follows: 117 (acetic acid), 120 (acetic-d4 acid), 131 (propionic acid), 134 (propionic-3,3,3-d3 acid), 145 (butyric acid), and 152 (butyric-d7 acid). The dwell time for each monitored m/z was 50ms, with a total cycle rate of 3Hz. Concentrations were normalized to mg of cecal contents and statistical significance was determined by $p < 0.05$ using a Welch's T-test.

DNA Extraction and 16S Library Preparation – Genomic DNA was extracted from feces using a bead-beating protocol (Turnbaugh et al., 2009). Briefly, mouse fecal pellets (~50 mg) or cecal contents were re-suspended in a solution containing 500µl of extraction buffer [200mM Tris (pH 8.0), 200mM NACL, 20mM EDTA], 210µl of 20% SDS, 500µl phenol:chloroform:isoamyl alcohol (pH 7.9, 25:24:1) and 500µl of 0.1-mm diameter zirconia/silica

beads. Cells were mechanically disrupted using a bead beater (BioSpec Products, Barlesville, OK; maximum setting for 3 min at room temperature), followed by extraction with phenol:chloroform:isoamyl alcohol and precipitation with isopropanol. NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel, Bethlehem, PA) was used to remove contaminants. Isloated DNA was stored at -80°C until further processing.

Amplification 16S rRNA genes (V4) was done from DNA by PCR using unique 8-bp barcodes on the forward and reverse primers and fused with Illumina sequencing adapters (Kozich et al., 2013). Each sample was amplified in a reaction volume of 25µl using KAPA HiFi HotStart DNA polymerase (KAPA Biosystems, Wilmington, MA), 10µM of each primer and ~25ng of genomic DNA. PCR was carried out under the following conditions: initial denaturation for 3 min at 95°C, followed by 25 cycles of denaturation for 30 s at 95°C, annealing for 30 s at 55°C and elongation for 30 s at 72°C, and a final elongation step for 5 min at 72°C. PCR products were purified with the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, Bethlehem, PA) and then quantified using Qubit dsDNA HS Assay kit (Invitrogen, Oregon, USA). Samples were pooled and sequenced on the Illumina MiSeq 2x250bp platform.

Microbiota Analysis in QIIME – Demultiplexing of 16S rRNA gene sequences, quality control and operational taxonomic unit (OTU) binning were performed using Quantitative Insights Into Microbial Ecology (QIIME) version 1.9.1 (Caporaso et al., 2010b). Quality filtered reads were trimmed of Illumina adaptor and index sequences. Sequences were then clustered in OTUs using a de novo OTU picking protocol based on 97% identity using UCLUST (Edgar, 2010) against the Greengenes reference database (McDonald et al., 2012). Representative sequences (most abundant sequence in OTUs) were picked, aligned to GreenGenes Core reference alignment (DeSantis et al., 2006) using PyNAST (Caporaso et al., 2010a). Taxonomic assignments were associated with OTUs based on the taxonomy associated with the Greengenes reference sequence defining each OTU. UniFrac distances between samples were calculated using the Greengenes reference tree. Greengenes reference sequences, trees and taxonomy data used in the analysis can be found at: http://greengenes.secondgenome.com/downloads/database/13_5

The resulting biom-formatted OTU table was filtered to remove singletons. Microbial composition at each taxonomic level (phylum, class, order, family, genus) was defined using the summarize_taxa function in QIIME on the rarefied OTU tables. The relative abundance of each taxon was calculated by dividing the sequences pertaining to a specific taxon by the total number of sequences for that sample. Principal Coordinates Analysis was performed in QIIME v.1.9.1 using UniFrac distances calculated from the Greengenes reference tree.

Triglyceride measurement – Liver triglycerides (TG) were quantified following the Bligh and Dyer extraction method (Bligh and Dyer, 1959). Briefly, 10-30 mg frozen liver tissue was homogenized using a 40X dilution with 1X PBS. Total lipids were extracted from the liver homogenate in methanol-chloroform (2:1). The organic extract with dried and reconstituted in 10% Triton X-100 in isopropanol. Triglyceride content was determined by colorimetric assay, purchased from Wako Diagnostics (Richmond, VA), according to the manufacturer's instructions and expressed as µg of triglycerides per milligram of protein. Total protein was quantified using the Quick Start Bradford Assay (BioRad).

