ELUCIDATION OF HOST GENETICS – GUT MICROBIOME INTERACTIONS, AND THE CONSEQUENCES ON METABOLIC DISEASE

By

Julia Hill Kemis

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Cellular and Molecular Pathology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2018

Date of final oral examination: 11/19/2018

This dissertation is approved by the following members of the Final Oral Committee:
Federico E. Rey, Assistant Professor, Bacteriology
Alan D. Attie, Professor, Biochemistry
Cameron Currie, Professor, Bacteriology
Garret Suen, Associate Professor, Bacteriology
C. -L. Eric Yen, Associate Professor, Nutritional Sciences

ABSTRACT

The population of microbes that inhabit the mammalian intestine have profound effects on host physiology. The gut microbiome varies substantially among healthy individuals, and its composition is shaped by a complex interplay of environmental and genetic factors. Alterations in its composition are associated with the development of metabolic diseases, including obesity and type 2 diabetes. Therefore, manipulation of the intestinal microbiome ecosystem is a promising target for emerging therapies. However, it remains largely unknown how host genetics interacts with environmental factors (e.g. diet) to shape microbiota profiles, and how these interactions may contribute to metabolic disease susceptibility.

The objective of this thesis research was to investigate the effects of host genetic variation on gut microbiota composition, evaluate how these interactions influence host diet-induced metabolic phenotypes, and to identify genetic variants that influence the abundance of gut microbes.

In **Chapter 2**, I evaluate the relative contributions of host genetics and diet on gut microbiota composition and metabolic phenotypes using a panel of eight genetically diverse inbred mouse strains. In a controlled laboratory environment, I found gut microbiota composition and metabolic phenotypes are shaped by both genetics and diet. Guided by the results of this screen, I went on to demonstrate that in a gnotobiotic mouse model transplantation of genotype-associated microbiota can alter pancreatic islet function and confer sustained metabolic phenotypes despite chronic high-fat high-sucrose (HF/HS) feeding.

In **Chapter 3**, I identify host genetic loci that influence gut microbiota and bile acid profiles. I performed quantitative trait loci (QTL) mapping to find genetic variants associated with abundance of gut microbes and bile acid levels using the Diversity Outbred (DO) mouse stock,

which is derived from the eight strains profiled in Chapter 2. I found novel genetic variants associated with both microbial taxa and bile acids, including an association between the intestinal bile acid transporter, *Slc10a2*, the abundance of *Turicibacter sp*. and plasma cholic acid levels. Subsequent investigation revealed direct interactions between *Turicibacter sp.* and bile acids *in vitro*, supporting a role of genetics in elucidating host-microbe interactions.

Together, this thesis work contributes to our understanding host-microbe interactions and provides a foundation for future mechanistic studies.

ACKNOWLEDGEMENTS

Thank you to my advisor Federico Rey for your guidance and for the opportunity to grow as an independent scientist. Thank you for creating an enriching lab environment and for continuously challenging me. You have worked tirelessly to establish the lab and I am eager to see where the Rey Lab goes in the future.

Thank you to my committee members, Alan Attie, Cameron Currie, Garret Suen and Eric Yen, for the support and insightful discussions over years, both in committee meetings and out.

Thank you to the current and former members of the Rey Lab for shaping my graduate experience. I am grateful for their friendship, guidance, feedback, knowledge, and excellent sense of humor. It has been a privilege to watch the lab evolve over the years and to be a part of the wonderful community it is today.

Thank you to all of our collaborators I've been privileged to work with over the years. To say science takes a team is an understatement, and without them, this work wouldn't be possible.

Thank you to the countless mentors and friends who have guided and inspired me over the years. Specifically, thank you to Kathy Jacobson and Kirsten Beyer for nurturing my interest in scientific research and providing me with two exemplary role models during college.

Thank you to my family for the unwavering support and unconditional love that has laid the foundation for my success. Thank you to my extended family for the support, inspiration and for encouraging me to effectively communicate my research to a broad audience. Thank you to my cousin, Kari, for always being my rock and my best friend. Thank you to my brother, Christian, for challenging me and making everything more fun. Thank you to my parents, for working tirelessly to help me achieve my goals and for always having my back. I am incredibly grateful for your endless enthusiasm, confidence, and love.

Finally, thank you to my husband, Jason, for always pushing me to follow my passions and for providing unconditional support and infinite patience. You help me keep things in perspective and always know how to make me laugh, even when I really don't want to. I am eternally grateful for the sacrifices you have made to make my aspirations become reality.

TABLE OF CONTENTS

CHAPTER 1: Host – Microbe Interactions and the Contributions of the Gut Microbiome to Disease

INTRODUCTION

The mammalian gut harbors trillions of commensal microorganisms, comprised of bacteria, viruses, archaea and eukaryotes (fungi and protists), which together constitute the intestinal microbiota (Sender et al., 2016). These microorganisms have evolved with the host to establish mutualistic symbioses, where they reside in nutrient rich intestine and as reciprocity they expand the host metabolic repertoire, allowing for the breakdown of otherwise indigestible carbohydrates and the synthesis of essential nutrients (Qin et al., 2010; Sommer and Bäckhed, 2013). It has become clear that the microbes that inhabit the intestine play a critical role in determining many aspects of health. Alterations in microbiota composition are implicated in a spectrum of metabolic (Karlsson et al., 2013a), immunological (Petersen and Round, 2014) and cognitive disorders (Petra et al., 2015; Vuong et al., 2017). To date, most research has focused on identifying individual gut microbes and microbial communities associated with healthy and diseased states, but the underlying mechanisms remain largely elusive.

Sequencing-based studies of fecal microbial communities have also revealed substantial inter-individual differences in microbiota composition (Human Microbiome Project Consortium, 2012; Qin et al., 2010; Spor et al., 2011). Microbiomes can be characterized by 16S rRNA gene sequencing and metagenomic sequencing, which allow for the quantification of bacterial taxa and gene functions, respectively (Goodrich et al., 2014a). Thousands of organisms are capable of colonizing the human intestine (Lozupone et al., 2012) and their combined genomes contain >100-fold more genes than are encoded in the human genome (Qin et al., 2010). The observed variation in gut microbiota communities is driven by a multitude of variables including environmental factors (e.g. maternal seeding, diet) and host genetics (Costello et al., 2009; Falony et al., 2016; Goodrich et al., 2014b; Zhernakova et al., 2016).

Due to the contributions of the gut microbiota to various diseased states, it has become increasing important to identify and understand the specific factors that govern microbiota composition. However, the extent to which host genetics shapes microbiota composition and the causal variants remain poorly characterized. In this review, we summarize the importance of the gut microbiome for host health and the current research deciphering relative contributions of environmental and genetic factors in shaping the composition of these microbial communities. Additionally, we discuss how genetic-driven variation in the microbiome influences the host and highlight the role of metabolites in mediating host-microbe interactions.

THE ENTERIC MICROBIOTA IN HEALTH AND DISEASE

Contributions to host development

The presence of the enteric microbiota is instrumental for proper host development. The importance of these microbes in development is evident from comparative studies with germ-free animals which have extensive developmental defects. Germ-free mice are devoid of any microorganisms and can be used to evaluate the contributions of microbes to clinical phenotypes. Studies of germ-free animals demonstrate the gut microbiota modulates many aspects of development ranging from bone-mass density (Sjögren et al., 2012), intestinal angiogenesis (Reinhardt et al., 2012), intestinal architecture and mucus layer properties (Hooper and Gordon, 2001; Petersson et al., 2011; Sharma et al., 1995), and innate and adaptive immune systems (Cebra, 1999; Ivanov et al., 2008; Macpherson and Harris, 2004). Best described in germ-free animals are the roles of the gut microbiota in development and maturation of the intestinal epithelium and the immune system.

Germ-free animals have stark differences in intestinal morphology compared to those fully colonized. Most noticeably, germ-free rodents have an enlarged cecum (Wostmann, 1981). Other

morphological differences include reduced intestinal surface area (Gordon and Bruckner-Kardoss, 1961) and villus thickness (Reinhardt et al., 2012), as well as impaired brush border differentiation (Abrams et al., 1963). Furthermore, the absence of a microbiota impairs regulation of cell turnover and promotion of cell renewal (Smith et al., 2007). In addition to intestinal morphology, gut microbes serve an important role in maintaining mucosal barrier integrity (Natividad and Verdu, 2013). In fact, the addition of specific organisms to the intestinal ecosystem can improve intestinal barrier function including *Bacteroides thetaiotaomicron* (Hooper et al., 2001), *Akkermansia muciniphila* (Reunanen et al., 2015) and *Lactobacillus plantarum* (Zhou et al., 2010).

Human and animal studies have also shown a direct role of microbiota in the maturation and function of the immune system (Brestoff and Artis, 2013). Intestinal microbes contribute to immune system by promoting the development of lymphoid structures, as well as through the modulating of activation and differentiation of several lymphocyte populations (Round and Mazmanian, 2009). Moreover, the microbiome is required for development of completely functional IgA-producing cells (Kawamoto et al., 2012) and germ-free mice have limited IgA plasmablasts in their gut lamina propria (Crabbé et al., 1970). However, the number of IgA plasma cells is greatly expanded after colonization (Crabbé et al., 1968). These deficiencies in immune and intestinal development leave the germ-free host susceptible to invasion by opportunistic pathogens (Smith et al., 2007).

Immune implications of gut microbiome

In addition to stimulating the development of the host immune system, gut bacteria are instrumental for maintaining immune homeostasis. There is substantial evidence that perturbations the intestinal microbiota contribute to the development of numerous inflammatory disorders including chronic inflammatory bowel diseases (Devkota et al., 2012; Sokol et al., 2008),

rheumatoid arthritis (Vaahtovuo et al., 2008), asthma (Arrieta et al., 2015), and type 1 diabetes (Greiner et al., 2014; Wen et al., 2008). Bacterial species interact with several cell types including epithelial cells, dendritic cells, and T cells to influence host immune responses. Commensal microbes maintain immune homeostasis through both immune-stimulatory and immune-modulatory effects. Starting at birth, the early intestinal colonizers instigate development and maturation of the immune system (Brestoff and Artis, 2013). Microbes stimulate the immune system by components of their cell wall and their metabolites. For example, microbial production of short chain fatty acids (SCFAs) induce immune cell activation, cytokine production and T-lymphocyte proliferation (Corrêa-Oliveira et al., 2016). Commensal microbes help maintain intestinal barrier function through the recruitment of immune cells to the mucosa (Macpherson and Harris, 2004), as well as stimulate protective epithelial functions such as the secretion of mucus and antimicrobial peptides (Hooper and Macpherson, 2010). Gut microbes also serve an immune-modulatory role to prevent overactivation of inflammatory and allergic responses. For example, microbial production of the short chain fatty acid butyrate induces colonic regulatory T cells (T_{Reg} cells) (Furusawa et al., 2013). T_{Reg} cells regulate the immune system through the induction of anti-inflammatory cytokines IL-10 and IL-35. The induction of T_{Reg} cells by commensal organisms is crucial to limiting inflammation and disease. These immune-modulatory effects determine the robustness of the host immune response and influence host health.

Metabolic syndrome and the role of the microbiome

There is substantial evidence that the gut microbiota influences development of metabolic and cardiovascular diseases. Comparative studies between health and diseased individuals identified alterations in the microbiota composition in obesity (Ley et al., 2006; Turnbaugh et al., 2009a), type 2 diabetes (T2D) (Karlsson et al., 2013b; Qin et al., 2012), nonalcoholic fatty liver

disease (Henao-Mejia et al., 2012), and atherosclerosis (Wang et al., 2011). Lower microbial diversity is also associated with many of these metabolic syndromes (Le Chatelier et al., 2013; Turnbaugh et al., 2009a). Greater microbial species richness, or the number of species present, is associated with a healthier intestinal ecosystem considered more stable and less susceptible to invasion by new species (Lozupone et al., 2012). This may be an indication of dietary differences between lean and obese individuals, given Hadza hunter-gathers consume food with high-fiber content and have greater microbial species richness compared with individuals living in Westernized nations (Schnorr et al., 2014). Moreover, greater richness in microbial species and genes has been observed in lean compared to obese individuals (Turnbaugh et al., 2009a).

In addition to these associations, studies using germ-free mice have demonstrated a causal role of the microbiome in metabolic disease development. Germ-free mice are protected from high-fat diet-induced obesity and have significantly less body fat mass than conventionally-raised mice independent of food-intake (Bäckhed et al., 2004). This resistance is attributed to altered fatty acid metabolism and reduced energy harvest from dietary substrates (Bäckhed et al., 2007). Transplant studies have demonstrated a causal role of the microbiota in obesity (Ridaura et al., 2013), insulin sensitivity (Vijay-Kumar et al., 2010), and insulin secretion from pancreatic islets (Kreznar et al., 2017; Perry et al., 2016). Interestingly, there appears to be a similar causal relationship in humans, as demonstrated from a study by Vrieze and colleagues that found transfer of intestinal microbes from lean donors to recipients with metabolic syndrome improved glucose metabolism and insulin sensitivity (Vrieze et al., 2010). Therefore, manipulation of the gut microbiome may be an effective therapy for treating metabolic disorders. However, the organisms responsible for these metabolic changes are relatively unknown.

The gut microbiota influences the development of metabolic disorders in part by altering host energy harvest and inflammation. Intestinal microbes provide important metabolic capabilities that influence the efficiency of energy harvest from diet and how it is stored (Bäckhed et al., 2004; Turnbaugh et al., 2006). Microbiota from obese individuals has an increased capacity for energy harvest compared with lean individuals and this metabolic capability can be transplanted into germ-free animals (Turnbaugh et al., 2006, 2009a). By-products of microbial metabolism also influence host energy balance by acting as ligands for various host receptors. For example, short chain fatty acids (SCFAs), which are derived from bacterial fermentation of dietary fibers, are involved in energy regulation through altering host epigenetics (Krautkramer et al., 2016) and through interactions with G protein-coupled receptors (GPCRs) (Gao et al., 2009; Lin et al., 2012). GPCRs stimulate the release of anorexigenic intestinal hormones glucagon-like peptide-1 (GLP-1) and peptide YY (PYY), thereby modulating host energy homeostasis. Systemic and adipose tissue inflammation are hallmarks of obesity, insulin resistance and T2D (Osborn and Olefsky, 2012), and this inflammatory state may be attributed to microbiota composition. Bacterial cell wall components like lipopolysaccharide (LPS) and peptidoglycan cause inflammation (Rietschel et al., 1998). Increased plasma levels of LPS have also been observed in patients with metabolic syndrome (Creely et al., 2007). Consumption of a high-fat diet increases intestinal permeability, leading to an increase in the translocation of microbial-derived components into the circulatory system and systemic inflammation (Cani et al., 2007). Together, these data suggest microbiota may contribute to metabolic diseases through several mechanisms, such as increased energy harvest from diet and altered systemic and adipose tissue inflammation.

ENVIRONMENTAL FACTORS SHAPING MICROBIOTA COMPOSITION

Early life establishment

The assembly of the intestinal microbiome is initiated at birth, with rapid colonization by microbes in the surrounding environment. Vertical transmission of the microbiome from mother to infant is facilitated by delivery mode, which determines colonization patterns of the infant microbiome (Bokulich et al., 2016; Dominguez-Bello et al., 2010; Yassour et al., 2016). Infants born via vaginal delivery acquire bacterial communities that resemble their mother's vaginal microbiome, while caesarian (CS) delivered infants harbor a microbiota that is the most similar to their mother's skin (Dominguez-Bello et al., 2010). Infants delivered vaginally were also had a greater abundance of *Lactobacillus*, while CS-delivered infants had a greater abundance of *Staphylococcus* (Dominguez-Bello et al., 2010). In addition to delivery mode, breastmilk shapes the infant microbiota by providing a continuous supply of potentially probiotic bacteria to the infant gut (Fernández et al., 2013), along with secreted maternal antibodies that provide protection from harmful species (Rogier et al., 2014). The presence of various oligosaccharides in breastmilk have been shown to select for specific gut microbes (Zivkovic et al., 2011). Additionally, antibiotic administration is considered one of the most significant factors affecting the infant microbiota and results in a loss of microbial diversity (Jernberg et al., 2010). Although the effects of these early life events on microbiota composition are lost within the first year of life (Rutayisire et al., 2016). The microbiome is also more susceptible to perturbations during infancy before a stable microbial community has developed. This coincides with a critical host developmental period where microbial colonizers facilitate/orchestrate immune and metabolic development. Therefore, disruption of the infant microbiome during this critical development window may have lasting effects on the host (Ajslev et al., 2011; Cox et al., 2014; Kelly et al., 2007). Antibiotic perturbations of the infant microbiota alter the host's metabolic activity, resulting in growth promotion and an

increased risk of obesity (Cox et al., 2014). Additionally, early life alterations in the microbiota are associated with immunological changes in intestinal mucosa (Renz et al., 2012) and the function of intestinal natural killer T cells is shaped by age-sensitive contact with commensal microbes (Olszak et al., 2012). Taken together, early life events have a profound impact on shaping the microbiota composition, and in turn, host health.

Lifestyle factors

The composition of the gut microbiota is also shaped by a range of lifestyle factors including medications, social relationships, personal habits and location. Medication such as antibiotics (Antonopoulos et al., 2009; Cox et al., 2014), proton-pump inhibitors (Imhann et al., 2016), metformin (Forslund et al., 2015) and antidepressants (Zhernakova et al., 2016) shift microbiota profiles. Medications have differential effects on the microbiota profiles. Many of these studies report decrease in species richness, which may have detrimental effects on the host. However, other xenobiotics, like metformin, appear to shift the microbiota to a more beneficial state. Metformin treatment in individuals with T2D increased the abundance of bacteria capable of producing SCFAs, which are linked with many metabolic benefits (Kasubuchi et al., 2015). Other lifestyle factors such as smoking (Biedermann et al., 2013) and exercise (Allen et al., 2018) also influence the microbiota. Geographical location contributes to variation as shown by differences among geographically discrete populations (Rehman et al., 2016; Yatsunenko et al., 2012) and among individuals living in rural vs urban environments (Tyakht et al., 2013). Recently, there has been an increase in studies focused on understanding the role of social relationships in shaping the microbiota (Herd et al., 2018). Co-habitation has been identified in several studies as having a stronger effect than relatedness in determining the similarity of microbiota composition among individuals (Dill-McFarland et al., 2018; Rothschild et al., 2018; Song et al., 2013). For

example, married individuals had greater bacterial diversity and richness than those living alone (Dill-McFarland et al., 2018). Interestingly, social dynamics appear to influence the microbiome as couples reporting close relationship had greater bacterial diversity than ones who reported a somewhat close relationship (Dill-McFarland et al., 2018).

Diet drives composition and function of gut microbiome

Diet exerts a strong effect on the gut microbiota and is arguably one of the most significant determinants of microbiota composition and function. Dietary components that are not digested and absorbed by the host pass through the intestine where they serve as primary energy sources for bacteria. The microbiome is highly responsive to alterations in dietary patterns (David et al., 2014; Muegge et al., 2011). Substantial changes in diet have been shown to have profound consequences on the overall composition and metabolic capabilities of the microbiota, such as switching from vegetarian to an omnivorous diet (David et al., 2014) or from a low-fat plant-rich diet to a high-fat high-sugar diet (Turnbaugh et al., 2009b). In fact, diet-induced changes in population structure can occur in a single day (Carmody et al., 2015). Dietary preferences also enrich for specific bacteria taxa. For example, *Prevotella* is enriched in individuals consuming high-fiber diets (De Filippo et al., 2010), while *Bacteroides* is higher in humans consuming high-protein diets (Yatsunenko et al., 2012). The influence of a “Western” high-fat diet on microbiota composition and function has been particularly well characterized for its role in diet-induced obesity. Western-style diets cause substantial changes in microbiota composition and function, as demonstrated by studies in humans and rodent models. A frequently observed theme in response to a Western diet is a shift in the ratio of the major phyla Bacteroidetes and Firmicutes (Carmody et al., 2015; Kreznar et al., 2017; Parks et al., 2013; Turnbaugh et al., 2009b). Moreover, diets high in fat are associated with decreased overall microbiome diversity (Turnbaugh et al., 2008).

Microbes enriched in response to a Western diet also allow the host to harvest more energy (Ley et al., 2005; Turnbaugh et al., 2006), as shown by an enrichment of microbial pathways involved in nutrient processing.

HOST GENETICS SHAPES MICROBIOME

Interpersonal differences in the overall microbiota composition and the inter-individual variation in bacterial taxa can be partially attributed to host genetics. Early investigations focused on comparisons of overall microbiome composition (β-diversity) among related and unrelated individuals. One study found the overall similarity of the gut microbiome increased with closer degrees of relatedness in families (Erwin G. Zoetendal, Antoon D. L. Ak, 2001) . Twin studies corroborated these findings, where the microbiota composition between twins (both monozygotic (MZ) and dizygotic (DZ)) were more similar to one another than to unrelated individuals (Turnbaugh et al., 2009a; Yatsunenko et al., 2012). In fact, a comparison of 416 twin pairs found that MZ twins are more similar to each other than DZ twins (Goodrich et al., 2014b). However, these differences between MZ and DZ twins can only be discerned with a sufficient sample size. Interestingly, metagenomic analysis of the microbiome of TwinsUK cohort found that the similarity among twins decreased once they lived apart (Xie et al., 2016). This suggesting that environment may mask the contributions of genetics on shaping the microbiota and the effects of genetic variation is most pronouced among individuals in a shared environment.

Since contributions of genetics is confounded by environmental factors, experiments using inbred mouse have proven to be valuable to evaluate the extent to which genetic variation shapes microbial communities. Mouse models are well suited for discerning the influence of host genetics because many confounding environmental factors can be carefully controlled (Spor et al., 2011). Additionally, different breeding approaches can be utilized to maximize genetic diversity and

mimic the variation found among humans. For example, the Collaborative Cross (CC) progenitor strains are comprised of 5 classical inbred strains and 3 wild derived strains, which together encompass the genetic diversity found in human population (Churchill et al., 2004). Three separate characterizations of these 8 strains by different research groups found microbial signatures unique to each strain, where the microbiome within strains was more similar than between strains (Kovacs et al., 2011; Kreznar et al., 2017; O'Connor et al., 2014). Similarly, another study performed by Org and colleagues profiled fecal samples from 113 mouse strains that comprise the Hybrid Mouse Diversity Project (HMDP) population. Again, they also found greater similarity in microbiota structure among mice of the same genotype (Org et al., 2015).

In addition to comparing overall structure of related and unrelated populations, many studies have utilized heritability measurements to identify specific bacterial taxa influenced by genetics. Here, heritability is defined as the extent to which the total phenotypic variation for a trait is attributed to genetic rather than environmental factors. Studies of the heritability of the gut microbiota in different organisms have collectively identified a subset of organisms that appear to be influenced by host genetics, including *Turicibacter*, *Oscillospira*, *Lactobacillus*, *Lactococcus*, *Roseburia* and *Akkermansia* (Benson et al., 2010; Goodrich et al., 2016; Org et al., 2015). Interesting, all of these genera are part of the Firmicutes phyla, with the exception of *Akkermansia* which is in the Verrucomicrobia phyla, indicating bacteria in this phylum are particularly sensitive to host genetics. Heritability estimates are substantially higher in mice compared to humans where environmental factors are controlled. In mice, heritability estimates for individual taxa range from 26% – 86% (O'Connor et al., 2014; Org et al., 2015), whereas heritability only accounts for 1.9% to 8.1% of the overall variation in microbiota composition in humans (Goodrich et al., 2016; Rothschild et al., 2018). Despite these observed differences in heritability, the congruence of

heritable taxa between studies provides strong evidence that specific, identifiable taxa are responsive to host genotype across populations.