Total cholesterol measurement – 10-30mg of liver was homogenized in a 7:11:0.1 mixture of chloroform:isopropanol:NP-40 at a ratio of 20µl/mg tissue. Hepatic homogenate was centrifuged for 10 minutes at 15,000g. 100µl of the supernatant was dried under N2 gas. Dried samples were re-suspended in 1 volume of 10% Triton X-100 in isopropanol. Re-suspended samples were dried a second time until a vicious gel remained on the bottom and re-suspended in 1/2 volume water. Total cholesterol was measured using the Cholesterol E kit from Wako Diagnostics (Richmond, VA) as follows. Standards were prepared according to package insert instructions. 2µl of each standard and prepared sample was added to a 96-well plate in triplicate. 200µl of the Color Reagent Solution was added to each well. The plate was incubated for 5 minutes at 37˚C prior to the absorbance immediately being read at 540nm.

Western blot – 2 µg of mouse liver histone extract was used for each sample. Extracts were electrophoresed through a 16% polyacrylamide Tris-Glycine gel and transferred to a nitrocellulose membrane. Membranes were blocked for 1 hour at room temperature in 1% PBST (1X phosphate buffered saline + 0.1% Tween-20). Primary antibodies were diluted in 5% milk-PBST and incubated overnight at 4C with rocking, followed by 3 washes with PBST and incubation with secondary antibodies (diluted in 5% milk-PBST) for 1 hour at room temperature with rocking. Blots were washed three times in PBST and once with PBS prior to scanning on an Odyssey Infrared Imager (model no. 9120). Densitometry was performed using Image J v1.50i. All blots were normalized to a total histone H3 loading control and statistical significance was determine by Welch's T-test with a cutoff of $p = 0.05$. Antibodies used: H3 K9ac (Active Motif 39918) 1:1000; H3 K27me3 (Active Motif 61017) 1:1000; pan-acetyl(K) (Cell Signaling

9814) 1:1000; DyLight 680, anti-mouse IgG (Thermo Scientific 35518) 1:6000; DyLight 800, anti-rabbit IgG (Thermo Scientific 35571) 1:10,000.

Additional Information – Cluster 2, which contains genes coregulated in ConvR animals irrespective of diet, was enriched for pathways involved in insulin, SREBP, and PPAR signaling, among other metabolic processes, and antigen processing and presentation (Fig. 3A). Genes involved in insulin signaling and sensitivity and the PI3K/Akt pathway are downregulated in ConvR vs. GF livers on both chow and HF/HS feeding, including *Irs1* (insulin receptor substrate 1), *Irs2* (insulin receptor substrate 2), *Sos1* (son of sevenless 1), *Igfbp2* (insulin-like growth factor binding protein 2), and *Angptl4* (angiopoietin-like 4) (Fig. 3B). The expression of *Lpl* (lipoprotein lipase), a target of insulin signaling and SREBP, is also decreased in ConvR animals on chow and HF/HS-feeding (Fig. 3B). Finally, expression of *G6pc* (glucose 6-phosphatase), a key enzyme in glucose metabolism and target of SREBP, decreased in both chow and HF/HS-fed mice (Fig. 3B). In contrast, expression of *Pten* (phosphatase and tensin homolog), a negative regulator of insulin signaling, increased in both chow and HF/HS-fed ConvR mouse livers vs. GF controls (Fig. 3B). Expression of *SCD1* (stearoyl-CoA desaturase 1), a known SREBP and PPAR target gene, increased in chow-fed colonized mice but not in HF/HS-fed ConvR mice (Fig. 3B).

It is noteworthy that there were also a number of DE genes in cluster 2 that may regulate histone PTM states. Expression of *Acly* (ATP citrate lyase), which is essential for glucose-driven histone acetylation in mammalian cells (Wellen et al., 2009), decreased modestly in the livers of both chow and HF/HS-fed ConvR mice vs. GF controls (Fig. 3B). This downregulation may be due to decreased need for glucose-derived acetyl-CoA in the presence of increased SCFA-derived acetyl-CoA in colonized mice. Expression of *Nampt* (nicotinamide phosphoribosyltransferase, Visfatin), a reported adipokine which catalyzes the rate-limiting step in NAD^+ biosynthesis, increased as a function of colonization (Fig. 3B). In contrast, expression of *Afmid* (arylformamidase), the second step of the kynurenine pathways of tryptophan degradation, which is channeled into NAD^+ biosynthesis, was modestly decreased in ConvR mouse livers (Fig. 3B). These observations suggest that NAD^+ salvage pathways were upregulated while *de novo* NAD^+ biosynthesis was suppressed upon gut colonization. Finally, *Gnmt* (glycine N-methyltransferase) expression decreased in both chow and HF/HS-fed ConvR mouse livers relative to GF controls. This enzyme methylates glycine to produce S-adenosyl homocysteine (SAH), a known inhibitor of histone methyltransferases (HMTs). Thus, decreased expression of *Gnmt* may lead to less inhibition of HMTs and increased histone methylation. In summary, cluster 2 genes were enriched for processes involved in signaling by growth factors and cytokines and metabolism of carbohydrates, lipids, and cholesterol.