Genetic mapping approaches can be applied to identify specific genetic loci associated with bacterial abundance and community diversity. Two approaches can be used, quantitative trait loci (QTL) analysis or genome wide association studies (GWAS). QTL analysis can be applied to intercross populations where kinship is known. GWAS relies on regression at measured markers and a larger sample size is required. Benson et al. (Benson et al., 2010) was the first to identify host genetic variants associated with microbial abundance using the fourth generation (G4) of an advanced intercross mouse population by QTL mapping. In total, 13 loci were significantly associated with microbial abundance. Notably, this study provided valuable insight into the underlying genetic architecture that shapes microbiota composition. For example, the researchers identified several genomic regions where a single locus associated with multiple microbial traits, indicating that host genetic variation can also influence population structure. A follow-up study using later generations of these intercrossed mice (G10) identified an additional 42 microbial QTL and replicated four QTL identified in G4 (Leamy et al., 2014). Several other QTL studies using different mouse populations and breeding schemes have identified even more regions associated with bacterial abundance (McKnite et al., 2012; Snijders et al., 2017; Wang et al., 2015). Another study by Org et al. (Org et al., 2015) used a GWAS approach with 110 HMDP strains and identified 7 loci associated with microbial abundance. While the QTL overlaps among these studies are extremely limited, some of the taxa with the strongest associations were similar including Lachnospiraceae, *Ruminococcus, Lactobacillus, Turicibacter, Bacteroides* and *Oscillospira*. Interestingly, several of these associations occurred with taxa like Lachnospiraceae, *Ruminococcus* and *Turicibacter*, which were also identified as highly heritable (Benson et al., 2010; Goodrich et

al., 2016). Moreover, several of these microbial QTL were also linked with obesity, lipid levels and markers of immune response (Leamy et al., 2014; McKnite et al., 2012; Org et al., 2015; Snijders et al., 2017).

Human GWAS studies have also shed important insight into genetic determinants of both gut microbiota composition and function. The first microbial GWAS studies were limited by sample size, but still identified several interesting associations between variants in metabolism-related genes (Blekhman et al., 2015; Davenport et al., 2015). For example, a variant near *PLD1*, a gene previously associated with body mass index (Ng et al., 2012) was associated with the abundance of the genus *Akkermansia*, which is known to affect obesity (Everard et al., 2013). Blekham et al. (Blekhman et al., 2015) used a subset of 93 individuals from the Human Microbiome Project for whom they had both genotype and metagenomics data. A significant association between increased *Bifidobacterium* genus and a SNP (rs56064699) located in the lactase (*LCT*) gene was found in this population. Interestingly, the association variant in *LCT* and *Bifidobacterium* abundance was observed in larger population cohorts (Goodrich 2016, Bonder 2016, Rothschild 2018). This association was recently replicated in an expanded analysis using 298 HMP participants (Kolde et al.).

Additional associations have been identified by large-scale population studies using German (Wang et al., 2016), Dutch (Bonder et al., 2016) and Canadian cohorts (Turpin et al., 2016). All three cohorts were comprised of more than 1,000 unrelated individuals and each study included replication cohorts. Despite the large cohorts, the observed associations had small effect sizes, demonstrating the genetic architecture underlying microbiome traits is highly complex. Turpin et al. (Turpin et al., 2016) identified 6 significant associations with microbial taxa. In the German cohort, Wang et al. (Wang et al., 2016) analyzed bacterial β-diversity and discovered an

association with a variant in the *VDR* gene that encodes the vitamin D receptor. They went on to show significant shifts in microbiota composition in *Vdr*$^{-/-}$ mice relative to control mice. The Dutch cohorts replicated several findings from a previous study of UK twins (Goodrich et al., 2016), including strong heritability of Methanobacteriaceae—a family that belongs to the Archaea, and the bacterial genus *Blautia* (Bonder et al., 2016). Moreover, they found significant associations between microbial functional groups and variants in C-type lectins, which are proteins involved in modulating innate immunity.

A common theme that emerges from these genetic mapping studies is the association between the microbiome and variants in immune related genes. The reciprocal role of the immune system in modulating microbiota composition is evident from knockout of innate and adaptive immune genes in mice, including *Tlr5* (Vijay-Kumar et al., 2010), *Nod2* (Rehman et al., 2011), *Myd88* (Larsson et al., 2012), *Card9* (Lamas et al., 2016) and *Rag1* (Dimitriu et al., 2013). Innate immune genes are responsible for sensing microbes and triggering down-stream cell signaling pathways, while adaptive immune genes maintain immune homeostasis through antigen recognition and immunological memory. The genetic mapping studies discussed above corroborate the importance of immune genes in shaping intestinal microbial communities. For instance, Benson et al (Benson et al., 2010) discovered QTLs for Coriobacteriaceae and *Lactococcus* on mouse chr 10 where the loci contained multiple genes involved in mucosal immunity including *Irak3*, *Il22*, and lysozyme genes *Liz1* and *Liz2*. Abundance of Rikensenellaceae and *Roseburia* were associated with *Irak4* (McKnite et al., 2012; Org et al., 2015), which encodes for a kinase that activates TLR- and T cell-receptor signaling pathways (Suhir and Etzioni, 2010). In humans, microbial functions associated with polymorphisms in genes known immune genes NOD1 and NOD2, as well as genes implicated in IBD risk (CCL2, DAP2,

IL23R)(Bonder et al., 2016). These studies highlight the importance of the bidirectional interaction of the microbiome and host immune system.

HOST GENOTYPE-SHAPED MICROBIOMES ALTER DISEASE SUSCEPTIBILITY

Heterogeneity in disease susceptibility can in part be explained by host genotype-driven differences in the gut microbiome. There is evidence that gut bacteria act as a causal link between the genetic and phenotypic diversity among genetically diverse inbred mouse strains (Kasahara et al., 2018; Kreznar et al., 2017; Parks et al., 2013). For example, the eight CC progenitor strains show substantial variation in metabolic phenotypes to atherogenic and Western-style high-fat high-sucrose diets when raised in the same environment (Kreznar et al., 2017; O'Connor et al., 2014). Some of these strains are highly responsive to the Western diet and become obese and glucose intolerant, while others remain lean and insulin sensitive even after 22 weeks of dietary challenge (Kreznar et al., 2017). Strikingly, microbiota transplantation of microbiota from strains with disparate phenotypes replicates aspects of donor metabolic phenotypes in recipient mice. Cecal contents from Western diet-responsive and -resistant mice were transplanted into germ-free recipient animals of the same genotype as the diet-responsive strain and fed the Western diet for 16 weeks. Despite the dietary challenge, germ-free mice that received the microbiota of the diet-resistant strain gained significantly less weight than the mice that received the diet-response microbiota.

A separate study using the Hybrid Mouse Diversity Panel (HMDP) also found a causal role of host genotype-shaped microbiota on cardiovascular disease development. The HMDP consists of ~100 inbred strains that exhibit diverse microbiota community structure and have varying susceptibility to obesity and atherosclerosis (Bennett et al., 2015; Parks et al., 2013). Groups of germ-free *ApoE*$^{-/-}$ mice were colonized with cecal microbiota from four HMDP strains that showed

disparate atherosclerosis phenotypes. Researchers found that the microbiota successfully conferred cardiovascular phenotypes into recipient mice, as mice that received the microbiota from athero-prone HMDP strains developed larger aortic lesion sizes than germ-free animals that received the microbiota from athero-resistant HMDP strains (Kasahara et al., 2018). For the mouse populations used for these studies, environmental factors are the same and the only variation among these animals is genotype, so differences in the microbiota composition can mostly be attributed to genetic variation. These transplantation studies demonstrate the importance of host genotype-shaped microbiota in modulating susceptibility to metabolic and cardiovascular diseases.

METABOLITES: AT THE INTERSECTION OF HOST-MICROBR INTERACTIONS

Variation in host genetics and gut microbiota composition shape metabolic disease development in part by the production and modification of metabolites. Microbial metabolism of dietary substrates produces a myriad of metabolites with differential effects on host physiology (Nicholson et al., 2012). For example, intestinal bacteria metabolize dietary choline to trimethylamine (TMA), which is processed by the host hepatic enzyme FMO3 to produce the pro-atherogenic metabolic trimethylamine-*N*-oxide (TMAO) (Wang et al., 2011). Through fermentation reactions, the gut microbiota can metabolize complex polysaccharides to produce SCFAs. Acetate, butyrate, and propionate are the most abundant SCFAs in the distal gut and they have different effects on the host (den Besten et al., 2013). In general, reduced levels of total SCFAs and SCFA producing bacteria are associated with obesity (Ridaura et al., 2013) and T2D (Karlsson et al., 2013b; Qin et al., 2012). Butyrate has been shown to improve intestinal barrier function and ameliorate atherosclerosis development (Kasahara et al., 2018). Furthermore, acetate can directly influence glucose homeostasis through stimulation of insulin secretion from pancreatic islets (Perry et al., 2016; Priyadarshini et al., 2015). Gut microbes also synthesize essential

vitamins, such as B and K, that are substrates for metabolic reactions (Hill, 1997; LeBlanc et al., 2013). Interestingly, several of these microbially-derived metabolites have also been shown to directly interact with the host genome by altering host epigenetic status through histone acetylation and methylation (Krautkramer et al., 2016; Romano et al., 2017).

Bile acid (BA) metabolites are of particular interest because their composition and abundance are shaped by both the host and intestinal bacteria. BAs reciprocally modulate gut microbiota composition through alterations of the chemical and physical properties of the intestine (Islam 2011, Zheng 2017). Primary BAs are synthesized in the liver from cholesterol and are secreted into the duodenum to aid in the digestion of lipids and facilitate nutrient absorption (Russell, 2009). Gut microbes can metabolize primary BAs through several chemical reactions (deconjugation, dehydrogenation, epimerization and dehydroxylation) to produce secondary BAs (Ridlon et al., 2006), which in turn have varying effects on host physiology and health (Kuipers et al., 2014; Ridlon et al., 2016). BAs act as hormones to regulate lipid, glucose, lipoprotein and energy homeostasis (Li and Chiang, 2014; Zhou and Hylemon, 2014). Alterations in the size and composition of BA pools are associated many diseases like T2D (Handelsman, 2011), obesity (Ryan et al., 2014), IBD (Devkota et al., 2012), and colon cancer (Ajouz et al., 2014). Furthermore, BA metabolizing capabilities of the microbiome is associated with altered host metabolism as demonstrated by loss of bile salt hydrolase (BSH) activity in conventionally-raised and mono-colonized mice (Joyce et al., 2014; Yao et al., 2018)

Recently, several genetic studies have found evidence of interactions between host genetics and the microbiome through the regulation of bile acid metabolism. Much of this work has focused on genetic alterations to gene encoding the vitamin D receptor (VDR), which is involved in bile acid sensing and homeostasis. VDR is a known receptor for secondary bile acids and its activation

can inhibit bile acid synthesis (Makishima et al., 2002). Loss of VDR in mice significantly alters the composition of the microbiota where *Lactobacillus* is depleted and *Clostridium* and *Bacteroides* are enriched (Jin et al., 2015). In humans, polymorphisms in VDR were associated with β-diversity and abundance of *Parabacteroides* (Wang et al., 2016). The authors went on to validate the role of VDR in determining microbial community structure by showing loss of *Vdr* in mice significant affects β-diversity. Additionally, GWAS studies by Blekhman et al. (Blekhman et al., 2015) and Bonder et al. (Bonder et al., 2016) found strong associations between the host variants and bacterial pathways for bile acid metabolism. Together, these studies provide evidence for host genome-gut microbiome interactions regulated by variation in bile acid related genes. Additional investigations are warranted to further elucidate molecular mechanisms.

REFERENCES

Abrams, G.D., Bauer, H., and Sprinz, H. (1963). Influence of the normal flora on mucosal morphology and cellular renewal in the ileum. A comparison of germ-free and conventional mice. Lab. Invest. *12*, 355–364.

Ajouz, H., Mukherji, D., and Shamseddine, A. (2014). Secondary bile acids: an underrecognized cause of colon cancer. World J. Surg. Oncol. *12*, 164.

Ajslev, T.A., Andersen, C.S., Gamborg, M., Sørensen, T.I.A., and Jess, T. (2011). Childhood overweight after establishment of the gut microbiota: the role of delivery mode, pre-pregnancy weight and early administration of antibiotics. Int. J. Obes. *35*, 522–529.

Allen, J.M., Mailing, L.J., Niemiro, G.M., Moore, R., Cook, M.D., White, B.A., Holscher, H.D., and Woods, J.A. (2018). Exercise Alters Gut Microbiota Composition and Function in Lean and Obese Humans. Med. Sci. Sport. Exerc. *50*, 747–757.

Antonopoulos, D.A., Huse, S.M., Morrison, H.G., Schmidt, T.M., Sogin, M.L., and Young, V.B. (2009). Reproducible Community Dynamics of the Gastrointestinal Microbiota following Antibiotic Perturbation. Infect. Immun. *77*, 2367–2375.

Arrieta, M.-C., Stiemsma, L.T., Dimitriu, P.A., Thorson, L., Russell, S., Yurist-Doutsch, S., Kuzeljevic, B., Gold, M.J., Britton, H.M., Lefebvre, D.L., et al. (2015). Early infancy microbial and metabolic alterations affect risk of childhood asthma. Sci. Transl. Med. *7*, 307ra152.

Bäckhed, F., Ding, H., Wang, T., Hooper, L. V, Koh, G.Y., Nagy, A., Semenkovich, C.F., and Gordon, J.I. (2004). The gut microbiota as an environmental factor that regulates fat storage. Proc. Natl. Acad. Sci. U. S. A. *101*, 15718–15723.

Bäckhed, F., Manchester, J.K., Semenkovich, C.F., and Gordon, J.I. (2007). Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. Proc. Natl. Acad. Sci. *104*, 979–984.

Bennett, B.J., Davis, R.C., Civelek, M., Orozco, L., Wu, J., Qi, H., Pan, C., Packard, R.R.S., Eskin, E., Yan, M., et al. (2015). Genetic Architecture of Atherosclerosis in Mice: A Systems Genetics Analysis of Common Inbred Strains. PLoS Genet. *11*, e1005711.

Benson, A.K., Kelly, S.A., Legge, R., Ma, F., Low, S.J., Kim, J., Zhang, M., Oh, P.L., Nehrenberg, D., Hua, K., et al. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. Proc. Natl. Acad. Sci. U. S. A. *107*, 18933–18938.

den Besten, G., van Eunen, K., Groen, A.K., Venema, K., Reijngoud, D.-J., and Bakker, B.M. (2013). The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. J. Lipid Res. *54*, 2325–2340.

Biedermann, L., Zeitz, J., Mwinyi, J., Sutter-Minder, E., Rehman, A., Ott, S.J., Steurer-Stey, C., Frei, A., Frei, P., Scharl, M., et al. (2013). Smoking Cessation Induces Profound Changes in the Composition of the Intestinal Microbiota in Humans. PLoS One *8*, e59260.

Blekhman, R., Goodrich, J.K., Huang, K., Sun, Q., Bukowski, R., Bell, J.T., Spector, T.D., Keinan, A., Ley, R.E., Gevers, D., et al. (2015). Host genetic variation impacts microbiome composition across human body sites. Genome Biol. *16*, 191.

Bokulich, N.A., Chung, J., Battaglia, T., Henderson, N., Jay, M., Li, H., D Lieber, A., Wu, F., Perez-Perez, G.I., Chen, Y., et al. (2016). Antibiotics, birth mode, and diet shape microbiome maturation during early life. Sci. Transl. Med. *8*, 343ra82.

Bonder, M.J., Kurilshikov, A., Tigchelaar, E.F., Mujagic, Z., Imhann, F., Vila, A.V., Deelen, P., Vatanen, T., Schirmer, M., Smeekens, S.P., et al. (2016). The effect of host genetics on the gut microbiome. Nat. Genet. *48*, 1407–1412.

Brestoff, J.R., and Artis, D. (2013). Commensal bacteria at the interface of host metabolism and the immune system. Nat. Immunol. *14*, 676–684.

Cani, P.D., Amar, J., Iglesias, M.A., Poggi, M., Knauf, C., Bastelica, D., Neyrinck, A.M., Fava, F., Tuohy, K.M., Chabo, C., et al. (2007). Metabolic endotoxemia initiates obesity and insulin resistance. Diabetes *56*, 1761–1772.

Carmody, R.N., Gerber, G.K., Luevano, J.M., Gatti, D.M., Somes, L., Svenson, K.L., and Turnbaugh, P.J. (2015). Diet dominates host genotype in shaping the murine gut microbiota. Cell Host Microbe *17*, 72–84.

Cebra, J.J. (1999). Influences of microbiota on intestinal immune system development. Am. J. Clin. Nutr. *69*, 1046s–1051s.

Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.-M., Kennedy, S., et al. (2013). Richness of human gut microbiome correlates with metabolic markers. Nature *500*, 541–546.

Churchill, G.A., Airey, D.C., Allayee, H., Angel, J.M., Attie, A.D., Beatty, J., Beavis, W.D., Belknap, J.K., Bennett, B., Berrettini, W., et al. (2004). The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat. Genet. *36*, 1133–1137.

Corrêa-Oliveira, R., Fachi, J.L., Vieira, A., Sato, F.T., and Vinolo, M.A.R. (2016). Regulation of immune cell function by short-chain fatty acids. Clin. Transl. Immunol. *5*, e73.

Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., and Knight, R. (2009). Bacterial Community Variation in Human Body Habitats Across Space and Time. Science (80-.). *326*, 1694–1697.

Cox, L.M., Yamanishi, S., Sohn, J., Alekseyenko, A.V., Leung, J.M., Cho, I., Kim, S.G., Li, H., Gao, Z., Mahana, D., et al. (2014). Altering the Intestinal Microbiota during a Critical Developmental Window Has Lasting Metabolic Consequences. Cell *158*, 705–721.

Crabbé, P.A., Bazin, H., Eyssen, H., and Heremans, J.F. (1968). The normal microbial flora as a major stimulus for proliferation of plasma cells synthesizing IgA in the gut. The germ-free intestinal tract. Int. Arch. Allergy Appl. Immunol. *34*, 362–375.

Crabbé, P.A., Nash, D.R., Bazin, H., Eyssen, H., and Heremans, J.F. (1970). Immunohistochemical observations on lymphoid tissues from conventional and germ-free mice. Lab. Invest. *22*, 448–457.

Creely, S.J., McTernan, P.G., Kusminski, C.M., Fisher, ff. M., Da Silva, N.F., Khanolkar, M., Evans, M., Harte, A.L., and Kumar, S. (2007). Lipopolysaccharide activates an innate immune system response in human adipose tissue in obesity and type 2 diabetes. Am. J. Physiol. Metab. *292*, E740–E747.

Davenport, E.R., Cusanovich, D.A., Michelini, K., Barreiro, L.B., Ober, C., and Gilad, Y. (2015). Genome-Wide Association Studies of the Human Gut Microbiota. PLoS One *10*, e0140301.

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A. V., Devlin, A.S., Varma, Y., Fischbach, M.A., et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome. Nature *505*, 559–563.

Devkota, S., Wang, Y., Musch, M.W., Leone, V., Fehlner-Peach, H., Nadimpalli, A., Antonopoulos, D.A., Jabri, B., and Chang, E.B. (2012). Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in Il10-/- mice. Nature *487*, 104–108.

Dill-McFarland, K., Tang, Z.-Z., Kemis, J., Kerby, R., Chen, G., Palloni, A., Sorenson, T., Rey, F., and Herd, P. (2018). Social relationships, social isolation, and the human gut microbiota. BioRxiv 428938.

Dimitriu, P.A., Boyce, G., Samarakoon, A., Hartmann, M., Johnson, P., and Mohn, W.W. (2013). Temporal stability of the mouse gut microbiota in relation to innate and adaptive immunity. Environ. Microbiol. Rep. *5*, 200–210.

Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., and Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc. Natl. Acad. Sci. *107*, 11971–11975.

Erwin G. Zoetendal, Antoon D. L. Ak, A.D.L.A.W.M.A.V.J.A.G.M. de V.W.M. de V. (2001). The Host Genotype Affects the Bacterial Community in the Human Gastronintestinal Tract. Microb. Ecol. Health Dis. *13*, 129–134.

Everard, A., Belzer, C., Geurts, L., Ouwerkerk, J.P., Druart, C., Bindels, L.B., Guiot, Y., Derrien, M., Muccioli, G.G., Delzenne, N.M., et al. (2013). Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. Proc. Natl. Acad. Sci. U. S. A. *110*, 9066–9071.

Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurilshikov, A., Bonder, M.J., Valles-Colomer, M., Vandeputte, D., et al. (2016). Population-level analysis of gut microbiome variation. Science *352*, 560–564.

Fernández, L., Langa, S., Martín, V., Maldonado, A., Jiménez, E., Martín, R., and Rodríguez, J.M. (2013). The human milk microbiota: Origin and potential roles in health and disease. Pharmacol. Res. *69*, 1–10.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., and Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc. Natl. Acad. Sci. *107*, 14691–14696.

Forslund, K., Hildebrand, F., Nielsen, T., Falony, G., Le Chatelier, E., Sunagawa, S., Prifti, E., Vieira-Silva, S., Gudmundsdottir, V., Krogh Pedersen, H., et al. (2015). Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature *528*, 262–266.

Furusawa, Y., Obata, Y., Fukuda, S., Endo, T.A., Nakato, G., Takahashi, D., Nakanishi, Y., Uetake, C., Kato, K., Kato, T., et al. (2013). Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. Nature *504*, 446–450.

Gao, Z., Yin, J., Zhang, J., Ward, R.E., Martin, R.J., Lefevre, M., Cefalu, W.T., and Ye, J. (2009). Butyrate improves insulin sensitivity and increases energy expenditure in mice. Diabetes *58*, 1509–1517.

Goodrich, J.K., Di Rienzi, S.C., Poole, A.C., Koren, O., Walters, W.A., Caporaso, J.G., Knight, R., and Ley, R.E. (2014a). Conducting a Microbiome Study. Cell *158*, 250–262.

Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J.T., et al. (2014b). Human genetics shape the gut microbiome. Cell *159*, 789–799.

Goodrich, J.K., Davenport, E.R., Beaumont, M., Jackson, M.A., Knight, R., Ober, C., Spector, T.D., Bell, J.T., Clark, A.G., and Ley, R.E. (2016). Genetic Determinants of the Gut Microbiome in UK Twins. Cell Host Microbe *19*, 731–743.

Gordon, H.A., and Bruckner-Kardoss, E. (1961). Effect of normal microbial flora on intestinal surface area. Am. J. Physiol. Content *201*, 175–178.

Greiner, T.U., Hyötyläinen, T., Knip, M., Bäckhed, F., and Orešič, M. (2014). The gut microbiota modulates glycaemic control and serum metabolite profiles in non-obese diabetic mice. PLoS One *9*, e110359.

Handelsman, Y. (2011). Role of bile acid sequestrants in the treatment of type 2 diabetes. Diabetes Care *34 Suppl 2*, S244-50.

Henao-Mejia, J., Elinav, E., Jin, C., Hao, L., Mehal, W.Z., Strowig, T., Thaiss, C.A., Kau, A.L., Eisenbarth, S.C., Jurczak, M.J., et al. (2012). Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. Nature *482*, 179–185.

Herd, P., Palloni, A., Rey, F., and Dowd, J.B. (2018). Social and population health science approaches to understand the human microbiome. Nat. Hum. Behav. 1.

Hill, M.J. (1997). Intestinal flora and endogenous vitamin synthesis. Eur. J. Cancer Prev. *6 Suppl 1*, S43-5.

Hooper, L. V., and Macpherson, A.J. (2010). Immune adaptations that maintain homeostasis with the intestinal microbiota. Nat. Rev. Immunol. *10*, 159–169.

Hooper, L. V, and Gordon, J.I. (2001). Commensal host-bacterial relationships in the gut. Science *292*, 1115–1118.

Hooper, L. V, Wong, M.H., Thelin, A., Hansson, L., Falk, P.G., and Gordon, J.I. (2001). Molecular analysis of commensal host-microbial relationships in the intestine. Science *291*, 881–884.

Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214.

Imhann, F., Bonder, M.J., Vich Vila, A., Fu, J., Mujagic, Z., Vork, L., Tigchelaar, E.F., Jankipersadsing, S.A., Cenit, M.C., Harmsen, H.J.M., et al. (2016). Proton pump inhibitors affect the gut microbiome. Gut *65*, 740–748.

Ivanov, I.I., Frutos, R. de L., Manel, N., Yoshinaga, K., Rifkin, D.B., Sartor, R.B., Finlay, B.B., and Littman, D.R. (2008). Specific microbiota direct the differentiation of IL-17-producing T-helper cells in the mucosa of the small intestine. Cell Host Microbe *4*, 337–349.

Jernberg, C., Lofmark, S., Edlund, C., and Jansson, J.K. (2010). Long-term impacts of antibiotic exposure on the human intestinal microbiota. Microbiology *156*, 3216–3223.

Jin, D., Wu, S., Zhang, Y.-G., Lu, R., Xia, Y., Dong, H., and Sun, J. (2015). Lack of Vitamin D Receptor Causes Dysbiosis and Changes the Functions of the Murine Intestinal Microbiome. Clin. Ther. *37*, 996–1009.e7.

Joyce, S.A., MacSharry, J., Casey, P.G., Kinsella, M., Murphy, E.F., Shanahan, F., Hill, C., and Gahan, C.G.M. (2014). Regulation of host weight gain and lipid metabolism by bacterial bile acid modification in the gut. Proc. Natl. Acad. Sci. U. S. A. *111*, 7421–7426.

Karlsson, F., Tremaroli, V., Nielsen, J., and Backhed, F. (2013a). Assessing the Human Gut Microbiota in Metabolic Diseases. Diabetes *62*, 3341–3349.

Karlsson, F.H., Tremaroli, V., Nookaew, I., Bergström, G., Behre, C.J., Fagerberg, B., Nielsen, J., and Bäckhed, F. (2013b). Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature *498*, 99–103.

Kasahara, K., Krautkramer, K.A., Org, E., Romano, K.A., Kerby, R.L., Vivas, E.I., Mehrabian, M., Denu, J.M., Bäckhed, F., Lusis, A.J., et al. (2018). Interactions between Roseburia intestinalis and diet modulate atherogenesis in a murine model. Nat. Microbiol. 1.

Kasubuchi, M., Hasegawa, S., Hiramatsu, T., Ichimura, A., and Kimura, I. (2015). Dietary gut microbial metabolites, short-chain fatty acids, and host metabolic regulation. Nutrients *7*, 2839–2849.

Kawamoto, S., Tran, T.H., Maruya, M., Suzuki, K., Doi, Y., Tsutsui, Y., Kato, L.M., and Fagarasan, S. (2012). The Inhibitory Receptor PD-1 Regulates IgA Selection and Bacterial Composition in the Gut. Science (80-.). *336*, 485–489.

Kelly, D., King, T., and Aminov, R. (2007). Importance of microbial colonization of the gut in early life to the development of immunity. Mutat. Res. Mol. Mech. Mutagen. *622*, 58–69.

Kolde, R., Franzosa, E.A., Rahnavard, G., Hall, A.B., Vlamakis, H., Stevens, C., Daly, M.J., Xavier, R.J., and Huttenhower, C. Host genetic variation and its microbiome interactions within the Human Microbiome Project.

Kovacs, A., Ben-Jacob, N., Tayem, H., Halperin, E., Iraqi, F.A., and Gophna, U. (2011). Genotype is a stronger determinant than sex of the mouse gut microbiota. Microb. Ecol. *61*, 423–428.

Krautkramer, K.A., Kreznar, J.H., Romano, K.A., Vivas, E.I., Barrett-Wilt, G.A., Rabaglia, M.E., Keller, M.P., Attie, A.D., Rey, F.E., and Denu, J.M. (2016). Diet-Microbiota Interactions Mediate Global Epigenetic Programming in Multiple Host Tissues. Mol. Cell *64*, 982–992.

Kreznar, J.H., Keller, M.P., Traeger, L.L., Rabaglia, M.E., Schueler, K.L., Stapleton, D.S., Zhao, W., Vivas, E.I., Yandell, B.S., Broman, A.T., et al. (2017). Host Genotype and Gut Microbiome Modulate Insulin Secretion and Diet-Induced Metabolic Phenotypes. Cell Rep. *18*, 1739–1750.

Kuipers, F., Bloks, V.W., and Groen, A.K. (2014). Beyond intestinal soap--bile acids in metabolic control. Nat. Rev. Endocrinol. *10*, 488–498.

Lamas, B., Richard, M.L., Leducq, V., Pham, H.-P., Michel, M.-L., Da Costa, G., Bridonneau, C., Jegou, S., Hoffmann, T.W., Natividad, J.M., et al. (2016). CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands. Nat. Med. *22*, 598–605.

Larsson, E., Tremaroli, V., Lee, Y.S., Koren, O., Nookaew, I., Fricker, A., Nielsen, J., Ley, R.E., and Bäckhed, F. (2012). Analysis of gut microbial regulation of host gene expression along the length of the gut and regulation of gut microbial ecology through MyD88. Gut *61*, 1124–1131.

Leamy, L.J., Kelly, S.A., Nietfeldt, J., Legge, R.M., Ma, F., Hua, K., Sinha, R., Peterson, D.A., Walter, J., Benson, A.K., et al. (2014). Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. Genome Biol. *15*, 552.

LeBlanc, J.G., Milani, C., de Giori, G.S., Sesma, F., van Sinderen, D., and Ventura, M. (2013). Bacteria as vitamin suppliers to their host: a gut microbiota perspective. Curr. Opin. Biotechnol. *24*, 160–168.

Ley, R.E., Bäckhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D., and Gordon, J.I. (2005). Obesity alters gut microbial ecology. Proc. Natl. Acad. Sci. U. S. A. *102*, 11070–11075.

Ley, R.E., Turnbaugh, P.J., Klein, S., and Gordon, J.I. (2006). Human gut microbes associated with obesity. Nature *444*, 1022–1023.

Li, T., and Chiang, J.Y.L. (2014). Bile Acid Signaling in Metabolic Disease and Drug Therapy. Pharmacol. Rev. *66*, 948–983.

Lin, H. V., Frassetto, A., Kowalik Jr, E.J., Nawrocki, A.R., Lu, M.M., Kosinski, J.R., Hubert, J.A., Szeto, D., Yao, X., Forrest, G., et al. (2012). Butyrate and Propionate Protect against Diet-Induced Obesity and Regulate Gut Hormones via Free Fatty Acid Receptor 3-Independent Mechanisms. PLoS One *7*, e35240.

Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., and Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. Nature *489*, 220–230.

Macpherson, A.J., and Harris, N.L. (2004). Interactions between commensal intestinal bacteria and the immune system. Nat. Rev. Immunol. *4*, 478–485.

Makishima, M., Lu, T.T., Xie, W., Whitfield, G.K., Domoto, H., Evans, R.M., Haussler, M.R., and Mangelsdorf, D.J. (2002). Vitamin D Receptor As an Intestinal Bile Acid Sensor. Science (80-.). *296*, 1313–1316.

McKnite, A.M., Perez-Munoz, M.E., Lu, L., Williams, E.G., Brewer, S., Andreux, P.A., Bastiaansen, J.W.M., Wang, X., Kachman, S.D., Auwerx, J., et al. (2012). Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. PLoS One *7*, e39191.

Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., González, A., Fontana, L., Henrissat, B., Knight, R., Gordon, J.I., Gonzalez, A., et al. (2011). Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science *332*, 970–974.

Natividad, J.M.M., and Verdu, E.F. (2013). Modulation of intestinal barrier by intestinal microbiota: Pathological and therapeutic implications. Pharmacol. Res. *69*, 42–51.

Ng, M.C.Y., Hester, J.M., Wing, M.R., Li, J., Xu, J., Hicks, P.J., Roh, B.H., Lu, L., Divers, J., Langefeld, C.D., et al. (2012). Genome-Wide Association of BMI in African Americans. Obesity *20*, 622–627.

Nicholson, J.K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., and Pettersson, S. (2012). Host-gut microbiota metabolic interactions. Science *336*, 1262–1267.

O'Connor, A., Quizon, P.M., Albright, J.E., Lin, F.T., and Bennett, B.J. (2014). Responsiveness of cardiometabolic-related microbiota to diet is influenced by host genetics. Mamm. Genome *25*, 583–599.

Olszak, T., An, D., Zeissig, S., Vera, M.P., Richter, J., Franke, A., Glickman, J.N., Siebert, R., Baron, R.M., Kasper, D.L., et al. (2012). Microbial Exposure During Early Life Has Persistent Effects on Natural Killer T Cell Function. Science (80-.). *336*, 489–493.

Org, E., Parks, B.W., Joo, J.W.J., Emert, B., Schwartzman, W., Kang, E.Y., Mehrabian, M., Pan, C., Knight, R., Gunsalus, R., et al. (2015). Genetic and environmental control of host-gut microbiota interactions. Genome Res. *25*, 1558–1569.

Osborn, O., and Olefsky, J.M. (2012). The cellular and signaling networks linking the immune system and metabolism in disease. Nat. Med. *18*, 363–374.

Parks, B.W., Nam, E., Org, E., Kostem, E., Norheim, F., Hui, S.T., Pan, C., Civelek, M., Rau, C.D., Bennett, B.J., et al. (2013). Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. Cell Metab. *17*, 141–152.

Perry, R.J., Peng, L., Barry, N.A., Cline, G.W., Zhang, D., Cardone, R.L., Petersen, K.F., Kibbey, R.G., Goodman, A.L., and Shulman, G.I. (2016). Acetate mediates a microbiome-brain-β-cell axis to promote metabolic syndrome. Nature *534*, 213–217.

Petersen, C., and Round, J.L. (2014). Defining dysbiosis and its influence on host immunity and disease. Cell. Microbiol. *16*, 1024–1033.

Petersson, J., Schreiber, O., Hansson, G.C., Gendler, S.J., Velcich, A., Lundberg, J.O., Roos, S., Holm, L., and Phillipson, M. (2011). Importance and regulation of the colonic mucus barrier in a mouse model of colitis. Am. J. Physiol. Liver Physiol. *300*, G327–G333.

Petra, A.I., Panagiotidou, S., Hatziagelaki, E., Stewart, J.M., Conti, P., and Theoharides, T.C. (2015). Gut-Microbiota-Brain Axis and Its Effect on Neuropsychiatric Disorders With Suspected Immune Dysregulation. Clin. Ther. *37*, 984–995.

Priyadarshini, M., Villa, S.R., Fuller, M., Wicksteed, B., Mackay, C.R., Alquier, T., Poitout, V., Mancebo, H., Mirmira, R.G., Gilchrist, A., et al. (2015). An Acetate-Specific GPCR, FFAR2, Regulates Insulin Secretion. Mol. Endocrinol. *29*, 1055–1066.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. Nature *464*, 59–65.

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature *490*, 55–60.

Rehman, A., Sina, C., Gavrilova, O., Hasler, R., Ott, S., Baines, J.F., Schreiber, S., and Rosenstiel, P. (2011). Nod2 is essential for temporal development of intestinal microbial communities. Gut *60*, 1354–1362.

Rehman, A., Rausch, P., Wang, J., Skieceviciene, J., Kiudelis, G., Bhagalia, K., Amarapurkar, D., Kupcinskas, L., Schreiber, S., Rosenstiel, P., et al. (2016). Geographical patterns of the standing and active human gut microbiome in health and IBD. Gut *65*, 238–248.

Reinhardt, C., Bergentall, M., Greiner, T.U., Schaffner, F., Östergren-Lundén, G., Petersen, L.C., Ruf, W., and Bäckhed, F. (2012). Tissue factor and PAR1 promote microbiota-induced intestinal vascular remodelling. Nature *483*, 627–631.

Renz, H., Brandtzaeg, P., and Hornef, M. (2012). The impact of perinatal immune development on mucosal homeostasis and chronic inflammation. Nat. Rev. Immunol. *12*, 9–23.

Reunanen, J., Kainulainen, V., Huuskonen, L., Ottman, N., Belzer, C., Huhtinen, H., de Vos, W.M., and Satokari, R. (2015). Akkermansia muciniphila Adheres to Enterocytes and Strengthens the Integrity of the Epithelial Cell Layer. Appl. Environ. Microbiol. *81*, 3655–3662.

Ridaura, V.K., Faith, J.J., Rey, F.E., Cheng, J., Duncan, A.E., Kau, A.L., Griffin, N.W., Lombard, V., Henrissat, B., Bain, J.R., et al. (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science *341*, 1241214.

Ridlon, J.M., Kang, D.-J., and Hylemon, P.B. (2006). Bile salt biotransformations by human intestinal bacteria. J. Lipid Res. *47*, 241–259.

Ridlon, J.M., Harris, S.C., Bhowmik, S., Kang, D.-J., and Hylemon, P.B. (2016). Consequences of bile salt biotransformations by intestinal bacteria. Gut Microbes *7*, 22–39.

Rietschel, E.T., Schletter, J., Weidemann, B., EL-Samalouti, V., Mattern, T., Zahringer, U., Seydel, U., Brade, H., Flad, H.-D., Kusumoto, S., et al. (1998). Lipopolysaccharide and Peptidoglycan: CD14-Dependent Bacterial Inducers of Inflammation. Microb. Drug Resist. *4*, 37–44.

Rogier, E.W., Frantz, A.L., Bruno, M.E.C., Wedlund, L., Cohen, D.A., Stromberg, A.J., and Kaetzel, C.S. (2014). Secretory antibodies in breast milk promote long-term intestinal homeostasis by regulating the gut microbiota and host gene expression. Proc. Natl. Acad. Sci. *111*, 3074–3079.

Romano, K.A., Martinez-Del Campo, A., Kasahara, K., Chittim, C.L., Vivas, E.I., Amador-Noguez, D., Balskus, E.P., and Rey, F.E. (2017). Metabolic, Epigenetic, and Transgenerational Effects of Gut Bacterial Choline Consumption. Cell Host Microbe *22*, 279–290.e7.

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P.I., Godneva, A., Kalka, I.N., Bar, N., et al. (2018). Environment dominates over host genetics in shaping human gut microbiota. Nature *555*, 210–215.

Round, J.L., and Mazmanian, S.K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. Nat. Rev. Immunol. *9*, 313–323.

Russell, D.W. (2009). Fifty years of advances in bile acid synthesis and metabolism. J. Lipid Res. *50 Suppl*, S120-5.

Rutayisire, E., Huang, K., Liu, Y., and Tao, F. (2016). The mode of delivery affects the diversity and colonization pattern of the gut microbiota during the first year of infants' life: a systematic review. BMC Gastroenterol. *16*, 86.

Ryan, K.K., Tremaroli, V., Clemmensen, C., Kovatcheva-Datchary, P., Myronovych, A., Karns, R., Wilson-Pérez, H.E., Sandoval, D.A., Kohli, R., Bäckhed, F., et al. (2014). FXR is a molecular target for the effects of vertical sleeve gastrectomy. Nature *509*, 183–188.

Schnorr, S.L., Candela, M., Rampelli, S., Centanni, M., Consolandi, C., Basaglia, G., Turroni, S., Biagi, E., Peano, C., Severgnini, M., et al. (2014). Gut microbiome of the Hadza hunter-gatherers. Nat. Commun. *5*, 3654.

Sender, R., Fuchs, S., and Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol. *14*, e1002533.

Sharma, R., Schumacher, U., Ronaasen, V., and Coates, M. (1995). Rat intestinal mucosal responses to a microbial flora and different diets. Gut *36*, 209–214.

Sjögren, K., Engdahl, C., Henning, P., Lerner, U.H., Tremaroli, V., Lagerquist, M.K., Bäckhed, F., and Ohlsson, C. (2012). The gut microbiota regulates bone mass in mice. J. Bone Miner. Res. *27*, 1357–1367.

Smith, K., McCoy, K.D., and Macpherson, A.J. (2007). Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. Semin. Immunol. *19*, 59–69.

Snijders, A.M., Langley, S.A., Kim, Y.-M., Brislawn, C.J., Noecker, C., Zink, E.M., Fansler, S.J., Casey, C.P., Miller, D.R., Huang, Y., et al. (2017). Influence of early life exposure, host genetics and diet on the mouse gut microbiome and metabolome. Nat. Microbiol. *2*, 16221.

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L.G., Gratadoux, J.-J., Blugeon, S., Bridonneau, C., Furet, J.-P., Corthier, G., et al. (2008). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proc. Natl. Acad. Sci. *105*, 16731–16736.

Sommer, F., and Bäckhed, F. (2013). The gut microbiota--masters of host development and physiology. Nat. Rev. Microbiol. *11*, 227–238.

Song, S.J., Lauber, C., Costello, E.K., Lozupone, C.A., Humphrey, G., Berg-Lyons, D., Caporaso, J.G., Knights, D., Clemente, J.C., Nakielny, S., et al. (2013). Cohabiting family members share microbiota with one another and with their dogs. Elife *2*, e00458.

Spor, A., Koren, O., and Ley, R. (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. Nat. Rev. Microbiol. *9*, 279–290.

Suhir, H., and Etzioni, A. (2010). The Role of Toll-Like Receptor Signaling in Human Immunodeficiencies. Clin. Rev. Allergy Immunol. *38*, 11–19.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. Nature *444*, 1027–1031.

Turnbaugh, P.J., Bäckhed, F., Fulton, L., and Gordon, J.I. (2008). Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. Cell Host Microbe *3*, 213–223.

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., et al. (2009a). A core gut microbiome in obese and lean twins. Nature *457*, 480–484.

Turnbaugh, P.J., Ridaura, V.K., Faith, J.J., Rey, F.E., Knight, R., and Gordon, J.I. (2009b). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. Sci. Transl. Med. *1*, 6ra14-6ra14.

Turpin, W., Espin-Garcia, O., Xu, W., Silverberg, M.S., Kevans, D., Smith, M.I., Guttman, D.S., Griffiths, A., Panaccione, R., Otley, A., et al. (2016). Association of host genome with intestinal microbial composition in a large healthy cohort. Nat. Genet.

Tyakht, A. V., Kostryukova, E.S., Popenko, A.S., Belenikin, M.S., Pavlenko, A. V., Larin, A.K., Karpova, I.Y., Selezneva, O. V., Semashko, T.A., Ospanova, E.A., et al. (2013). Human gut microbiota community structures in urban and rural populations in Russia. Nat. Commun. *4*, 2469.

Vaahtovuo, J., Munukka, E., Korkeamäki, M., Luukkainen, R., and Toivanen, P. (2008). Fecal microbiota in early rheumatoid arthritis. J. Rheumatol. *35*, 1500–1505.

Vijay-Kumar, M., Aitken, J.D., Carvalho, F.A., Cullender, T.C., Mwangi, S., Srinivasan, S., Sitaraman, S. V, Knight, R., Ley, R.E., and Gewirtz, A.T. (2010). Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. Science *328*, 228–231.

Vrieze, A., Holleman, F., Zoetendal, E.G., de Vos, W.M., Hoekstra, J.B.L., and Nieuwdorp, M. (2010). The environment within: how gut microbiota may influence metabolism and body composition. Diabetologia *53*, 606–613.

Vuong, H.E., Yano, J.M., Fung, T.C., and Hsiao, E.Y. (2017). The Microbiome and Host Behavior. Annu. Rev. Neurosci. *40*, 21–49.

Wang, J., Kalyan, S., Steck, N., Turner, L.M., Harr, B., Künzel, S., Vallier, M., Häsler, R., Franke, A., Oberg, H.-H., et al. (2015). Analysis of intestinal microbiota in hybrid house mice reveals evolutionary divergence in a vertebrate hologenome. Nat. Commun. *6*, 6440.

Wang, J., Thingholm, L.B., Skiecevičienė, J., Rausch, P., Kummen, M., Hov, J.R., Degenhardt, F., Heinsen, F.-A., Rühlemann, M.C., Szymczak, S., et al. (2016). Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. Nat. Genet. *48*, 1396–1406.

Wang, Z., Klipfell, E., Bennett, B.J., Koeth, R., Levison, B.S., Dugar, B., Feldstein, A.E., Britt, E.B., Fu, X., Chung, Y.-M., et al. (2011). Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature *472*, 57–63.

Wen, L., Ley, R.E., Volchkov, P.Y., Stranges, P.B., Avanesyan, L., Stonebraker, A.C., Hu, C., Wong, F.S., Szot, G.L., Bluestone, J.A., et al. (2008). Innate immunity and intestinal microbiota in the development of Type 1 diabetes. Nature *455*, 1109–1113.

Wostmann, B.S. (1981). The Germfree Animal in Nutritional Studies. Annu. Rev. Nutr. *1*, 257–279.

Xie, H., Guo, R., Zhong, H., Feng, Q., Lan, Z., Qin, B., Ward, K.J., Jackson, M.A., Xia, Y., Chen, X., et al. (2016). Shotgun Metagenomics of 250 Adult Twins Reveals Genetic and Environmental Impacts on the Gut Microbiome. Cell Syst. *3*, 572–584.e3.

Yao, L., Craven Seaton, S., Ndousse-Fetter, S., Adhikari, A.A., Dibenedetto, N., Mina, A.I., Banks, A.S., Bry, L., and Devlin, A.S. (2018). A selective gut bacterial bile salt hydrolase alters host metabolism.

Yassour, M., Vatanen, T., Siljander, H., Hämäläinen, A.-M., Härkönen, T., Ryhänen, S.J., Franzosa, E.A., Vlamakis, H., Huttenhower, C., Gevers, D., et al. (2016). Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. Sci. Transl. Med. *8*, 343ra81.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2012). Human gut microbiome viewed across age and geography. Nature *486*, 222–227.

Zhernakova, A., Kurilshikov, A., Bonder, M.J., Tigchelaar, E.F., Schirmer, M., Vatanen, T., Mujagic, Z., Vila, A.V., Falony, G., Vieira-Silva, S., et al. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science *352*, 565–569.

Zhou, H., and Hylemon, P.B. (2014). Bile acids are nutrient signaling hormones. Steroids *86*, 62–68.

Zhou, Y., Qin, H., Zhang, M., Shen, T., Chen, H., Ma, Y., Chu, Z., Zhang, P., and Liu, Z. (2010). Lactobacillus plantarum inhibits intestinal epithelial barrier dysfunction induced by unconjugated bilirubin. Br. J. Nutr. *104*, 390–401.

Zivkovic, A.M., German, J.B., Lebrilla, C.B., and Mills, D.A. (2011). Human milk glycobiome and its impact on the infant gastrointestinal microbiota. Proc. Natl. Acad. Sci. *108*, 4653–4658.

CHAPTER 2: Host Genotype and Gut Microbiome Modulate Insulin Secretion and Diet-Induced Metabolic Phenotypes

The work presented in this chapter has been published:

Kreznar JH[*], *Keller MP*[*], *Traeger LL, Rabaglia ME, Schueler KL, Stapleton DS, Zhao W, Vivas EI, Yandell BS, Broman AT, Hagenbuch B, Attie AD*[+], *and Rey FE*[+]

[*] *indicates lead author,* [+] *indicates corresponding author*

Data and supplemental tables available online.