In contrast to cluster 2 where DE genes in chow and HF/HS-fed animals were co-regulated, DE genes in clusters 4 and 6 were oppositely regulated across these two dietary conditions. Cluster 4 genes were enriched for KEGG pathways involving cholesterol and retinol metabolism, amino acid metabolism, and inflammatory processes. A large group of genes involved in cholesterol synthesis and homeostasis were present in this cluster, including *Mvd* (mevalonate decarboxylase), *Lss* (lanosterol synthase), *Sqle* (squalene epoxidase), *Fdps* (farnesyl diphosphate), *Insig1* (insulin induced gene 1), *Idi1* (isopentenyl-diphosphate delta isomerase 1), *Fdft1* (farnesyl diphosphate farnesyl transferase 1), *Pcsk9* (proprotein convertase subtilisin/kexin type 9), *Cyp39a1* (cytochrome P450 family 39, subfamily a, polypeptide 1), *Cyp51* (cytochrome P450, family 51), *Cyp2c70* (cytochrome P450, family 2, subfamily c, polypeptide 70), *Apoa1* (apolipoprotein A-1), and *Apom* (apolipoprotein m). There was also a group of co-regulated genes involved in inflammation, which includes *IL1b* (interleukin 1 beta) and the lectins *Clec4a2* (C-type lectin domain family 4, member a2) and *Clec4n* (C-type lectin domain faily 4, member n), which is known to induce Th17 differentiation. Thus a significant portion of cluster 4 genes were involved in cholesterol homeostasis and immune function.

Cluster 6 genes were enriched for metabolic pathways involving lipids, amino acids, and small molecule metabolites (Fig. 3A). Expression of *Plin4* (perlipin 4) robustly increased in both chow and HF/HS-fed livers, and interacted with other genes involved in lipid homeostasis whose expression increased modestly, including *Plin5* (perilipin 5), *Fitm2* (fat storage-inducing transmembrane protein 2), and *Pex11a* (peroxisomal biogenesis factor 11 alpha) (Fig. 3D). Other DE genes in cluster 6 involved in lipid metabolism include *Acat3* (acetyl-coenzyme A acetyltransferase 3), *Fabp2* (fatty acid binding protein 2), and *Apoc2* (apolipoprotein C-II) (Fig. 3D). Notably, expression of *Pparg* (peroxisome proliferator activated receptor gamma), which plays a role in liver lipid synthesis and storage, increased more in HF/HS-fed vs. chow-fed ConvR livers relative to their GF controls (Fig. 3D).

Cluster 6 also contains a group of DE genes involved in folate regulation, which may affect availability of the methyl donor SAM for histone methylation. This includes *Gldc* (glycine decarboxylase), *Sardh* (sarcosine dehydrogenase), and *Dmgdh* (dimethylglycine dehydrogenase), and *Amt* (aminomethyltransferase) (Fig. 3D). The expression of *Sardh, Dmgdh,* and *Amt* decreased modestly in colonized mice relative to GF controls, irrespective of diet, whereas expression of Gldc only decreased in ConvR chow-fed mice but remained relatively unchanged in ConvD chow-fed and ConvR HF/HS-fed mice (Fig. 3D). Finally, a group of cluster 6 genes were also involved in metabolism of organic and amino acids, including *Acss3* (acyl-CoA synthetase short-chain family member 3),

Aldh9a1 (aldehyde dehydrogenase 9, subfamily A1), *Amt* (aminomethyltransferase), and *Aldh3a2* (aminomethyltransferase) (Fig. 3D).

A significant number of hepatic SREBP and PPAR target genes involved in lipid, glucose, bile acid, and amino acid metabolism and inflammation are DE in colonized mice vs. their GF controls (Rakhshandehroo et al., 2010). PPAR targets in cluster 1 include *Mogat1* (monoacylglycerol O-aceyltransferase) and *Cidec* (cell death-inducing DFFA-like effector C), whose expression increased in HF/HS-fed but not chow fed mice. Scd1 (stearoyl CoA desaturase), *Csad* (cysteine sulfinic acid decarboxylase), *Pah* (phenylalanine hydroxylase) were differentially expressed in cluster 2. Notably, expression of *Csad,* which is involved in production of taurine, a key component of conjugated bile acids, increased robustly across all conditions (Fig. 3B). Clusters 3-5 include *UCP2* (uncoupling protein 2), *Slc10a2* (solute carrier family 10 (sodium/bile acid cotransporter), member 2), *Orm2* (orosomucoid 2), and *Saa2* (serum amyloid A2). Finally, 7 DE genes in cluster 6 are PPAR targets: the lipid metabolism genes *Aldh9a1* (aldehyde dehydrogenase 9 family, member A1), *Pex11a* (peroxisomal biogenesis factor 11 alpha), *Fabp2* (fatty acid binding protein 2), *Mgll* (monoglyceride lipase); two cholesterol and bile acid transporters *Abcb4* (ATP-binding cassette, sub-family B, member 4) and *Abcg8* (ATP-binding cassette, sub-family G, member 8); and the biotransformation enzyme *Ugt1a9* (UDP glucuronosyltransferase 1 family, polypeptide A9). These 7 genes all displayed higher expression in HF/HS-fed mice vs. chow fed counterparts (Fig. 3A, 3D). *Pparg* is itself a target of SREBP and its expression increased more in HF/HS fed mice vs. chow fed. Other SREBP targets with increased expression include *Me1* (malic enzyme 1), *Scd1*, and *Pcsk9*, and those with decreased expression include *Lpl, Acly, Fdft1, Fdps,* and *Cyp51*.