ABSTRACT

Genetic variation drives phenotypic diversity and influences the predisposition to metabolic disease. Here, we characterize the metabolic phenotypes of eight genetically distinct inbred mouse strains in response to a high-fat/high-sucrose diet. We found significant variation in diabetes-related phenotypes and gut microbiota composition among the different mouse strains in response to the dietary challenge and identified taxa associated with these traits. Follow-up microbiota transplant experiments showed that altering the composition of the gut microbiota modifies strain-specific susceptibility to diet-induced metabolic disease. Animals harboring microbial communities with enhanced capacity for processing dietary sugars and for generating hydrophobic bile acids showed increased susceptibility to metabolic disease. Notably, differences in glucose-stimulated insulin secretion between different mouse strains were partially recapitulated via gut microbiota transfer. Our results suggest that the gut microbiome contributes to the genetic and phenotypic diversity observed among mouse strains and provide a link between the gut microbiome and insulin secretion.

INTRODUCTION

The intestinal microbiota exerts a profound influence on development, physiology and health (Clemente et al. 2012; Sommer & Bäckhed 2013; Tremaroli & Bäckhed 2012). Although there is substantial interpersonal variation in the composition of the gut microbiota among unrelated healthy subjects, sequencing studies have revealed distal gut community patterns associated with different pathological states, including obesity and diabetes (Ridaura et al. 2013; Qin et al. 2012; Karlsson et al. 2013). Remarkably, alterations in the intestinal microbiota composition have been shown to modulate insulin sensitivity (Vrieze et al. 2010) —a key feature in metabolic disease and type 2 diabetes (T2D), and thus play a role in diabetes susceptibility. s

Dietary components that are not efficiently absorbed in the proximal intestine reach the distal gut where they are metabolized by gut microbes. Intestinal microbes impact our health in part by generating numerous metabolites from our diet. Short-chain fatty acids (SCFA), mainly acetate, propionate and butyrate, are produced through bacterial fermentation of dietary carbohydrates. SCFA serve as energy and signaling molecules in the intestine and peripheral organs (Besten et al. 2013). Specifically, SCFA are important regulators of both energy and glucose homeostasis (Besten et al. 2013; Koh et al. 2016). For example, butyrate improves insulin sensitivity (Gao et al. 2009; Hartstra et al. 2015) and T2D patients have reduced levels of butyrate-producing bacteria (Qin et al. 2012). Additionally, acetate modulates insulin secretion from β-cells (Priyadarshini et al. 2015; Perry et al. 2016). While primarily associated with metabolic benefits, increased concentrations of butyrate and acetate have been found in the cecum of obese mice, suggesting an increased ability of the microbiome to harvest energy from the diet (Turnbaugh et al. 2006).

Gut microbes also impact host physiology by modifying bile acids (BA) synthesized by the host (Houten et al. 2006; Kuipers et al. 2014; Ryan et al. 2014; Sayin et al. 2013). In addition to their role in emulsifying lipids, BA function as hormones through their ability to activate nuclear hormone receptors (D. J. Parks et al. 1999) and G-coupled protein receptors (Kawamata et al. 2003). They modulate glucose homeostasis, lipid metabolism, energy expenditure, and intestinal motility (Kuipers et al. 2014). Primary BA are synthesized from cholesterol in the liver (Russell 2009), stored in the gallbladder, and secreted into the duodenum upon ingestion of a meal. The gut microbiota catalyzes the production of secondary BA via deconjugation, dehydrogenation, epimerization, and dehydroxylation of primary BA (Ridlon et al. 2006). BA with different modifications vary in their ability to activate receptors and affect host physiology (Makishima et al. 1999; Kuipers et al. 2014). Subjects with T2D have altered circulating BA profiles. Treatment of T2D subjects with compounds that increase fecal excretion of BA and modify BA composition improves their glycemic status (Handelsman 2011).

Mouse genetics can be employed to explore the relationships between diet, host genetics, and metabolic responses (O'Connor et al. 2014; B. W. Parks et al. 2013; Ussar et al. 2015). The Collaborative Cross (CC) is a systems genetics mouse resource that consists of a panel of recombinant inbred lines and an outbred stock derived from eight genetically diverse founder strains. These include five classical inbred strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ and NZO/HILtJ), and three wild-derived strains (CAST/EiJ, PWK/PhJ, WSB/EiJ) (Churchill et al. 2004; Roberts et al. 2007; Aylor et al. 2011).

We examined the metabolic phenotypes and gut microbiota composition of the eight CC founder strains in response to chronic consumption of two defined diets: a high-fat/high-sucrose diet (HF/HS) and a control diet. We found remarkable variation in diabetes-related phenotypes

and gut microbiota composition as a function of host genotype and diet, and we identified bacterial taxa that correlate with metabolic traits, including body weight, glucose, and insulin levels. Germ-free (GF) mice were colonized with microbiota derived from two founder strains that exhibited divergent metabotypes, C57BL/6J and CAST/EiJ. The transplanted animals were maintained on the HF/HS diet and then subjected to metabolomic and metagenomic analyses. We identified functional differences attributable to the two transplanted microbial communities, including insulin secretion responses and susceptibility to diet-induced metabolic disease.

RESULTS

Host metabolic responses to diet are influenced by genetic background

We assessed the variability of diet-induced metabolic responses of the eight genetically diverse CC founder strains: A/J, C57BL/6J (B6), 129S1/SvImJ (129), NOD/ShiLtJ (NOD), NZO/HILtJ (NZO), CAST/EiJ (CAST), PWK/PhJ (PWK), WSB/EiJ (WSB). All mice were obtained from the Jackson Laboratory, maintained in the same vivarium and fed the same diet, so that the only known difference among the strains is genetics. We placed four-week-old male mice from each strain on either a control or a high-fat high-sucrose (HF/HS) diet for 22 weeks (Table S1).

The CC founder strains displayed a wide range of body weight and metabolic responses to the dietary challenge (Figure 2.1 and S2.1). Two-way ANOVA analysis of the clinical traits revealed a significant strain effect for fasting insulin ($F = 14.94$, $p < 0.0001$). We also observed significant strain-diet interactions for body weight ($F = 3.19$, $p < 0.01$) and fasting glucose ($F = 2.81$, $p < 0.01$). Significant strain and diet effects were also seen for hepatic triglyceride content ($F = 10.96$, $p < 0.0001$; $F = 11.92$, $p < 0.001$, respectively) effects. Liver triglyceride content showed high inter-strain variation, with 129 having the most significant response to diet ($p < 0.05$) (Figure 2.1D). NZO mice were the only strain to become overtly diabetic (glucose levels >300 mg/dl) as a consequence of HF/HS feeding. With the exception of NZO mice, which did not survive past 18 weeks on the HF/HS diet, B6 mice were the most responsive to diet. HF/HS-fed B6 mice became obese ($p < 0.01$) and developed insulin resistance and glucose intolerance after ~8 weeks (Figure 2.1A and S2.1A-C). In addition to differences in diet responsiveness, the strains varied in both absolute levels of insulin and change in insulin levels over time, suggesting a significant divergence in insulin sensitivity among the strains (Figure S2.1B).

To assess whole-body glucose homeostasis and more directly evaluate the underlying role of the pancreatic islets in the control of plasma insulin, we measured plasma glucose and insulin during an oral glucose tolerance test (oGTT). Both plasma glucose and insulin during the oGTT varied dramatically between the strains. We computed the area under the curve (AUC) for each trait to determine the overall excursion in glucose and insulin that occurred during the oGTT (Figure 2.1E-F and S2.2). We observed a wide inter-strain range of responses in plasma insulin during the oGTT ($F = 12.84$, $p < 0.0001$) (Figure 2.1F and 2.2B). Changes in plasma insulin may reflect altered insulin secretion from β-cells, peripheral insulin resistance, reduced insulin clearance, or any combination thereof. 129 and WSB showed diet-induced glucose intolerance, but minimal changes in their insulin response during the oGTT (Figure 2.1E-F and S2.2A), suggesting that their glucose intolerance may be driven by altered insulin secretion and/or enhanced insulin clearance. Remarkably, insulin secretion and glucose tolerance were completely unaffected by the HF/HS diet in CAST. Furthermore, the kinetics of the glucose and insulin responses were more rapid in CAST than in all other strains (Figure S2.2), suggesting that CAST mice may employ different pathways underlying glucose-stimulated insulin secretion and whole-body glucose disposal.

Diet and host genotype influence microbiota composition

Gut microbes influence the development of metabolic disease. We characterized the cecal microbiomes of the eight CC founder strains by 16S rRNA sequencing. We compared the cecal microbiomes employing UniFrac, a phylogenetic distance metric used to measure differences in bacterial community structure (Lozupone & Knight 2005). Principal coordinates analysis (PCoA) of 16S rRNA unweighted UniFrac distances revealed a strong influence of strain (PERMANOVA, $p < 0.001$) and diet (PERMANOVA, $p < 0.001$) on microbial community composition (Figure

S2.3A). Consistent with previous studies, the effect of diet on gut microbial composition varied among the strains (O'Connor et al. 2014; B. W. Parks et al. 2013; Carmody et al. 2015), where B6, CAST and NOD mice showed the greatest microbiome response to diet (Figure S2.3A).

We detected eight bacterial phyla among the mice (Figure S2.3B). Bacteroidetes and Firmicutes dominated the gut of all strains on either diet, accounting for >90% of the sequenced reads. As reported by other studies, we observed a decrease in the Bacteroidetes:Firmicutes ratio and an increase in Proteobacteria in the HF/HS-fed mice (Ley et al. 2005; Hildebrandt et al. 2009). In fact, Proteobacteria showed the greatest fold change in abundance in response to diet: HF/HS feeding caused an average 5.4-fold change ($p < 0.0001$), although the relative increase varied among strains.

Microbial taxa correlate with metabolic phenotypes

To determine whether strain-dependent variability in microbiota composition was associated with the dramatic differences in the diabetes-related clinical traits, we computed Pearson's correlations between abundance of family-level taxa and the metabolic traits among the 8 CC founder mice (Figure 2.2A). We focused our analysis on families that were present in at least 7 of the founder strains. Bacteroidaceae was among the most negatively correlated with several metabolic phenotypes, including body weight, fasting plasma insulin and $AUC_{insulin}$ during the oGTT. The Bacteroidaceae family belongs to the Bacteroidetes phylum and is typically found at higher levels in fecal samples of lean *vs.* obese individuals (Ley et al. 2005; Turnbaugh et al. 2009). Conversely, Clostridiaceae and Rikenellaceae showed the strongest positive correlations with plasma insulin levels. Our analysis also identified strong positive correlations between fasting plasma glucose and the Streptococcaceae and Desulfovibrionaceae families. Members of these

families have previously been shown to be enriched in the fecal microbiome of patients with T2D (Qin et al. 2012; Karlsson et al. 2013).

Some of the correlations mentioned above varied significantly as a function of host diet and strain (Table S2). For example, the negative correlation observed between fasting insulin levels and Bacteroidaceae had a significant strain effect ($p < 0.0001$). We also observed a slight diet effect ($p < 0.001$), which is likely driven by the low abundance and high fasting insulin levels in the chow-fed NZO mice (Figure 2.2B). We also observed a significant diet effect for the relationship between Clostridiaceae and fasting insulin levels ($p < 0.05$), but there was also a strain difference that seems to be driven by NZO on chow diet ($p < 0.001$) (Figure 2.2C).

These results suggest that diet and genetic background are major determinants of gut microbial composition and metabolic disease. However, the relative contributions of host genetic variance *vs.* microbial-derived genetic variation across different mouse strains in the development of diet-induced metabolic phenotypes remain largely unknown.

The gut microbiome is a source of genetic variation that influences host-associated differences in diet-induced metabotypes

To directly test the influence of gut microbes on the metabolic phenotypes observed among the founder strains, we performed cecal transplants into germ-free B6 (B6-GF) hosts, leveraging two CC founder strains that showed disparate responsiveness to the HF/HS diet. The B6 strain became obese, insulin resistant, and glucose intolerant, whereas the CAST strain remained lean and insulin-sensitive despite HF/HS feeding (Figure 2.1).

As mentioned above, B6 and CAST mice had significantly different intestinal microbiota (PERMANOVA, $F = 4.86$, $p < 0.001$) (Figure S2.3A). B6 mice harbored a significantly greater abundance of microbial families with strong positive correlations with metabolic traits, such as

weight and insulin (i.e. Clostridiaceae, $p < 0.05$), while CAST mice had a greater representation of families with significant negative correlations (i.e. Bacteroidaceae, $p < 0.01$) (Figure 2.2A and S2.3C).

We transplanted cecal microbiota from either conventionally-raised B6 (B6-CR) or CAST (CAST-CR) donor mice into 9-week-old B6-GF recipient mice, to yield $B6_{B6}$ or $B6_{CAST}$ mice, respectively. Transplanted animals were housed by treatment group in separate vinyl gnotobiotic isolators and maintained on a HF/HS diet for 16 weeks following colonization (Figure 2.3A). A dietary treatment of 16 weeks allows robust development of metabolic phenotypes associated with consumption of HF/HS diet.

Recipient mice recapitulated microbial and metabolic phenotypes observed in the respective donor strains (Figure 2.3 and 2.4). $B6_{B6}$ mice gained ~25% more weight, had larger epididymal fat pad mass and showed greater hepatic triglyceride accumulation than $B6_{CAST}$ mice (Figure 2.3). Additionally, oGTT revealed that while the plasma glucose levels resulting from an orally administered bolus of glucose did not significantly differ between the two groups of transplanted mice (Figure 2.3E), the insulin responses were dramatically different (Figure 2.3F). The glucose challenge evoked a much larger insulin response in $B6_{B6}$ mice than in $B6_{CAST}$ mice. The low insulin response in $B6_{CAST}$ mice resembled the insulin response of the CAST-CR donors (Figure 2.1F and S2.1F). These results suggest that the effectiveness of insulin to maintain euglycemia was greater in the mice receiving the CAST microbiota than in mice receiving the B6 microbiota (Figure 2.3E-F).

16S rRNA gene profiling of the donor cecal inoculum and transplant recipient fecal samples show that recipient mice were successfully colonized with the donor's microbiota. $B6_{B6}$ and $B6_{CAST}$ mice assumed a phylogenetically similar composition to that of their respective donors

as confirmed by PCoA of unweighted UniFrac distances (Figure 2.4A). As seen in the founders, Bacteroidetes and Firmicutes comprised ~90% of the microbiome, although the abundance of Firmicutes was higher in $B6_{B6}$ ($p < 0.05$) (Figure 2.4B). We identified taxonomic differences in the microbiota composition between the two recipient groups using linear discriminant analysis (LDA) effect size (LEfSe) with LDA score >2 (Segata et al. 2011). We found 20 microbial families that were differentially enriched in the fecal microbiota of $B6_{B6}$ versus $B6_{CAST}$ mice. There were 12 microbial families that were enriched in $B6_{B6}$, of which 7 belonged to the Firmicutes phyla (Figure 2.4C). Some of the families differentially represented in the transplanted animals overlap with taxa that are significantly correlated with metabolic phenotypes in the founder strains (Figure 2.2). Notably, $B6_{B6}$ mice exhibited higher levels of Clostridiaceae ($p < 0.01$), which is positively associated with insulin secretion in the founder strains (Figure 2.2), whereas $B6_{CAST}$ mice had higher levels of Bacteroidaceae ($p < 0.01$), which is negatively associated with body weight and insulin secretion (Figure 2.2). These results are concordant with the metabolic phenotypes observed in the transplanted mice and suggest that the distinct microbial gut communities influence metabolic changes evoked by HF/HS feeding, including insulin secretion,

We characterized the functional potential of transplanted communities by sequencing and analyzing their metagenomes. Metagenomic analysis of the same samples further validated that the B6 and CAST-derived microbiota were distinct from one another, with donors clustering with their respective transplant recipients (Figure 2.4D). We identified several thousand genes differentially represented between the B6 and CAST microbiota (Table S4). This metagenomic analysis also revealed microbial functions that were enriched in each transplanted microbial community (Table S5). The most enriched microbial pathways in $B6_{B6}$ mice included genes involved in membrane transport, and carbohydrate and lipid metabolism (Figure 2.4E). For

example, the ABC transporters and phosphotransferase system (PTS) pathways were enriched in mice colonized with the B6 microbiota ($p < 0.01$). PTS are a class of transport systems involved in the uptake and phosphorylation of a variety of carbohydrates that can be subsequently fermented to SCFA (Deutscher et al. 2006). It has been previously reported that diet-induced obese mice have a concomitant enrichment of microbial pathways involved in PTS and elevated concentrations of SCFA (Turnbaugh et al. 2008), reflecting an increased capacity for energy harvest. Consistent with these results, targeted GC/MS analysis of SCFA in cecal contents disclosed that $B6_{B6}$ mice had an increased concentration of the major fermentation end-products, compared with $B6_{CAST}$ (Figure 2.4F). Conversely, $B6_{CAST}$ microbiota were enriched in genes related to the vitamin B12 (cobalamin) biosynthetic pathway (Figure S2.4A), synthesis of other B vitamins and enzyme co-factors, as well as lipopolysaccharide (LPS) biosynthesis (Figure 2.4E and S2.4B). A difference in LPS biosynthetic potential may reflect the composition of the $B6_{CAST}$ microbiota, which has a significantly higher relative abundance of gram negative Bacteroidetes than the $B6_{B6}$ microbiota (Figure S2.3B). Our findings mirror those described previously in T2D patients relative to diabetes-free control patients (Qin et al. 2012; Karlsson et al. 2013)—both the microbiota of T2D patients and our metabolically-diseased mice with B6 microbiota show enrichment in KEGG pathways involved in membrane transport, while diabetes-free patients and mice with the CAST microbiota exhibit enrichment in vitamin and co-factor biosynthesis.

B6 and CAST microbiota produce divergent bile acid profiles

Gut microbes impact host physiology in part by modulating the composition of the BA pool. We determined fecal BA profiles of the transplanted mice and HF/HS-fed B6-CR and CAST-CR mice by UPLC/MS-based quantification of primary and the most abundant secondary BA. The BA composition of $B6_{B6}$ mice closely resembled that of B6-CR donor mice, whereas $B6_{CAST}$

exhibited a BA profile that was intermediate between CAST-CR and B6-CR mice (Figure 2.5A). Microbiota composition was also a significant predictor of BA composition. Bray-Curtis dissimilarity-based PCA revealed clustering of the BA profiles by microbiota composition.

Although the $B6_{CAST}$ microbiota composition resembled that of CAST-CR (Figure 2.4A), there were significant differences in BA profiles between these groups, suggesting that variation in circulating BA is under the control of both host genetics and gut microbiota. For example, the primary BA cholic acid (CA), chenodeoxycolic acid (CDCA) and α-muricholic acid (α-MCA) were significantly higher in CAST-CR mice compared to $B6_{CAST}$ mice ($p < 0.01$, $p < 0.05$, $p < 0.01$, respectively) (Figure 2.5B). Moreover, taurine-conjugated muricholic acids (MCAs) were significantly higher in CAST-CR mice compared with $B6_{CAST}$ mice. In contrast, these differences in taurine conjugation were not present between B6-CR and $B6_{B6}$ mice. Taurine conjugation of MCAs is a host process (Ridlon et al. 2006), further highlighting the interaction of host genetics and microbiome in modulating host BA profiles.

B6-CR and $B6_{B6}$ mice had a significantly greater representation of hydrophobic BA species (e.g., deoxycholic acid, lithocholic acid (Figure 2.5B-C)), which are elevated in humans and mice with insulin resistance (Ryan et al. 2014; Prawitt et al. 2011). Microbial metabolism of bile acids generally leads to a more hydrophobic bile acid pool, which facilitates fecal elimination of bile acids. Bile salt hydrolases (BSH) are involved in the hydrolysis of conjugated BA, a necessary step for the production of secondary BA. Consistent with the results presented above, there were a higher number of distinct BSH genes in the B6 microbiota relative to CAST microbiota (13 annotated BSH genes highly abundant in the B6 microbiota relative to CAST *vs.* two annotated BSH genes highly abundant in the CAST microbiota relative to B6, Table S4). Furthermore, the two groups of recipient mice had vastly different fecal BA profiles. Chenodeoxycholic acid

(CDCA; $p < 0.05$), deoxycholic acid (DCA; $p < 0.01$), lithocholic acid (LCA; $p < 0.01$), ω-muricholic acid (ωMCA; $p < 0.05$), and tauro-ω-muricholic acid (TωMCA; $p < 0.05$) were all significantly higher in $B6_{B6}$ than in $B6_{CAST}$ (Figure 2.5B). DCA was the most abundant BA species in $B6_{B6}$ mice, and was also ~5-fold more abundant in B6-CR *vs.* CAST-CR mice. DCA contributes to microbial dysbiosis, a hallmark of metabolic disease, and is positively associated with higher levels of Firmicutes (Islam et al. 2011). Tauroursodeoxycholic acid (TUDCA) was >2-fold higher in CAST-CR mice compared to the transplanted animals, but was not detected in B6-CR mice. Interestingly, administration of TUDCA has been shown to decrease hepatic steatosis and improve insulin resistance in genetically obese mice (Kars et al. 2010; Ozcan et al. 2006), suggesting a potential protective role. These results reveal differences in BA profiles linked to both host genotype and gut microbial composition. They also suggest that the differential responses to prolonged HF/HS diet consumption between B6 and CAST mice could be mediated at least in part by differences in microbial BA metabolism.

Gut microbiota influences insulin secretion

The most dramatic phenotype difference we observed between $B6_{B6}$ and $B6_{CAST}$ mice was in insulin secretion, where $B6_{CAST}$ mice had a blunted insulin response during the oGTT (Figure 2.3E). This attenuated response in $B6_{CAST}$ mice may also reflect low insulin secretion from β-cells and/or increased insulin clearance. To determine whether the differential insulin response during the oGTT in the $B6_{B6}$ *vs.* $B6_{CAST}$ mice resulted from altered insulin secretion, we performed *ex vivo* insulin secretion assays on isolated islets. Islets were harvested from B6-GF mice 1 month after successful colonization with either CAST-CR or B6-CR cecum-derived microbiota (Figure S2.5).

The isolated islets partially recapitulated the reduced insulin secretion observed in the CAST-colonized mice *in vivo* (Figure 2.3E). The comparison between the B6-GF mice receiving B6 *vs.* CAST microbiota allowed us to estimate the contribution of the microbiota to the strain difference in insulin secretion (Figure 2.6A). Accordingly, the reduction in insulin secretion caused by CAST microbiota colonization in B6 mice was ~33%.

Circulating acetate is capable of modulating insulin secretion from pancreatic islets. Specifically, recent studies have shown that acetate directly enhances glucose-stimulated insulin secretion through activation of free fatty acid receptors on β-cells (Priyadarshini et al. 2015) and the parasympathetic nervous system (Perry et al. 2016). Therefore, we measured concentrations of SCFA in plasma and cecum, but found no differences in levels of acetate between $B6_{B6}$ and $B6_{CAST}$ mice (Figure S2.6A-B), suggesting that the divergent effects of the B6 and CAST microbiota on insulin secretion are unlikely to stem from differences in acetate.

Recent *in vitro* studies have also identified BA as important regulators of islet function (Düfer et al. 2012; Renga et al. 2010). We investigated the plasma BA profiles in the $B6_{B6}$ and $B6_{CAST}$ mice used for insulin secretion studies (Figure S2.6C-D). $B6_{CAST}$ BA profiles were composed of a significantly higher percentage of hydrophilic BA (Figure S2.6C). Consistent with a previous report (Sayin et al. 2013), BA profiles were dominated by taurine-conjugated species, with TωMCA and TβMCA being the two most abundant in both groups of animals (Figure S2.6D). In $B6_{B6}$ mice, the hydrophobic secondary BA DCA and LCA were significantly higher than in $B6_{CAST}$ mice (Figure S2.6D).

BA regulate insulin secretion through the activation of specific receptors in islets. For instance, BA can directly increase insulin secretion and production through activation of farnesoid X receptor (*Fxr*) in β-cells (Düfer et al. 2012; Renga et al. 2010). Expression of *Fxr* is increased

in an agonist-dependent manner (Lee et al. 2006). Remarkably, we found that expression of *Fxr* was significantly higher in $B6_{B6}$ islets compared with $B6_{CAST}$ islets (Figure 2.6B). These results suggest that the gut microbiota modulate BA-dependent signaling in pancreatic islets.

DISCUSSION

The collective genetic variance of the eight CC strains is roughly equivalent to that of the entire human population, with the three wild-derived strains (WSB, CAST and PWK) accounting for ~75% of the genetic diversity within the cohort (Roberts et al. 2007). Remarkably, these three wild-derived strains captured the full scope of dietary responsiveness observed across the panel (Figure 2.1 and S2.1). HF/HS feeding had no effect on any of the phenotypes measured in CAST mice, whereas it resulted in weight gain, glucose intolerance and insulin resistance in B6 mice. Additionally, the diet caused a simultaneous increase in weight and glucose in NZO mice. We also identified significant differences in the gut microbiota composition among strains and between diets. All animals were obtained from the same facility, and subject to the same environmental conditions throughout the study, and genetic differences among the mice is the only known variable. Together, these results support a role for host genetics to regulate the composition of the microbiota. However, it's important to note that although large population studies have identified highly heritable taxa, the genetic architecture underlying these taxa is highly complex with relatively small effect sizes that are difficult to replicate (Benson 2016).

From the CC founder panel, we identified B6 and CAST as the two strains with the most divergent phenotypes. Previous studies have exploited the differential response to diet-induced metabolic disease between B6 and CAST to identify genetic loci associated with metabolic disease (Mehrabian et al. 2000; Mehrabian et al. 1998). In these studies, the gut microbiome was not may have contributed to the metabolic differences between strains.

In order to dissect the contribution of the microbiome of B6 and CAST to their contrasting metabolic profiles, we resorted to fecal transplantation experiments. B6-GF mice colonized with the CAST microbiota were less affected by chronic HF/HS feeding relative to B6-GF mice

colonized with the B6 microbiota. The mice receiving the CAST microbiota secreted far less insulin in response to a glucose challenge, but were still able to maintain normal blood glucose levels.

We consistently identified microbial taxa in both the CC founders and transplant recipients associated with metabolic traits. Clostridiaceae showed the strongest positive correlation with plasma insulin levels and weight gain (Figure 2.2A). Clostridiaceae also had a strong positive correlation with AUC insulin, a proxy for pancreas function. OTUs within the Clostridiaceae family have previously been both positively and negatively associated with metabolic traits (Ussar et al. 2015; Karlsson et al. 2013), and a recent study showed a positive correlation between an increase in BMI and an increase of SCFA-producing Closdiria species in Danish infants (Bergström et al. 2014). In contrast to the elevated Clostridiaceae in mice with a B6 microbiota, Bacteroidaceae was significantly higher in CAST-CR and $B6_{CAST}$ mice (Figure S2.3C and 2.4C). Bacteroidaceae was negatively correlated with body weight, circulating insulin and $AUC_{insulin}$ (Figure 2.2A). A previous report found that daily oral administration of *Bacteroides uniformis*, a member of the Bacteroidaceae family, ameliorated metabolic dysfunction resulting from a high-fat diet (Gauffin Cano et al. 2012). This species also evoked a reduction in hepatic triglyceride levels, consistent with our observations that $B6_{CAST}$ mice have lower hepatic lipid levels compared to $B6_{B6}$ mice. Fecal abundance of members of the Bacteroidaceae family, including *Bacteroides vulgatus*, has also been reported to be lower in humans with T2D (X. Wu et al. 2010). Despite the high abundance of Bacteroidaceae in $B6_{CAST}$ mice, we did not observe complete protection from diet-induced metabolic disease that we observed in CAST-CR mice, suggesting that host factors, or taxa that failed to colonize transplanted mice (e.g., Verrucomicrobiaceae), contribute to the metabotype differences.

Vitamin B12 is exclusively produced by microbes (Martens et al. 2002) and several members of the Bacteroidaceae family transport, metabolize and produce vitamin B12 analogs (Goodman et al. 2009; Degnan et al. 2014; M. Wu et al. 2015). Metagenomic analysis of the microbial communities from mice with the CAST microbiota revealed microbial functional enrichment for pathways involved in the biosynthesis of vitamin B12 (Figure S2.4A), which is necessary for DNA synthesis, neurological function, hematopoiesis, epigenetic modifications, and propionate metabolism (Kibirige & Mwebaze 2013). Importantly, deficiencies in vitamin B12 are commonly observed in individuals with T2D and gestational diabetes (Kibirige & Mwebaze 2013; Krishnaveni et al. 2009), and B12 therapy improves insulin resistance and endothelial function in patients with metabolic syndrome by mechanisms that are not fully elucidated (Setola et al. 2004).

Our metagenomic analysis also revealed that genes involved in LPS production are enriched in the CAST-transplanted microbiome (Figure 2.4E and SF2.4B). This finding was surprising given that increased levels of LPS have been causally linked to the development of metabolic disease, yet $B6_{CAST}$ mice are partially protected from the effects of HF/HS feeding relative to $B6_{B6}$ animals (Figure 2.3). Taxonomic evaluation of the metagenomic data indicated that the Bacteroidetes phylum is the major contributor to the increased abundance of genes from this pathway (Table S4). This is relevant because unrelated bacteria generate structurally distinct LPS molecules with varying capacity to elicit an innate immune response (Whitfield & Trent 2014). Notably, a recent study showed that LPS derived from *E. coli* generates a strong inflammatory signal, whereas LPS derived from members of the Bacteroidetes phylum inhibited the host immune response (Vatanen et al. 2016). The differential ability of LPS sub-types to modify host physiology may explain why LPS has been shown to both stimulate (Nguyen et al. 2014) and attenuate (Amyot et al. 2012) insulin secretion. Studies aimed at testing the roles of LPS

derived from phylogenetically diverse taxa on metabolic disease and insulin secretion are warranted to further clarify how structural differences in this molecule affect host metabolism.

In addition to LPS, gut microbes produce SCFA, which are important energy and signaling molecules implicated in metabolic disease. For instance, butyrate has been shown to improve whole-body insulin sensitivity (Gao et al. 2009) and patients with T2D have reduced levels of butyrate-producing bacteria (Qin et al. 2010). SCFA are also elevated in individuals with diet-induced obesity, which is consistent with the elevated cecal SCFA levels in $B6_{B6}$ mice (Turnbaugh et al. 2008). Interestingly, SCFA are also known regulators of insulin sensitivity and secretion. Acetate can modulate insulin secretion from β-cells either directly through FFAR2 or via parasympathetic activation (Priyadarshini et al. 2015; Perry et al. 2016). However, we did not observe differences in concentrations of plasma or cecal acetate in the transplanted animals (Figure 2.4F and S2.6A-B). Therefore, it is unlikely that the differences in insulin secretion could be attributed to SCFA and consequentially implies there are multiple pathways through which the gut microbiota can module insulin secretion from β-cells.

Gut microbes are responsible for the production of the highly hydrophobic secondary BA DCA and LCA through the dehydroxylation of the primary BA, CA and CDCA, in the colon. Removal of glycine/taurine BA conjugates via BSH enzymes is a prerequisite for 7α/β-dehydroxylation of primary BA into secondary BA (Batta et al. 1990). Interestingly, there were 13 predicted BSH genes that were more abundant in the B6 metagenome but only two in the CAST metagenome. One possible interpretation of this result is that there may be more bacterial species present in the B6 microbiome that are able to deconjugate BA. Consistent with this, $B6_{B6}$ mice had significantly higher levels of secondary BA as well as hydrophobic BA species than $B6_{CAST}$ mice (Figure 2.5B-C and S2.6C-D), both of which are elevated in humans and mice with insulin

resistance (Ryan et al. 2014; Prawitt et al. 2011). Furthermore, DCA has been positively associated with higher levels of Firmicutes (Islam et al. 2011). This is consistent with our findings as B6-CR founders and $B6_{B6}$ had a significantly greater relative abundance of Firmicutes and fecal DCA than CAST-CR and $B6_{CAST}$ (Figure S2.3B and 2.5B). Conversely, $B6_{CAST}$ had a higher abundance of hydrophilic BA and the majority of the BA pool was comprised of the mouse primary BA, βMCA (Figure 2.5B-C).

The BA receptor *Fxr* is expressed in pancreatic β-cells and its activation via BA enhances insulin secretion (Kumar et al. 2012; Renga et al. 2010). Hydrophobic BA such as CDCA, DCA, LCA, and their taurine conjugates are known ligands of *Fxr*. The hydrophobic TCDCA increases insulin production and secretion through an FXR-dependent regulation of K_{ATP} channels (Düfer et al. 2012). Moreover, β-cell FXR activation in diabetic leptin receptor deficient (db/db) mice and NOD mice increases insulin secretion and delays the development of diabetes (Renga et al. 2010; Zhang et al. 2006). We detected higher levels of LCA and DCA in the feces and plasma of $B6_{B6}$ mice relative to $B6_{CAST}$ mice (Figure 2.5B and S2.6C-D), along with increased expression of *Fxr* in pancreatic islets from $B6_{B6}$ mice (Figure 2.6B). Altogether, this suggests that the gut microbiota and BA composition could modulate pancreatic function and insulin secretion.

We have highlighted four examples of microbial-derived products, vitamin B12, SCFAs, LPS, and BA, as plausible mediators of the microbiome effect on insulin secretion. However, there are thousands of other metabolites that were not characterized in our study and could also play an important role in regulating host metabolism. Future experiments using gnotobiotic mice colonized with defined communities that have different metabolic capabilities will provide mechanistic insights into the communication between gut microbes and the host.

EXPERIMENTAL PROCEDURES

Mouse husbandry. Animal care and study protocols were approved by the University of Wisconsin - Madison Animal Care and Use Committee.

Collaborative Cross (CC) mouse husbandry. Mice were housed on a 12 h-light:dark cycle. CC founder strains were obtained from The Jackson Laboratory (Bar Harbor, ME, USA) and were bred at University of Wisconsin, Madison. Mice were group housed by strain (2 mice/cage) and diet under a temperature- and humidity-controlled conditions, and received *ad libitum* access to water and food. After 4 weeks of age, mice were maintained on either a control (TD.08810, Envigo Teklad, 16.8%-kcal fat, 60.9% carbohydrate, 22.3% protein) or a high-fat high-sucrose diet (TD.08811, Envigo Teklad, 44.6%-kcal fat, 40.6% carbohydrate, 14.8% protein) (Table S1). Strains were housed within the same vivarium throughout the duration of the study.

Gnotobiotic mouse husbandry. C57BL/6J germ-free mice were bred and housed in the Microbial Sciences Building vivarium at University of Wisconsin-Madison to generate mice used in this study. B6-CR and B6-GF mice were housed in separate plastic flexible vinyl gnotobiotic isolators under temperature- and humidity-controlled conditions (12 hr light:dark). Fresh cecal contents were collected from 15-week old conventional B6-CR and CAST-CR mice maintained on the HF/HS diet (n = 2 to 3 mice per donor cecal microbiota samples). Cecal contents from B6 and CAST donor mice were resuspended in rich medium (1:100 w/vol) inside an anaerobic chamber. Suspensions were transferred into anaerobic sealed tubes and moved into gnotobiotic isolators. 9-week-old B6-GF male mice were inoculated via a single oral gavage with ~0.2 ml of cecal incoula (Turnbaugh et al. 2009). Each group of mice was housed in a controlled environment in separate plastic flexible vinyl gnotobiotic isolators under standard conditions. Recipient mice received sterilized water and HF/HS diet (TD.0.8811) *ad libitum* beginning one week before colonization.

Fasting plasma measurements. Following a 4h fast, blood was collected via retro-orbital bleed in EDTA-coated eppendorf tubes. Blood samples were centrifuged and plasma was collected and stored at -80°C until further analysis. Plasma glucose was quantified using the Thermo-Fisher Infinity Glucose Oxidase reagent (Pittsburgh, PA), insulin was quantified using the Millipore-Linco Sensitive Rat Insulin RIA (Billerica, MA), and triglycerides levels were quantified Thermo-Fisher Infinity Triglycerides reagents (Pittsburgh, PA).

Oral Glucose Tolerance Test (oGTT). Mice were fasted for four hours prior to testing and were challenged with an oral dose of 2 g/kg body weight glucose at time 0. Blood was collected via retro-orbital bleed at 0, 5, 15, 30, 60 and 120 minutes post glucose challenge. Blood samples were centrifuged and plasma collected and stored at -80°C until further analysis.

Triglyceride measurement. Liver triglycerides (TG) were quantified following the Bligh and Dyer extraction method (Bligh & Dyer 1959). Briefly, ~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 from Wako (Richmond, VA) according to the manufacturer's instructions and expressed in μg of triglycerides per milligram of protein.

Microbiome Sample Processing. Genomic DNA was extracted from feces and cecum 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 500 µl of extraction buffer [200 mM Tris (pH 8.0), 200 mM NaCL, 20 mM EDTA], 210 µl of 20% SDS, 500 µl phenol:chloroform:isoamyl alcohol (pH 7.9, 25:24:1) and 500 µl of 0.1-mm diameter zirconia/silica beads. Samples were mechanically disrupted using a bead beater (BioSpec Products, Barlesville, OK; maximum setting for 3 min at room temperature), followed by centrifugation, recovery of the aqueous phase, and precipitation with isopropanol. NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel, Bethlehem, PA) was used to remove contaminants. Isolated DNA was eluted in 5 mM Tris/HCl (pH 8.5) and was stored at -20°C until further use.

Collaborative cross founders: Amplicons of ~330 bp, spanning variable region 2 (V2) of the bacterial 16S rRNA gene, were generated by using modified primers 27F and 338R that incorporated sample specific barcodes (Muegge et al., 2011). A final library for sequencing was created by combining equimolar ratios of amplicons from the individual samples. The 16S rRNA amplicon mixture was subjected to 454 pyrosequencing conducted on a Roche GS Junior (Roche, Indianapolis, IN) with the Lib-L kit and Titanium chemistry.

Transplant: Amplification of 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 duplicate 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) (Caporaso, Kuczynski, et al. 2010) version 1.9.1. Quality filtered reads were trimmed of Illumina adaptor and barcode sequences. Sequences were then clustered in OTUs using an open-reference 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, Bittinger, et al. 2010). 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 (Lozupone & Knight 2005). 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. CC founder cecal samples sequenced by 454 pyrosequencing were rarefied to an even sampling depth of 900 reads, and 5 samples were removed from the dataset as assigned reads fell below the rarefaction point of 900 reads/sample. Donor and recipient samples from microbiota transplants were rarefied to an even sampling depth of 10,000 reads/sample. 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. OTUs representing less than 0.1% were removed for relative abundance assessments and

correlation analyses. Assessments of alpha-diversity and beta-diversity were also conducted on the rarefied OTU table in QIIME. Principal coordinate analysis (PCoA) was performed in QIIME using UniFrac distances calculated from the Greengenes reference tree. Permutation Multivariate Analysis of Variance (PERMANOVA) was used to compare strength of sample groups (diet, genotype) for founder PCoA using the compare_categories.py command in QIIME. Linear discriminant analysis (LDA) effect size (LEfSe) was used to identify taxa that discriminated between the fecal microbiota of transplant recipient mice using standard parameters ($p < 0.05$, LDA score 2.0)(Segata et al. 2011). For correlation analyses, only microbial families with at least one non-zero measurement for each strain on at least one diet were included. Correlations between microbiota and phenotypes and association testing were performed in R. Correlation coefficients and adjusted p-values are reported in Table S2.

Metagenomic analysis. Raw reads were pre-processed using the fastx toolkit (version 0.0.13) (Hannon Lab n.d.): raw reads were demultiplexed using fastx_barcode_splitter (specifying -bol -partial 2 and -mismatches 2), barcodes were trimmed using fastx_trimmer, (specifying –f 9 and –Q 33), and quality trimmed using fastq_quality_trimmer, (specifying -t 20 -l 30 and -Q 33). In order to filter out host contaminating reads in the metagenome samples, we identified paired and unpaired reads in our demultiplexed and trimmed files, and mapped them independently to the mouse genome assembly (Ensembl release 84, GRCm38.dna.toplevel) using Bowtie2 (v. 2.2.7) (Langmead & Salzberg 2012) with default settings (Table S3). From this output, we then identified reads that did not map to the mouse genome using samtools view (version 1.3) (Li et al. 2009), specifying –f 4 only for unpaired reads, or both –f4 and –f 8 for paired reads in addition to default

settings, and regenerated .fastq files containing only reads that did not map to the mouse genome (custom perl scripts).

In order to examine gene-level abundance differences among our samples, we utilized the mouse gut metagenome gene sequences available from the Mouse Gut Metagenome Project (downloaded from gigadb.org: http://gigadb.org/dataset/100114) (Xiao et al. 2015). "Decontaminated" paired and unpaired reads were independently mapped against genes in the mouse gut metagenome assembly with Bowtie2 using default settings (v. 2.2.7) (Langmead & Salzberg 2012). A table of raw read counts was generated using htseq-count command (v. 0.6.0) (Anders et al. 2014), specifying a 'mock'.gff file containing "gene" entries, whose lengths were lengths of genes, for example:

```
S-Fe10_GL0000040  mock  gene  1  1870  .  +  .  gene_id "S-Fe10_GL0000040";
```

The resulting raw read count table was filtered to exclude low abundance genes, defined here as genes with average raw read counts of less than 10 across all 12 samples (total number of 73,905 genes), and then input into DESeq2 (version 1.10.1) (Love et al. 2014), for library size normalization (default settings). To allow for comparison of individual gene abundances, counts were further normalized by gene length to give "reads per kilobase gene" (Table S4).

In order to examine the similarity of the B6-derived microbiota DESeq2 was also used to identify genes differentially abundant between the B6 and CAST-derived microbiota using default settings, and found a very large number of genes differentially abundant (29,283 in B6 > CAST microbiota; 10,742 in CAST > B6 microbiota, with FDR < 0.05, Table S4), indicating dramatic genic diversity between B6 and CAST-derived microbiota.

Functional (KEGG Orthology (KO)and eggNOG annotations) and taxonomic annotations corresponding to the detected genes from the aforementioned mouse gut metagenome were downloaded (gigadb.org: http://gigadb.org/dataset/100114)(Xiao et al. 2015). This was further expanded to include enzyme commission numbers (ECs) and KEGG pathway information, by mapping these functions from KEGG to the individual genes by way of KO annotations. Enrichment of these functional groups in genes of increased abundance in B6 or CAST-derived microbiota compared to background (all genes detected) was examined using a Fisher's exact test (p-value < 0.05).

Bile acid analysis. ~100 mg feces were homogenized in 500 µl 50:50 water:methanol. Next 500 µl of alkaline acetonitrile (5% ammonium hydroxide in acetonitrile) was added to the homogenate, which was then heated for 20 minutes at 75°C. 500 µl of the mixture was centrifuged at 11,000 RPM for 10 minutes and 250 µl of the supernatant was collected and evaporated under N2 gas. Samples were reconstituted in 50 µl 50:50 water:methanol and 2H4-CDCA was added to the samples for a final concentration of 2 µg/ml. For serum samples, 1 ml ice-cold acetonitrile was added to 50 µl serum and spiked with the internal standard for a final concentration of 2 µg/ml 2H4-CDCA. The mixture was vortexed, centrifuged at 15,000 x g for 10 minutes, and the supernatant was aspirated and evaporated under vacuum. The LC-MS/MS conditions used were as described (Youcai Zhang & Klaassen 2010).

Measurement of SCFAs. Flash-frozen cecal contents (100mg) or plasma (50 µl) were mixed with 20 µl 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 µl 33% HCl.

Two rounds of extraction using 1 ml diethyl ether were carried out by mixing for 10 minutes at room temperature following by centrifugation at 1932 x g for 10 minutes at 4°C. Extracts (60 μl) were then incubated at room temperature for 2 hours with 2 μl N-tert-Butyldimethylsilyl-N-methyltrifluoroacetamide (MTBSTFA, Sigma-Aldrich #394882). Derivatized samples (1 μl) were injected onto an Agilent 7890B/5977A GC/MSD instrument with a DB1-ms column. A linear temperature gradient was used, wherein the initial temperature of 80°C was held for 1 minute, then increased to 280°C at a rate of 15°C per minute prior to a final hold at 280°C for 5 minutes. The source temperature was set to 200°C and emission current to 300mA. The injector and transfer line temperatures were set to 250°C. Quantitation was performed using selected ion monitoring 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). Concentrations were normalized to g of cecal contents or ml plasma.

Islet isolation, ex vivo insulin secretion and RNA isolation. Intact pancreatic islets were isolated from mice using a collagenase digestion procedure (Rabaglia et al. 2005). Briefly, islets were carefully hand-picked under a stereo microscope to remove contaminating acinar tissue. For insulin secretion assays, single islets were placed in a well of a 96-well microtiter plate and used to determine the amount of insulin secreted in response to low (1.7 mM) or high (16.7 mM) glucose, KCl (40 mM, plus 1.7 mM glucose), or the incretin hormone GLP-1 (100 nM, plus either 8.3 or 16.7 mM glucose). From each mouse, 7 islets were used per secretory condition, and 5 mice were surveyed per strain (B6, A/J, WSB, CAST), or transplant group (B6B6, B6CAST). Insulin

secretion was monitored over a 45 min period. Insulin levels in the medium as well as that remaining within the islets was determined by ELISA.

Islets used for RNA isolation were washed twice with phosphate buffered saline (PBS) and centrifuged at 1500 rpm, 3 min RT. The PBS supernatant was removed and 350 µl RLT buffer (Qiagen, Hilden, Germany) was added. Islets were homogenized by hand for 1 min with a plastic micropestel (USA Scientific) and stored at -80oC until RNA purification. Total RNA was purified using the RNeasy Mini Kit (Qiagen, Hilden, Germany) following manufacturer's directions with on-column TURBO DNase treatment (Invitrogen, Carlsbad, CA).

Quantitative Real-Time PCR. SuperScript II Reverse Transcriptase with oligo(dT) primer (all from Invitrogen, Carlsbad, CA) was used to synthesize 20 µl cDNA templates from 100 ng purified RNA. cDNA was diluted 2X before use and qRT-PCR reactions were prepared in a 10µl volume using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, USA) and 400 nM specific primers targeting the gene of interest (FXR-F [5'-CCAACCTGGGTTTCTACCC-3']; FXR-R [5'-CACACAGCTCATCCCCTTT-3']). Reactions were run on a CFX96 Real-Time PCR System (Bio-Rad, Hercules, CA, USA). Relative gene expression was calculated by the ΔΔCt method using β-actin as an internal control.

Statistical Analysis. The data are expressed as mean ± SEM and analyzed using GraphPad Prism 6.0 (GraphPad Software, La Jolla, CA). Multiple groups were analyzed by one-way or two-way ANOVA followed by Bonferroni's multiple comparisons test. Significant differences between two groups were evaluated by two-tailed unpaired Student's t-test or Mann-Whitney U test for samples that were not normally distributed. Pearson's correlations between microbiota and phenotypes and

association testing were performed in R. The level of significance was set at $p < 0.05$; $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

ACCESSION NUMBERS

The data reported in this paper are accessible in the European Nucleotide Archive (ENA) database under accession ID PRJEB15120.