Finally, a number of clusters are enriched for genes involved in innate and adaptive immunity. This includes *IL1b* (interleukin 1, beta), *Clec4n* (C-type lectin domain, family 4, member n), and *Clec4a2* (C-type lectin domain, family 4, member a2) of cluster 4 (Fig. 3C). In cluster 5, the innate immunity genes *Saa1-3* (serum amyloid A 1, 2, and 3), *Orm2* (orosomucoid 2), and *Lcn2* (lipocalin 2), are upregulated 135- to 966-fold in ConvR mice on a chow diet, but only 1.6- to 4.2-fold in ConvR HF/HS-fed mice (Supplemental Fig. 3A). Interestingly, these genes are relatively unchanged in ConvD chow-fed mice, and expression of *Saa1* decreases ~3-fold.

References

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L., Serova, N., Davis, S., Soboleva, A., 2013. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Research 41, D991–5. doi:10.1093/nar/gks1193

Bligh, E.G., Dyer, W.J., 1959. A rapid method of total lipid extraction and purification. Canadian Journal of Biochemistry and Physiology 37, 911–917. doi:10.1139/o59-099

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L., Knight, R., 2010a. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics 26, 266–267. doi:10.1093/bioinformatics/btp636

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, 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., Knight, R., 2010b. QIIME allows analysis of high-throughput community sequencing data. Nat Chem Biol 7, 335–336. doi:10.1038/nmeth.f.303

Dennis, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., Lempicki, R.A., 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 4, P3. doi:10.1186/gb-2003-4-9-r60

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and Environmental Microbiology 72, 5069–5072. doi:10.1128/AEM.03006-05

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. doi:10.1093/bioinformatics/btq461

Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., Mering, von, C., 2009. STRING 8-a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Research 37, D412–D416. doi:10.1093/nar/gkn760

Kessner, D., Chambers, M., Burke, R., Agus, D., Mallick, P., 2008. ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 24, 2534–2536. doi:10.1093/bioinformatics/btn323

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Applied and Environmental Microbiology 79, 5112–5120. doi:10.1128/AEM.01043-13

Krautkramer, K.A., Reiter, L., Denu, J.M., Dowell, J.A., 2015. Quantification of SAHA-Dependent Changes in Histone Modifications Using Data-Independent Acquisition Mass Spectrometry. J. Proteome Res. 14, 3252–3262. doi:10.1021/acs.jproteome.5b00245

Leng, N., Dawson, J.A., Thomson, J.A., Ruotti, V., Rissman, A.I., Smits, B.M.G., Haag, J.D., Gould, M.N., Stewart, R.M., Kendziorski, C., 2013. EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics 29, 1035–1043. doi:10.1093/bioinformatics/btt087

Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323. doi:10.1186/1471-2105-12-323

MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., MacCoss, M.J., 2010. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26, 966–968. doi:10.1093/bioinformatics/btq054

Maile, T.M., Izrael-Tomasevic, A., Cheung, T., Guler, G.D., Tindell, C., Masselot, A., Liang, J., Zhao, F., Trojer, P., Classon, M., Arnott, D., 2015. Mass Spectrometric Quantification of Histone Post-translational Modifications by a Hybrid Chemical Labeling Method. Mol. Cell Proteomics 14, 1148–1158. doi:10.1074/mcp.O114.046573

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., Hugenholtz, P., 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J 6, 610–618. doi:10.1038/ismej.2011.139

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504. doi:10.1101/gr.1239303

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., Egholm, M., Henrissat, B., Heath, A.C., Knight, R., Gordon, J.I., 2009. A core gut microbiome in obese and lean twins. Nature 457, 480–484. doi:10.1038/nature07540