CONTRIBUTIONS

This project was conceived by Federico Rey, Alan Attie, Mark Keller and Julia Kemis (Kreznar). Kathryn Schueler, Donnie Stapleton, and Mary Rabaglia measured clinical phenotypes, collected fecal samples, and harvested tissues from conventionally-raised mice. Kathryn Schueler performed oral glucose tolerance tests (oGTT) on all conventionally-raised and gnotobiotic mice used in study. Mary Rabaglia isolated pancreatic islets and performed insulin secretion assays in gnotobiotic experiments. Eugenio Vivas bred gnotobiotic mice used for transplant experiments and maintained animals prior to colonization. Following colonization, Julia Kemis performed gnotobiotic husbandry and daily care. Measurement of clinical phenotypes (except oGTT), fecal sample collection and tissue harvest of gnotobiotic mice was performed by Julia Kemis. Assessment of microbiota composition in conventionally-raised and gnotobiotic mice performed by Julia Kemis. Microbiome function in gnotobiotic animals conducted by Lindsay Traeger. Bile acid measurements done by Wen Zhao and Bruno Hagenbuch at the University of Kansas. Data interpretation and statistical analysis performed by Julia Kemis, Brian Yandell and Aimee Teo Broman. The manuscript was written by Julia Kemis, Mark Keller, Alan Attie and Federico Rey.

ACKNOWLEDGEMENTS

The authors thank the University of Wisconsin Biotechnology Center DNA Sequencing Facility for providing sequencing and support services, and the University of Wisconsin Center for High Throughput Computing (CHTC) in the Department of Computer Sciences for providing computational resources, support, and assistance. We also thank the University of Wisconsin Mass Spectrometry Facility for technical assistance. Finally, we are grateful to Dr. Jeffrey Gordon for his support on this project.

This work was supported in part by grants NIH DK1018259-01 (F.E.R.), DK101573 (A.D.A), RR021940 (B.H.) and GM077336 (B.H.) and by the Clinical and Translational Science Award (CTSA) program, through the NIH National Center for Advancing Translational Sciences (NCATS) grants UL1TR000427 (A.T.B) and KL2TR000428 (F.E.R.). L.L.T. is supported by a NLM Computation and Informatics in Biology and Medicine Postdoctoral Fellowship (5T15LM007359). J.H.K. is supported by the NIH National Institute of Allergy and Infectious Diseases grant T32AI55397.

REFERENCES

Amyot, J. et al., 2012. Lipopolysaccharides impair insulin gene expression in isolated islets of Langerhans via Toll-Like Receptor-4 and NF-κB signalling. O. S. Shirihai, ed. *PloS one*, 7(4), p.e36200.

Anders, S., Pyl, P.T. & Huber, W., 2014. HTSeq - A Python framework to work with high-throughput sequencing data. *bioRxiv*, p.002824.

Aylor, D.L. et al., 2011. Genetic analysis of complex traits in the emerging Collaborative Cross. *Genome research*, 21(8), pp.1213–1222.

Batta, A.K. et al., 1990. Side chain conjugation prevents bacterial 7-dehydroxylation of bile acids. *The Journal of biological chemistry*, 265(19), pp.10925–10928.

Benson, A.K., 2016. The gut microbiome-an emerging complex trait. *Nature genetics*, 48(11), pp.1301–1302.

Bergström, A. et al., 2014. Establishment of intestinal microbiota during early life: a longitudinal, explorative study of a large cohort of Danish infants. *Applied and environmental microbiology*, 80(9), pp.2889–2900.

Besten, den, G. et al., 2013. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *Journal of lipid research*, 54(9), pp.2325–2340.

Bligh, E.G. & Dyer, W.J., 1959. A rapid method of total lipid extraction and purification. *Canadian journal of biochemistry and physiology*, 37(8), pp.911–917.

Caporaso, J.G., Bittinger, K., et al., 2010. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*, 26(2), pp.266–267.

Caporaso, J.G., Kuczynski, J., et al., 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7(5), pp.335–336.

Carmody, R.N. et al., 2015. Diet dominates host genotype in shaping the murine gut microbiota. *Cell host & microbe*, 17(1), pp.72–84.

Churchill, G.A. et al., 2004. The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nature genetics*, 36(11), pp.1133–1137.

Clemente, J.C. et al., 2012. The impact of the gut microbiota on human health: an integrative view. *Cell*, 148(6), pp.1258–1270.

Degnan, P.H., Taga, M.E. & Goodman, A.L., 2014. Vitamin B12 as a modulator of gut microbial ecology. *Cell metabolism*, 20(5), pp.769–778.

DeSantis, T.Z. et al., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology*, 72(7), pp.5069–5072.

Deutscher, J., Francke, C. & Postma, P.W., 2006. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiology and molecular biology reviews : MMBR*, 70(4), pp.939–1031.

Düfer, M. et al., 2012. Bile acids acutely stimulate insulin secretion of mouse β-cells via farnesoid X receptor activation and K(ATP) channel inhibition. *Diabetes*, 61(6), pp.1479–1489.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), pp.2460–2461.

Gao, Z. et al., 2009. Butyrate improves insulin sensitivity and increases energy expenditure in mice. *Diabetes*, 58(7), pp.1509–1517.

Gauffin Cano, P. et al., 2012. Bacteroides uniformis CECT 7771 ameliorates metabolic and immunological dysfunction in mice with high-fat-diet induced obesity. S. Bereswill, ed. *PloS one*, 7(7), p.e41079.

Goodman, A.L. et al., 2009. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell host & microbe*, 6(3), pp.279–289.

Handelsman, Y., 2011. Role of bile acid sequestrants in the treatment of type 2 diabetes. *Diabetes care*, 34 Suppl 2(Supplement_2), pp.S244–50.

Hannon Lab, FASTX Toolkit. Available at: http://hannonlab.cshl.edu/fastx_toolkit/.

Hartstra, A.V. et al., 2015. Insights into the role of the microbiome in obesity and type 2 diabetes. *Diabetes care*, 38(1), pp.159–165.

Hildebrandt, M.A. et al., 2009. High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology*, 137(5), pp.1716–24.e1–2.

Houten, S.M., Watanabe, M. & Auwerx, J., 2006. Endocrine functions of bile acids. *The EMBO journal*, 25(7), pp.1419–1425.

Islam, K.B.M.S. et al., 2011. Bile acid is a host factor that regulates the composition of the cecal microbiota in rats. *Gastroenterology*, 141(5), pp.1773–1781.

Karlsson, F.H. et al., 2013. Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature*, 498(7452), pp.99–103.

Kars, M. et al., 2010. Tauroursodeoxycholic Acid may improve liver and muscle but not adipose tissue insulin sensitivity in obese men and women. *Diabetes*, 59(8), pp.1899–1905.

Kawamata, Y. et al., 2003. A G protein-coupled receptor responsive to bile acids. *The Journal of biological chemistry*, 278(11), pp.9435–9440.

Kibirige, D. & Mwebaze, R., 2013. Vitamin B12 deficiency among patients with diabetes mellitus: is routine screening and supplementation justified? *Journal of diabetes and metabolic disorders*, 12(1), p.17.

Koh, A. et al., 2016. From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. *Cell*, 165(6), pp.1332–1345.

Kozich, J.J. et al., 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(17), pp.5112–5120.

Krishnaveni, G.V. et al., 2009. Low plasma vitamin B12 in pregnancy is associated with gestational "diabesity" and later diabetes. *Diabetologia*, 52(11), pp.2350–2358.

Kuipers, F., Bloks, V.W. & Groen, A.K., 2014. Beyond intestinal soap--bile acids in metabolic control. *Nature reviews. Endocrinology*, 10(8), pp.488–498.

Kumar, D.P. et al., 2012. Activation of transmembrane bile acid receptor TGR5 stimulates insulin secretion in pancreatic β cells. *Biochemical and biophysical research communications*, 427(3), pp.600–605.

Langmead, B. & Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), pp.357–359.

Lee, F.Y. et al., 2006. FXR, a multipurpose nuclear receptor. *Trends in biochemical sciences*, 31(10), pp.572–580.

Ley, R.E. et al., 2005. Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America*, 102(31), pp.11070–11075.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), pp.2078–2079.

Love, M.I., Huber, W. & Anders, S., 2014. *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2*, Cold Spring Harbor Labs Journals.

Lozupone, C. & Knight, R., 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and environmental microbiology*, 71(12), pp.8228–8235.

Makishima, M. et al., 1999. Identification of a nuclear receptor for bile acids. *Science (New York, N.Y.)*, 284(5418), pp.1362–1365.

Martens, J.H. et al., 2002. Microbial production of vitamin B12. *Applied microbiology and biotechnology*, 58(3), pp.275–285.

McDonald, D. et al., 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal*, 6(3), pp.610–618.

Mehrabian, M. et al., 2000. Genetic control of HDL levels and composition in an interspecific mouse cross (CAST/Ei x C57BL/6J). *Journal of lipid research*, 41(12), pp.1936–1946.

Mehrabian, M. et al., 1998. Genetic loci controlling body fat, lipoprotein metabolism, and insulin levels in a multifactorial mouse model. *The Journal of clinical investigation*, 101(11), pp.2485–2496.

Muegge, B.D. et al., 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science (New York, N.Y.)*, 332(6032), pp.970–974.

Nguyen, A.T. et al., 2014. Lipopolysaccharides-mediated increase in glucose-stimulated insulin secretion: involvement of the GLP-1 pathway. *Diabetes*, 63(2), pp.471–482.

O'Connor, A. et al., 2014. Responsiveness of cardiometabolic-related microbiota to diet is influenced by host genetics. *Mammalian genome : official journal of the International Mammalian Genome Society*, 25(11-12), pp.583–599.

Ozcan, U. et al., 2006. Chemical chaperones reduce ER stress and restore glucose homeostasis in a mouse model of type 2 diabetes. *Science (New York, N.Y.)*, 313(5790), pp.1137–1140.

Parks, B.W. et al., 2013. Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell metabolism*, 17(1), pp.141–152.

Parks, D.J. et al., 1999. Bile acids: natural ligands for an orphan nuclear receptor. *Science (New York, N.Y.)*, 284(5418), pp.1365–1368.

Perry, R.J. et al., 2016. Acetate mediates a microbiome-brain-β-cell axis to promote metabolic syndrome. *Nature*, 534(7606), pp.213–217.

Prawitt, J., Caron, S. & Staels, B., 2011. Bile acid metabolism and the pathogenesis of type 2 diabetes. *Current diabetes reports*, 11(3), pp.160–166.

Priyadarshini, M. et al., 2015. An Acetate-Specific GPCR, FFAR2, Regulates Insulin Secretion. *Molecular endocrinology (Baltimore, Md.)*, 29(7), pp.1055–1066.

Qin, J. et al., 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285), pp.59–65.

Qin, J. et al., 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*, 490(7418), pp.55–60.

Rabaglia, M.E. et al., 2005. Alpha-Ketoisocaproate-induced hypersecretion of insulin by islets from diabetes-susceptible mice. *American journal of physiology. Endocrinology and metabolism*, 289(2), pp.E218–24.

Renga, B. et al., 2010. The bile acid sensor FXR regulates insulin transcription and secretion. *Biochimica et biophysica acta*, 1802(3), pp.363–372.

Ridaura, V.K. et al., 2013. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science (New York, N.Y.)*, 341(6150), pp.1241214–1241214.

Ridlon, J.M., Kang, D.-J. & Hylemon, P.B., 2006. Bile salt biotransformations by human intestinal bacteria. *Journal of lipid research*, 47(2), pp.241–259.

Roberts, A. et al., 2007. The polymorphism architecture of mouse genetic resources elucidated using genome-wide resequencing data: implications for QTL discovery and systems genetics. *Mammalian genome : official journal of the International Mammalian Genome Society*, 18(6-7), pp.473–481.

Russell, D.W., 2009. Fifty years of advances in bile acid synthesis and metabolism. *Journal of lipid research*, 50 Suppl(Supplement), pp.S120–5.

Ryan, K.K. et al., 2014. FXR is a molecular target for the effects of vertical sleeve gastrectomy. *Nature*, 509(7499), pp.183–188.

Sayin, S.I. et al., 2013. Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. *Cell metabolism*, 17(2), pp.225–235.

Segata, N. et al., 2011. Metagenomic biomarker discovery and explanation. *Genome biology*, 12(6), p.R60.

Setola, E. et al., 2004. Insulin resistance and endothelial function are improved after folate and vitamin B12 therapy in patients with metabolic syndrome: relationship between homocysteine levels and hyperinsulinemia. *European Journal of Endocrinology*, 151(4), pp.483–489.

Sommer, F. & Bäckhed, F., 2013. The gut microbiota--masters of host development and physiology. *Nature reviews. Microbiology*, 11(4), pp.227–238.

Tremaroli, V. & Bäckhed, F., 2012. Functional interactions between the gut microbiota and host metabolism. *Nature*, 489(7415), pp.242–249.

Turnbaugh, P.J. et al., 2009. A core gut microbiome in obese and lean twins. *Nature*, 457(7228), pp.480–484.

Turnbaugh, P.J. et al., 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, 444(7122), pp.1027–1031.

Turnbaugh, P.J. et al., 2008. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell host & microbe*, 3(4), pp.213–223.

Ussar, S. et al., 2015. Interactions between Gut Microbiota, Host Genetics and Diet Modulate the Predisposition to Obesity and Metabolic Syndrome. *Cell metabolism*, 22(3), pp.516–530.

Vatanen, T. et al., 2016. Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. *Cell*, 165(4), pp.842–853.

Vrieze, A. et al., 2010. The environment within: how gut microbiota may influence metabolism and body composition. *Diabetologia*, 53(4), pp.606–613.

Whitfield, C. & Trent, M.S., 2014. Biosynthesis and export of bacterial lipopolysaccharides. *Annual review of biochemistry*, 83(1), pp.99–128.

Wu, M. et al., 2015. Genetic determinants of in vivo fitness and diet responsiveness in multiple human gut Bacteroides. *Science (New York, N.Y.)*, 350(6256), pp.aac5992–aac5992.

Wu, X. et al., 2010. Molecular characterisation of the faecal microbiota in patients with type II diabetes. *Current microbiology*, 61(1), pp.69–78.

Xiao, L. et al., 2015. A catalog of the mouse gut metagenome. *Nature Biotechnology*, 33(10), pp.1103–1108.

Zhang, Yanqiao et al., 2006. Activation of the nuclear receptor FXR improves hyperglycemia and hyperlipidemia in diabetic mice. *Proceedings of the National Academy of Sciences of the United States of America*, 103(4), pp.1006–1011.

Zhang, Youcai & Klaassen, C.D., 2010. Effects of feeding bile acids and a bile acid sequestrant on hepatic bile acid composition in mice. *Journal of lipid research*, 51(11), pp.3230–3242.

FIGURES

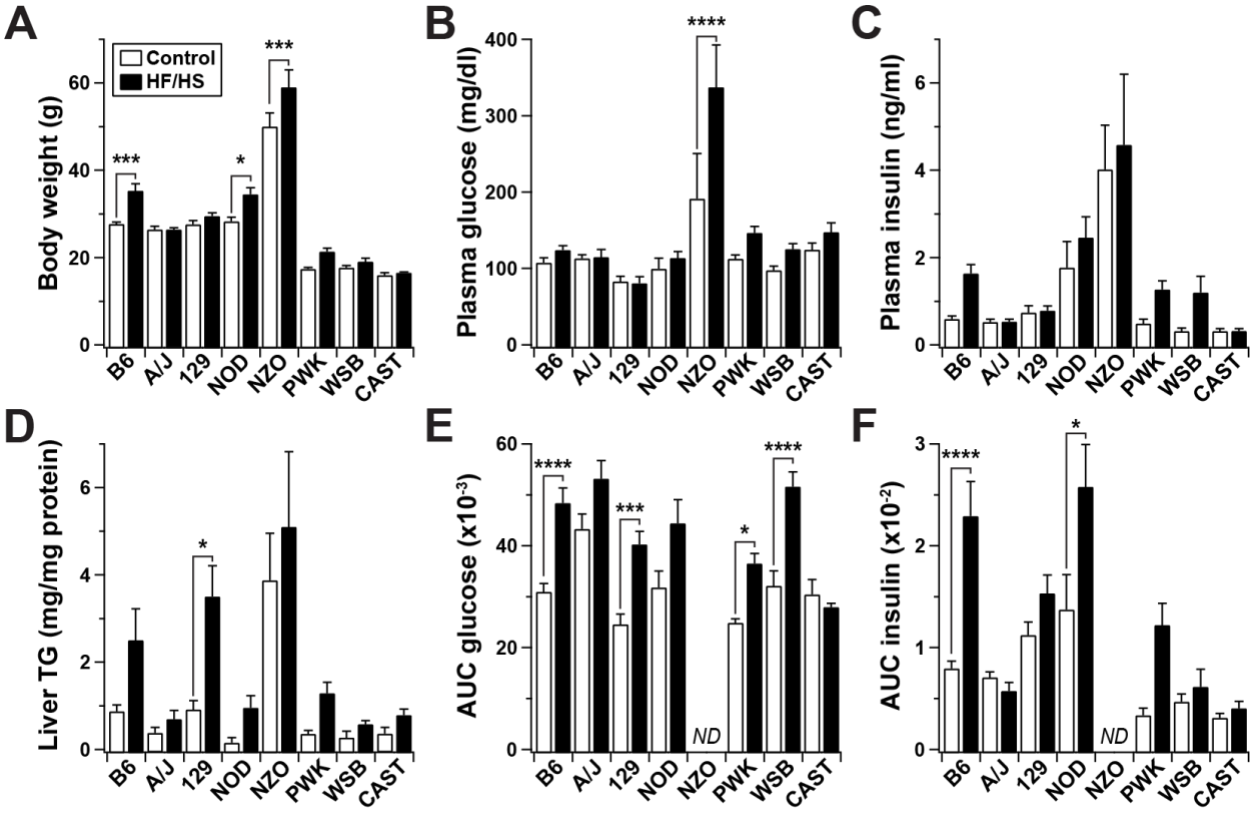


Figure 2.1. Segregation of metabolic syndrome among CC founder mice. Male mice were maintained on the high-fat/high-sucrose (HF/HS) or a control diet for 22 weeks beginning at 4 weeks of age. (A) Body weight, (B) fasting plasma glucose and (C) insulin, and (D) hepatic triglyceride content determined for all mice at 26 weeks of age. Areas under the curve (AUC) for (E) glucose and (F) insulin during oral glucose tolerance test (oGTT) conducted at 22 weeks of age. Insulin and glucose values were determined from plasma following a 4 hour fast. No data (ND) were collected for NZO mice during oGTT. In all panels, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$ by two-way ANOVA (diet and strain) with Bonferroni's multiple comparisons test to assess within-strain differences. Data are mean ± SEM, $n \geq 9$ mice/genotype/diet.

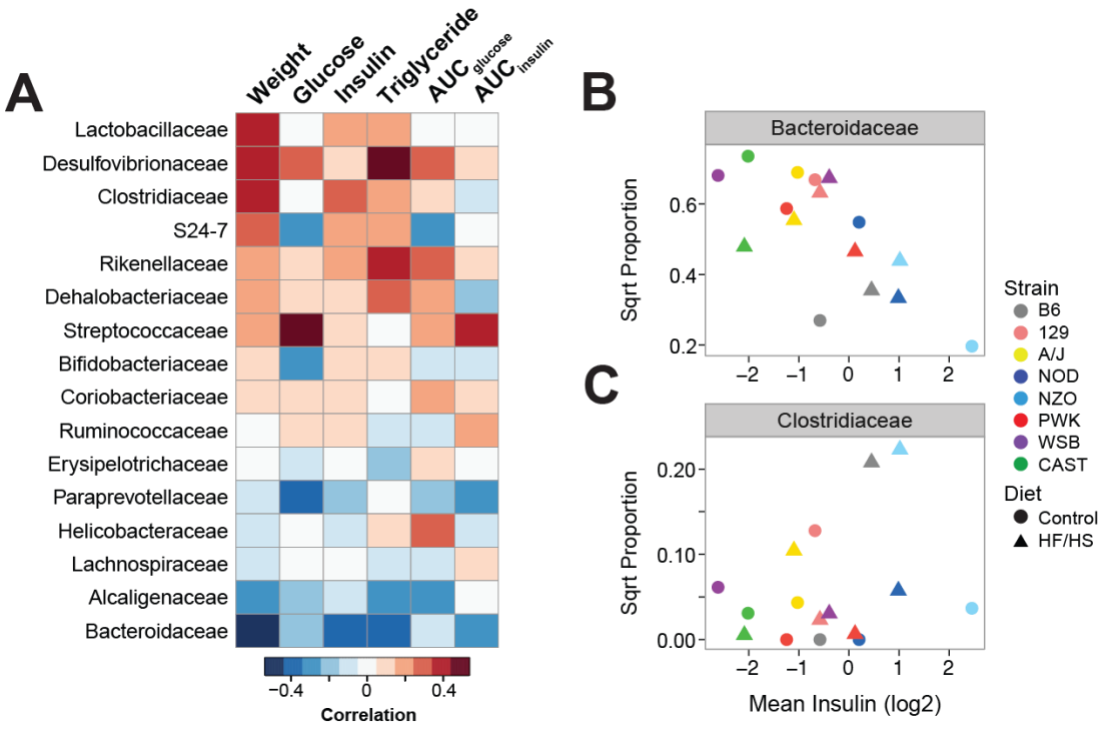


Figure 2.2. Gut microbial taxa correlate with metabolic phenotypes. (A) Heat map illustrates Pearson's pair-wise correlation between microbial families and diabetes-related clinical traits measured in the 8 CC founder mice ($n \geq 9$ mice/genotype/diet). Microbial families are ordered by their correlation to body weight. Red, positive correlation; blue, negative. Area under the curve (AUC) values for insulin and glucose were computed from oGTT conducted at 22 weeks; other metrics were collected at 26 weeks. Correlation coefficients and p-values found in Table S2. Contributions of strain and diet on the correlations observed between fasting insulin and (B) the Bacteroidaceae family, and (C) the Clostridiaceae family.

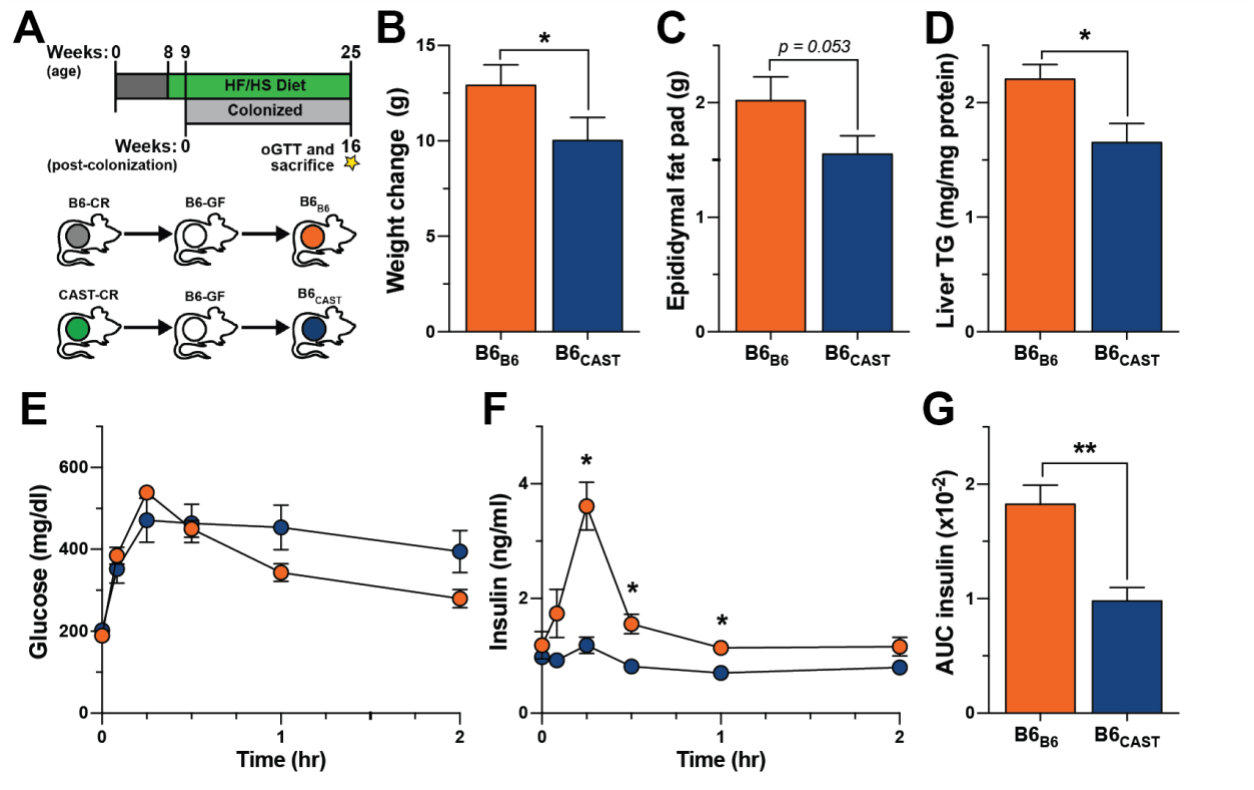
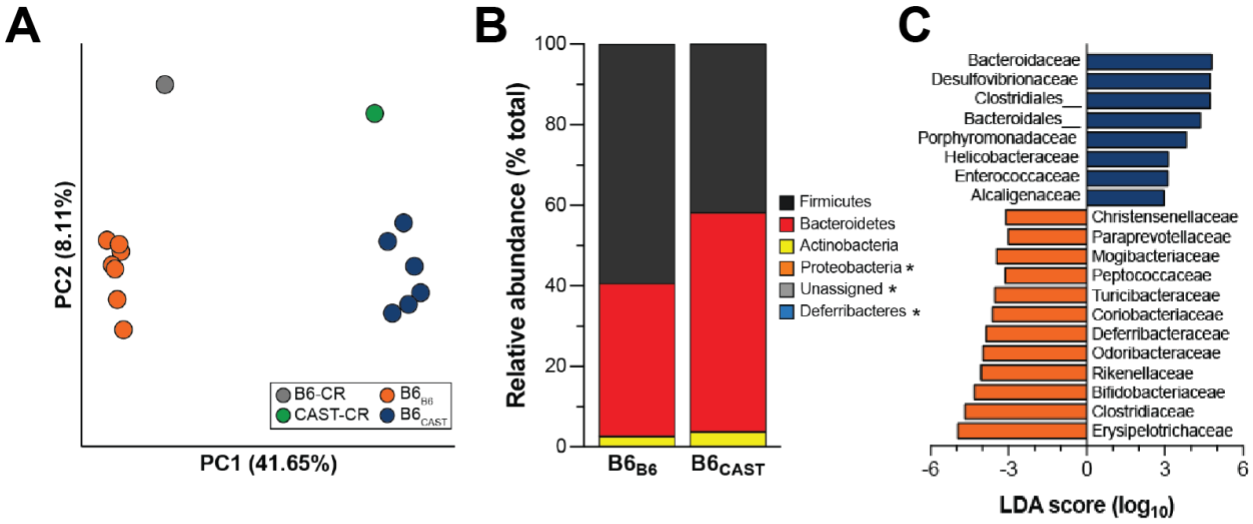
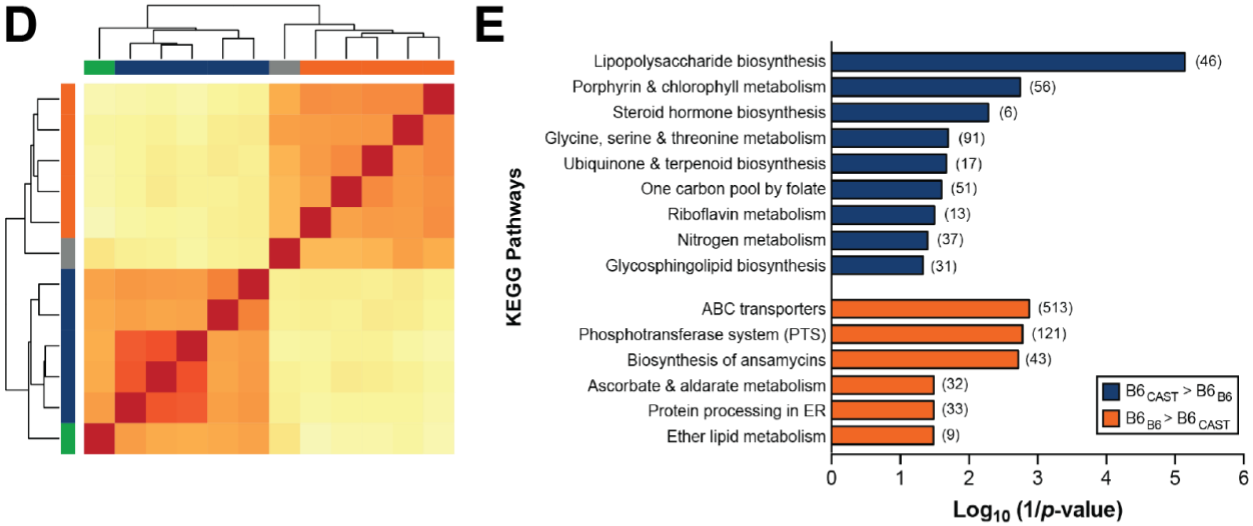


Figure 2.3. Divergent effects of B6 and CAST microbiomes on diet-induced metabolic phenotypes. (A) Transplant experimental design. (B) Total weight change, (C) epididymal fat pad mass and (D) quantification of hepatic triglyceride (TG) contents. (E and F) Glucose and insulin values during oGTT and (G) AUC insulin in $B6_{B6}$ and $B6_{CAST}$ mice. All measurements shown collected 16-weeks post-colonization. *$p < 0.05$, **$p < 0.01$ by Student's t-test. Data are mean ± SEM, n = 7 for $B6_{B6}$ and n = 6 for $B6_{CAST}$ mice.


A
PC2 (8.11%)
PC1 (41.65%)
B6-CR
CAST-CR
B6B6
B6CAST
B
Relative abundance (% total)
100
80
60
40
20
0
B6B6
B6CAST
Firmicutes
Bacteroidetes
Actinobacteria
Proteobacteria *
Unassigned *
Deferribacteres *
C
Bacteroidaceae
Desulfovibrionaceae
Clostridiales__
Bacteroidales__
Porphyromonadaceae
Helicobacteraceae
Enterococcaceae
Alcaligenaceae
Christensenellaceae
Paraprevotellaceae
Mogibacteriaceae
Peptococcaceae
Turicibacteraceae
Coriobacteriaceae
Deferribacteraceae
Odoribacteraceae
Rikenellaceae
Bifidobacteriaceae
Clostridiaceae
Erysipelotrichaceae
-6
-3
0
3
6
LDA score (log10)


D
E
KEGG Pathways
Lipopolysaccharide biosynthesis (46)
Porphyrin & chlorophyll metabolism (56)
Steroid hormone biosynthesis (6)
Glycine, serine & threonine metabolism (91)
Ubiquinone & terpenoid biosynthesis (17)
One carbon pool by folate (51)
Riboflavin metabolism (13)
Nitrogen metabolism (37)
Glycosphingolipid biosynthesis (31)
ABC transporters (513)
Phosphotransferase system (PTS) (121)
Biosynthesis of ansamycins (43)
Ascorbate & aldarate metabolism (32)
Protein processing in ER (33)
Ether lipid metabolism (9)
B6CAST > B6B6
B6B6 > B6CAST
0
1
2
3
4
5
6
Log10 (1/p-value)

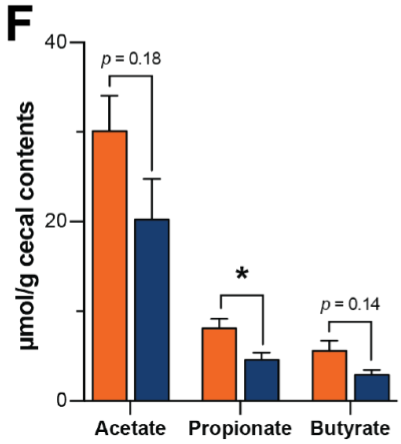

F
40
20
0
μmol/g cecal contents
p = 0.18
*
p = 0.14
Acetate
Propionate
Butyrate

Figure 2.4. Gut microbiota composition and function of transplant recipients. (A) Principal coordinate analysis (PCoA) of unweighted UniFrac distances for the fecal microbiota of transplant donors and recipients at sacrifice. Each circle represents an individual mouse. Percent variation explained by each PC is shown in parentheses. (B) Relative abundance of major microbial phyla ordered by increasing mean abundance; * denotes mean phyla abundance <1%. (C) Microbial families differentially enriched in either $B6_{CAST}$ (blue) or $B6_{B6}$ (orange) as determined by linear discriminant analysis (LDA) with effect size (LEfSe). (D) Clustering of mice based on relative abundance of KEGG metabolic pathways using euclidian distance measurement with complete linkage hierarchical clustering; B6-CR (grey), CAST-CR (green), $B6_{B6}$ (orange), $B6_{CAST}$ (blue). (E) KEGG categories enriched in either CAST (blue) or B6 (orange) transplanted microbiomes. (F) Targeted GC-MS analysis of cecal short-chain fatty acids; *$p < 0.05$ by Student's t-test. Data are mean ± SEM, n = 6-7 mice/recipient group and n = 2-3 mice/donor group. For metagenomics analysis n=5 mice/recipient group.

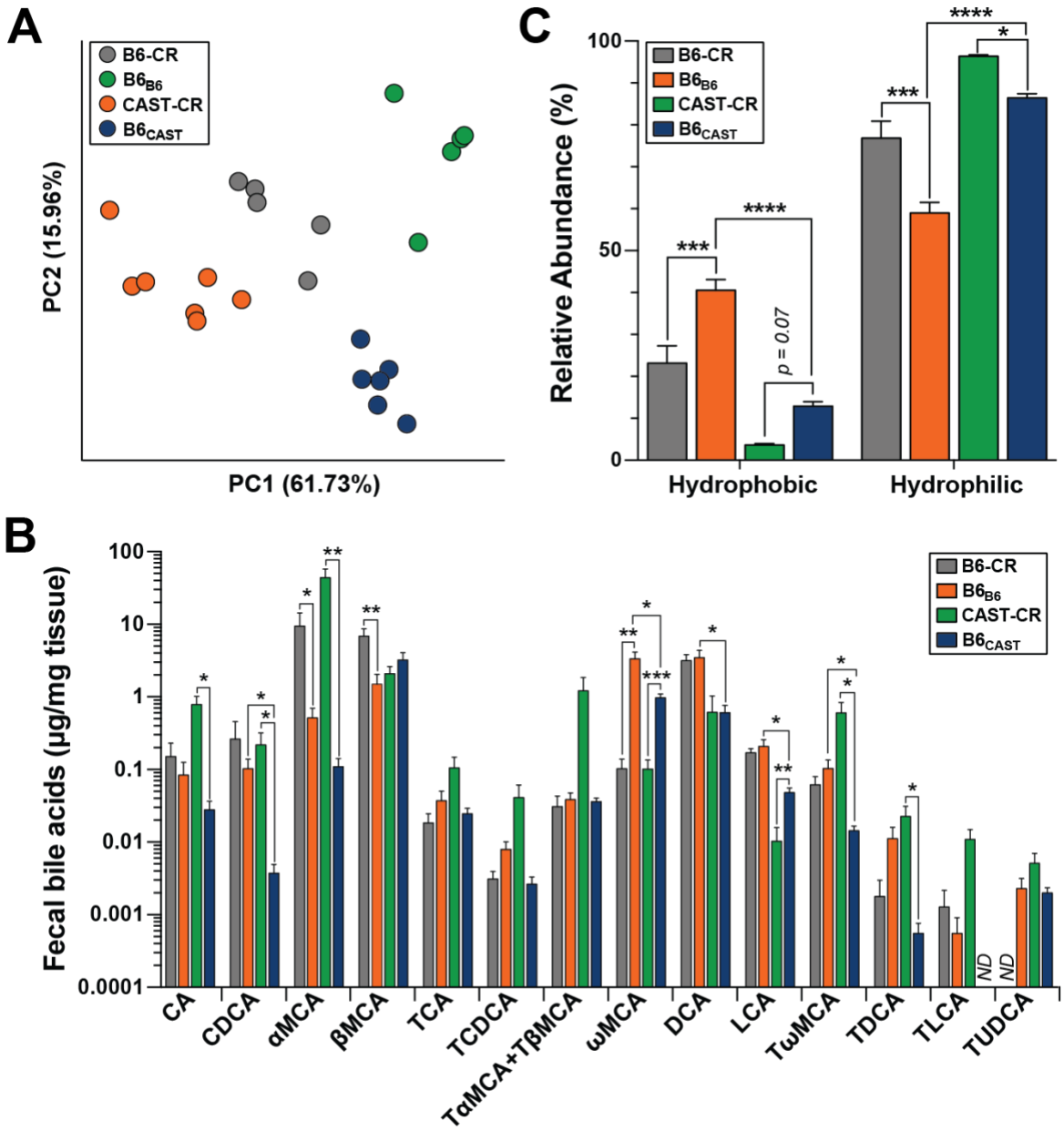


Figure 2.5. B6 and CAST microbiota produce different bile acid profiles. (A) Principal component analysis of the square root proportion of 14 major bile acid species (ng/mg). Each dot represents the bile acid profile of an individual mouse. Percent variation explained by each PC is shown in parentheses. (B) Abundance of fecal bile acids, and (C) relative abundance of hydrophobic and hydrophilic BA species determined by UPLC-MS/MS from fecal samples collected at 12-weeks

post-colonization. No data (ND). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ by one-way ANOVA with Bonferroni's multiple comparisons test. Data are mean ± SEM, n= 6-7 for transplant recipients, and n= 5 for CR mice.

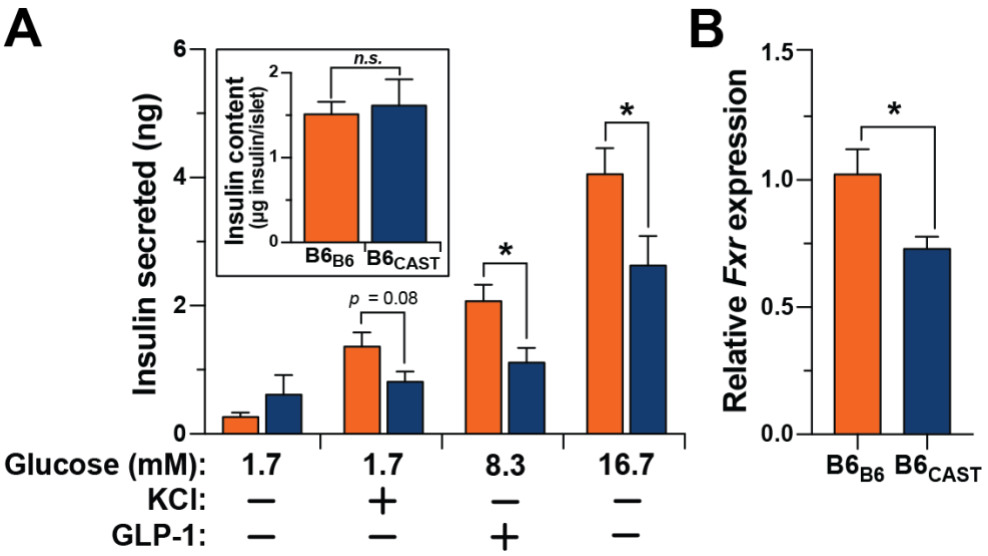
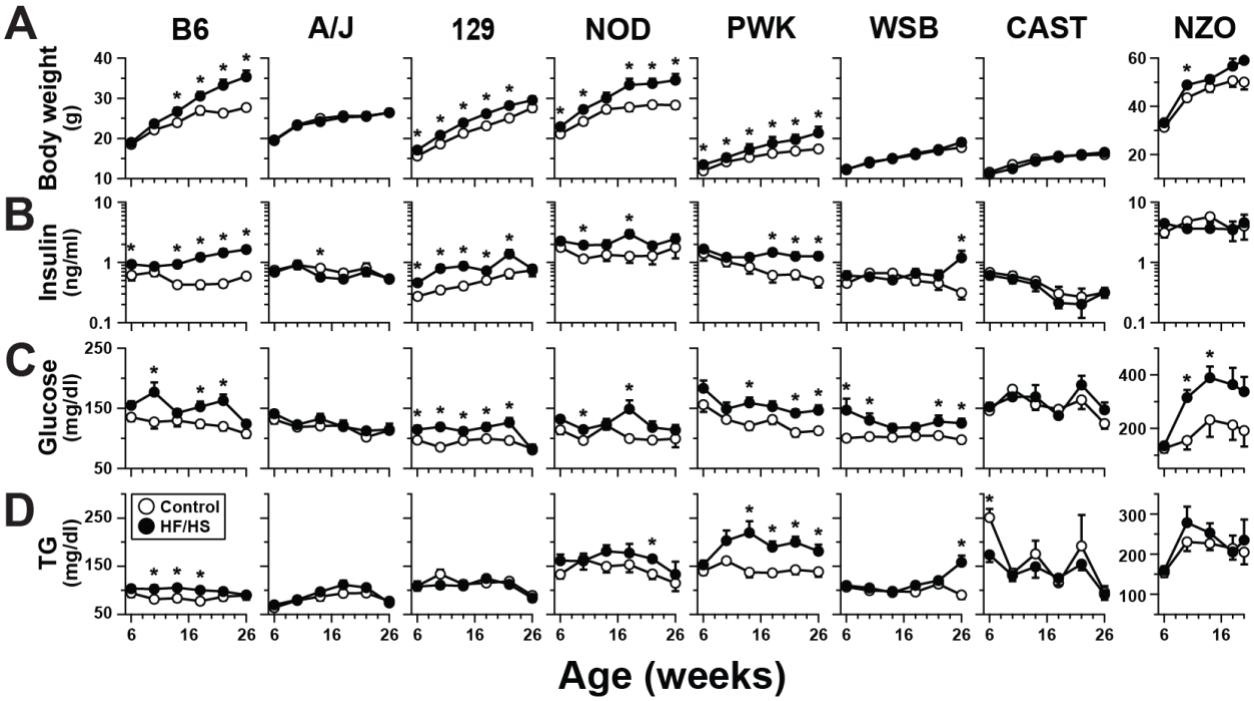
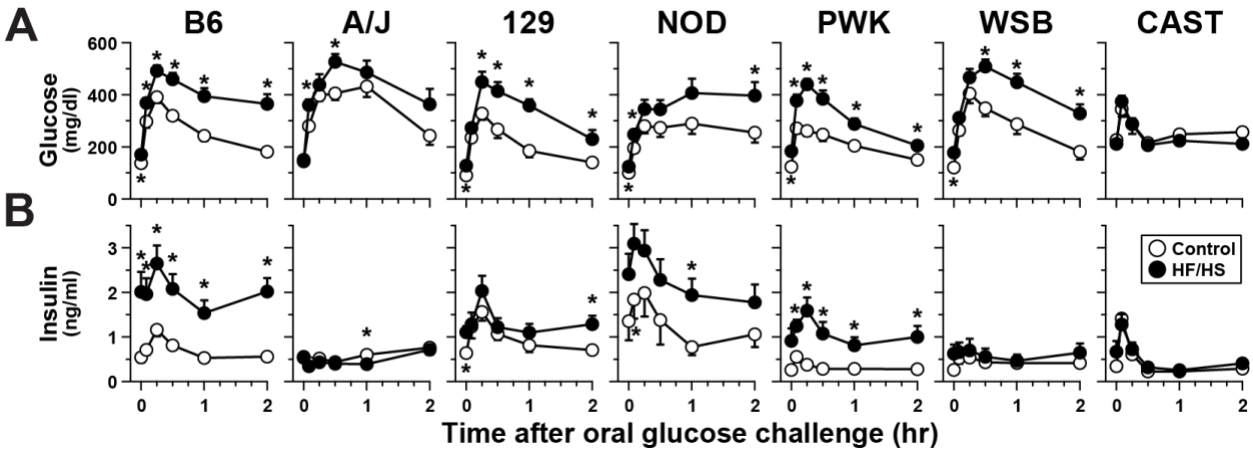


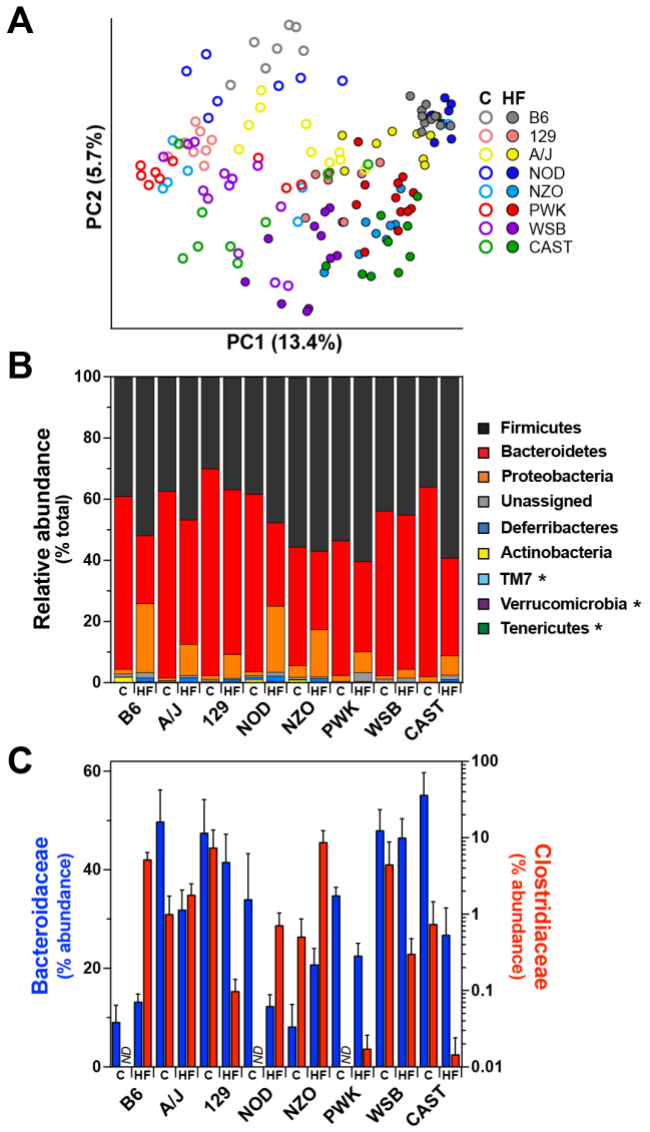
Figure 2.6. CAST and B6 microbiomes differentially regulate insulin secretion and *Fxr* expression in pancreatic islets. (A) Total islet insulin content and glucose-stimulated insulin secretion in response to low glucose (3.3 mM), low glucose plus KCl (40 mM), high glucose (16.7), and high glucose plus GLP-1 (100 mM) from islets isolated from $B6_{B6}$ and $B6_{CAST}$ mice. The number of islets and the insulin content per islet were not different between the groups. (B) Relative expression of *Fxr* mRNA from isolated islets. Supplememntal Figure 2.6 shows microbiota composition for donor and transplanted communities. *$p < 0.05$ by Student's t-test. Data are mean ± SEM, n = 5.



Supplemental Figure 2.1. **Segregation of metabolic syndrome among the founder strains of the Collaborative Cross (CC). Related to Figure 2.1.** (A) Body weight, (B) fasting plasma insulin, (C) glucose and (D) triglycerides (TG) were determined at various ages for CC founder mice fed either a high-fat/high-sucrose (HF/HS) or a control diet for 22 weeks. Note differences in Y-axis scale for NZO mice. *$p < 0.05$. data are mean ± SEM, $n \geq 9$ mice/genotype/diet.

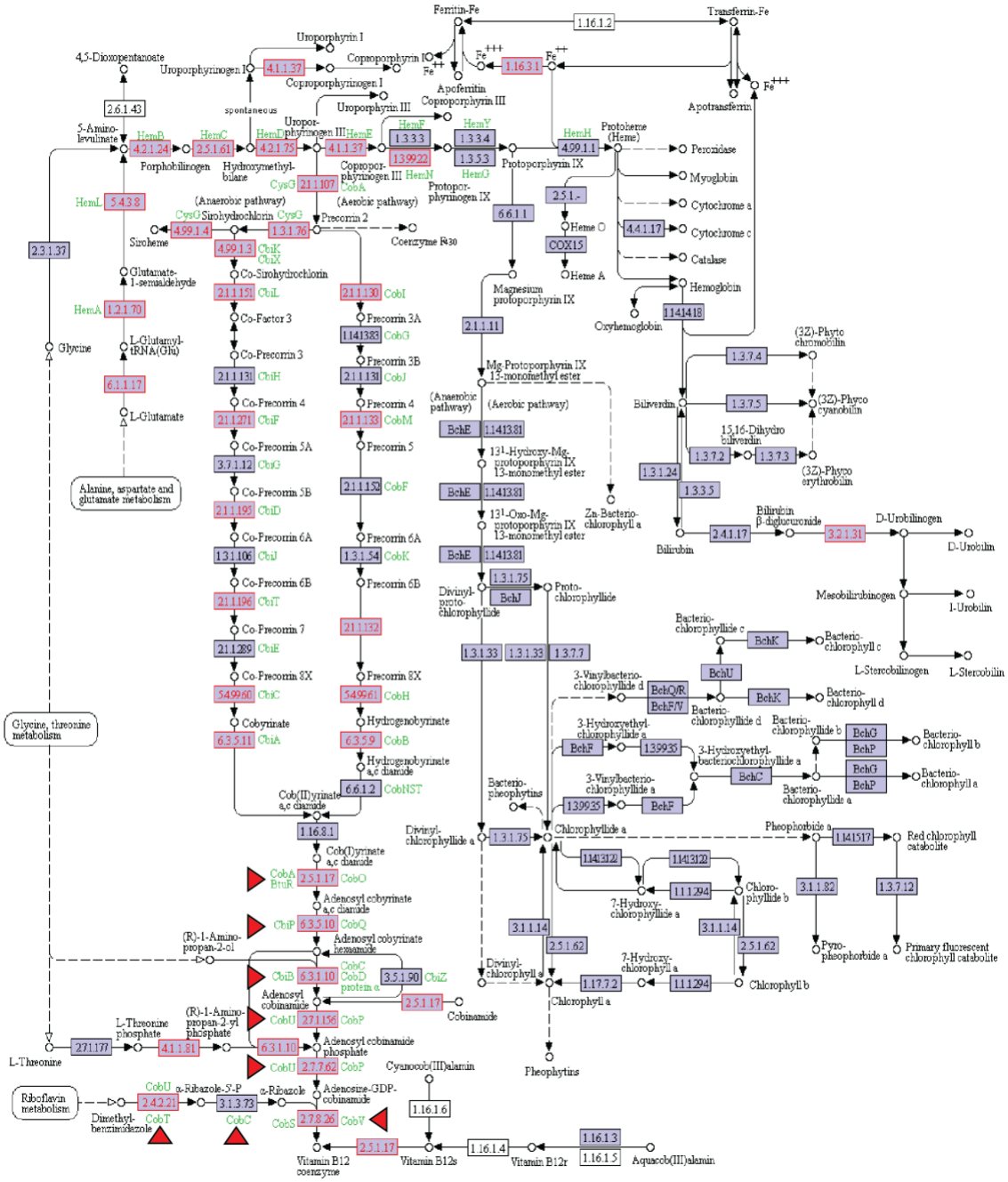


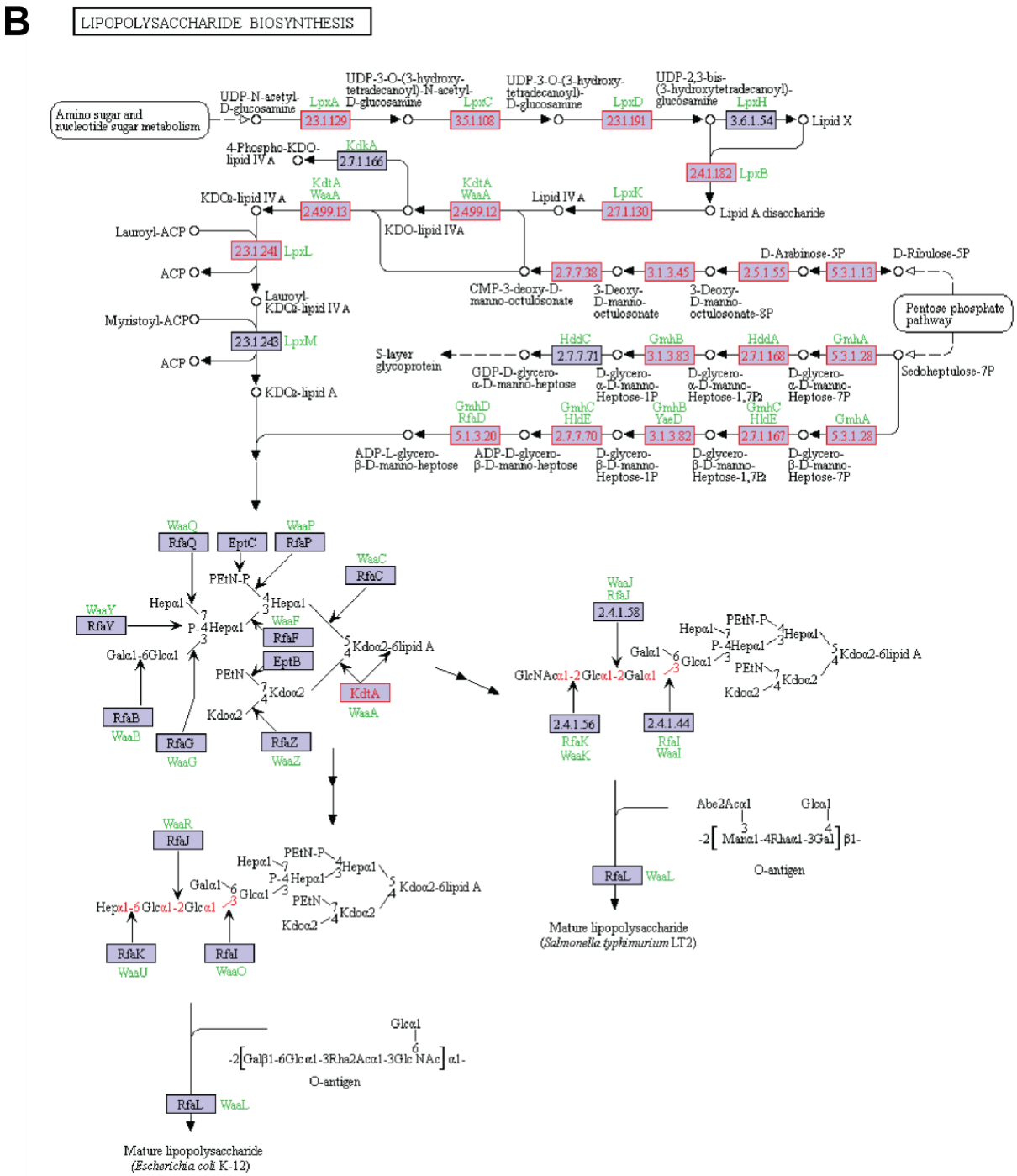
Supplemental Figure 2.2. Diet-induced glucose tolerance and insulin sensitivity differ among CC founder mice. Related to Figure 2.1. Male mice were maintained on either a control or high-fat/high-sucrose (HF/HS) diet. At 22 weeks of age, mice were given glucose bolus (2 g/kg body weight) via oral gavage following a 4-hour fast. Blood was collected via retro-orbital bleed at 0, 5, 15, 30, 60, and 120 minutes following the glucose bolus, and used to determine plasma (A) glucose and (B) insulin levels. NZO mice did not survive 22 weeks of age on HF/HS diet. *$p < 0.05$. Data are mean ± SEM, $n \geq 9$ mice/genotype/diet.



Supplemental Figure 2.3. Host genotype and diet affect microbial composition. Related to Figure 3.2. (A) Principal Coordinate Analysis (PCoA) of unweighted UniFrac distances for the cecal microbiota of the founder mice. Open symbols, control diet; filled symbols HF/HS diet. (B) Relative abundance of 8 major microbiota phyla identified in cecal contents from CC founder mice maintained on control (abv. C) or HF/HS (abv. HF) diet for 22 weeks. Phyla ordered by mean abundance; * denotes mean phyla abundance < 1%. (C) Relative abundance of Bacteroidaceae and Clostridicaceae in CC founder strains. Family not detected marked as ND. Data are mean ± SEM, n ≥ 9 mice/genotype/diet.

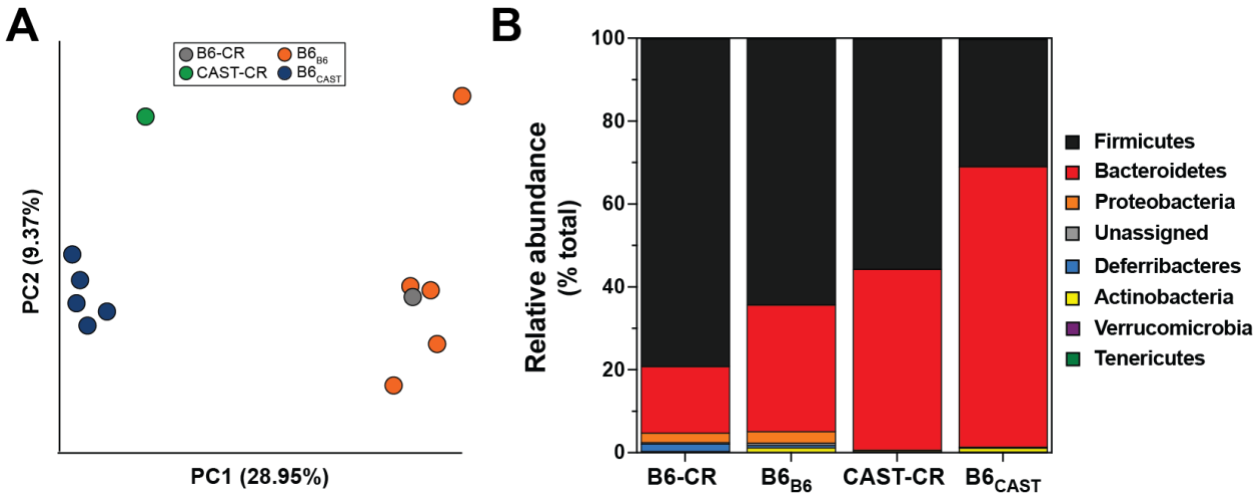
A


PORPHYRIN AND CHLOROPHYLL METABOLISM
Ferritin-Fe
Transferrin-Fe
Apoferritin
Apotransferrin
Uroporphyrin I
Coproporphyrin I
Uroporphyrinogen I
Coproporphyrinogen I
Uroporphyrin III
Coproporphyrin III
4,5-Dioxopentanoate
5-Amino-levulinate
Porphobilinogen
Hydroxymethyl-bilane
Uropor-phyrinogen III
Copropor-phyrinogen III
Protopor-phyrinogen IX
Protoporphyrin IX
Protoheme (Heme)
Peroxidase
Myoglobin
Cytochrome a
Cytochrome c
Catalase
Hemoglobin
Oxyhemoglobin
Heme O
Heme A
COX15
Magnesium protoporphyrin IX
Mg-Protoporphyrin IX 13-monomethyl ester
13¹-Hydroxy-Mg-protoporphyrin IX 13-monomethyl ester
13¹-Oxo-Mg-protoporphyrin IX 13-monomethyl ester
Zn-Bacterio-chlorophyll a
Divinyl-proto-chlorophyllide
Proto-chlorophyllide
Divinyl-chlorophyllide a
Chlorophyllide a
Pheophytins
Pheophorbide a
Red chlorophyll catabolite
Chloro-phyllide b
7-Hydroxy-chlorophyllide a
7-Hydroxy-chlorophyll a
Chlorophyll a
Chlorophyll b
Pyro-phosphorbide a
Primary fluorescent chlorophyll catabolite
Biliverdin
15,16-Dihydro biliverdin
(3Z)-Phyto chromobilin
(3Z)-Phyco cyanobilin
(3Z)-Phyco erythrobilin
Bilirubin
Bilirubin β-diglucuronide
D-Urobilinogen
D-Urobilin
Mesobilirubinogen
I-Urobilin
L-Stercobilinogen
L-Stercobilin
Bacterio-chlorophyllide c
Bacterio-chlorophyll c
3-Vinylbacterio-chlorophyllide d
Bacterio-chlorophyll d
Bacterio-chlorophyllide d
3-Hydroxyethyl-chlorophyllide a
3-Hydroxyethyl-bacteriochlorophyllide a
Bacterio-chlorophyllide b
Bacterio-chlorophyll b
3-Vinylbacterio-chlorophyllide a
Bacterio-chlorophyllide a
Bacterio-chlorophyll a
Bacterio-pheophytins
Glycine
L-Glutamyl-tRNA(Glu)
L-Glutamate
Glutamate-1-semialdehyde
Alanine, aspartate and glutamate metabolism
Glycine, threonine metabolism
Sirohydrochlorin
Siroheme
Precorrin 2
Coenzyme F430
Co-Sirohydrochlorin
Co-Factor 3
Co-Precorrin 3
Co-Precorrin 4
Co-Precorrin 5A
Co-Precorrin 5B
Co-Precorrin 6A
Co-Precorrin 6B
Co-Precorrin 7
Co-Precorrin 8X
Cobyrinate
Hydrogenobyrinate
Hydrogenobyrinate a,c diamide
Cob(II)yrinate a,c diamide
Cob(I)yrinate a,c diamide
Adenosyl cobyrinate a,c diamide
Adenosyl cobyrinate hexamide
(R)-1-Amino-propan-2-ol
Adenosyl cobinamide
(R)-1-Amino-propan-2-yl phosphate
Adenosyl cobinamide phosphate
Cobinamide
L-Threonine phosphate
L-Threonine
Cyanocob(III)alamin
Adenosine-GDP-cobinamide
α-Ribazole
α-Ribazole-5'-P
Riboflavin metabolism
Dimethyl-benzimidazole
Vitamin B12 coenzyme
Vitamin B12s
Vitamin B12r
Aquacob(III)alamin
Fe++
Fe+++
spontaneous
(Anaerobic pathway)
(Aerobic pathway)

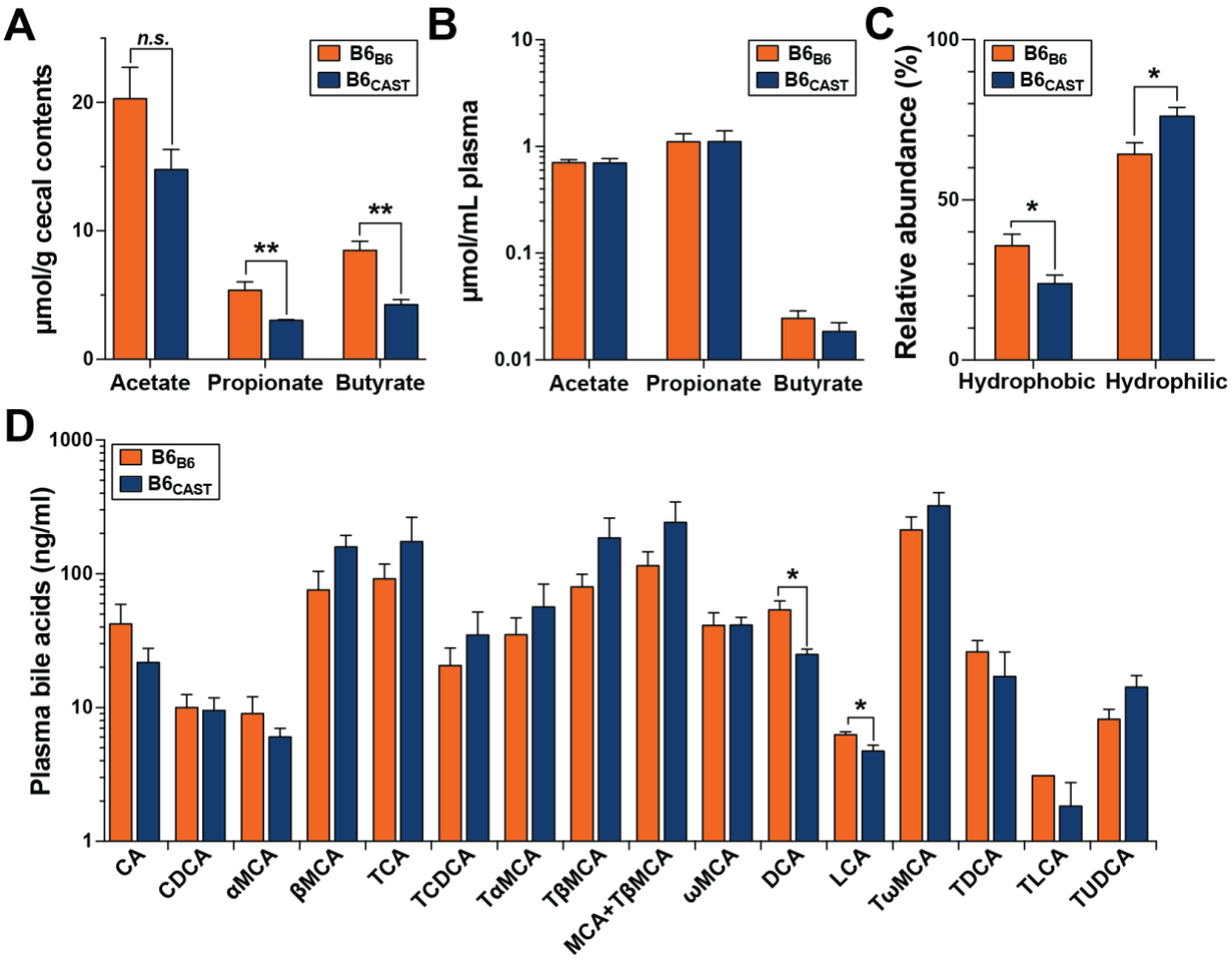


Supplemental Figure 2.4. Microbial pathways enriched in CAST-derived microbiota. Related to Figure 2.4. (A) Vitamin B12 biosynthesis is functionally enriched in CAST-derived

microbiota. KEGG pathway for “Porphyrin and Chlorophyll Metabolism” (map00860). Fifty-six genes within the pathway were more abundant in $B6_{CAST}$ than $B6_{B6}$ microbiota. ECs higher in the $B6_{CAST}$ microbiota compared to $B6_{B6}$ microbiota colored in red. The KO annotations for the 56 genes in the Porphyrin and Chlorophyll Metabolism pathway were input to the Reconstruct Module of KEGG Mapper. Red triangles indicate members of the module for vitamin B12 (cobalamin) biosynthesis, which is nearly complete; the missing block in this module corresponds to EC number 3.1.3.73, which is not boxed in red. (B) Lipopolysaccharide biosynthesis is functionally enriched in CAST-derived microbiota. KEGG pathway for “Lipopolysaccharide Biosynthesis” (map00540). Forty-six genes within the pathway were more abundant in $B6_{CAST}$ than $B6_{B6}$ microbiota. ECs higher in the $B6_{CAST}$ microbiota compared to $B6_{B6}$ microbiota colored in red.



Supplemental Figure 2.5. Microbiota composition of B6-CR, CAST-CR, $B6_{B6}$ and $B6_{CAST}$ mice used for insulin secretion studies. Related to Figure 2.6. (A) Principal Coordinate Analysis (PCoA) of unweighted Unifrac distances, and (B) relative abundance of microbial phyla. Data are mean ± SEM, n = 5 for $B6_{B6}$ and $B6_{CAST}$ mice and n = 1 for CR mice



Supplemental Figure 2.6. SCFA and BA measurements in $B6_{B6}$ and $B6_{CAST}$ mice. Related to Figure 2.6. (A) Cecal and (B) plasma SCFA concentrations from $B6_{B6}$ and $B6_{CAST}$ animals used for insulin secretion studies as determined by GC-MS. (C) Relative abundance of hydrophobic and hydrophilic bile acid (BA) species, and (D) abundance of major plasma BA species determined by UPLC-MS/MS. *$p < 0.05$, **$p < 0.01$. Data are mean ± SEM, n = 5

Supplemental Table 2.1. Related to Figures 2.1-5. Composition of diets used in this study

Low glycemic control diet (TD.08810)

Component	g/Kg
Casein	210.0
L-Cystine	3.0
INACTIVE Hi-Maize 220 (Resistant Starch)	500.0
Maltodextrin	100.0
Sucrose	39.1
Anhydrous Milkfat	20.0
Soybean Oil	20.0
Cellulose	35.0
Mineral Mix, AIN-93G-MX (94046)	35.0
Vitamin Mix, AIN-93-VX (94047)	15.0
Choline Bitartrate	2.8
TBHQ, antioxidant	0.0
Yellow Food Color	0.1

High-fat high-sucrose diet (TD.08811)

Component	g/Kg
Casein	195.0
L-Cystine	3.0
Sucose	340.0
Corn Starch	56.9
Maltodextrin	60.0
Anhydrous Milkfat	210.0
Soybean Oil	20.0
Cellulose	50.0
Mineral Mix, AIN-93G-MX (94046)	43.0
Vitamin Mix, AIN-93-VX (94047)	19.0
Choline Bitartrate	3.0
TBHQ, antioxidant	0.0
Green Food Color	0.1

CHAPTER 3: Genetic Determinants of Gut Microbiota Composition and Bile Acid Profiles in Mice

Kemis JH[*], *Linke V, Barrett K, Boehm F, Traeger LL, Keller MP, Rabaglia ME, Schueler KL, Stapleton DS, Yandell BS, Broman KW, Gatti DM, Churchill GA, Amador-Noguez D, Russel J, Coon J, Attie AD, Rey FE*[+].

[*] *indicates lead author,* [+] *indicates corresponding author*

ABSTRACT

The microbial communities that inhabit the distal gut vary widely among individuals. While host genetic variation is a known factor that influences gut microbiota composition, the mechanisms underlying this variation have not been fully elucidated. Bile acids (BAs) are hormones that are produced by the host and modified by gut bacteria. BAs can serve as environmental cues and nutrients for bacteria, but they can also have antibacterial effects. We hypothesized that host genetic variation in BA metabolism impact gut microbiota composition. To address this, we used the Diversity Outbred (DO) stock, a population of genetically distinct mice derived from eight founder strains. We characterized the fecal microbiota composition and plasma and cecal BA profiles of 400 DO mice fed a high-fat high-sucrose diet for ~22 weeks. Using quantitative trait loci (QTL) analysis, we identified several genomic regions associated with both microbial abundance and BA levels. These overlapping QTL included taxa previously associated with BAs, including *Akkermansia muciniphila* and the Peptostreptococcaeae family. Notably, we found overlapping QTL for *Turicibacter sp.* and plasma cholic acid that mapped to a locus containing the gene for the ileal bile acid transporter, *Slc10a2*. Mediation analysis and follow-up validation experiments suggest that differences in *Slc10a2* gene expression associated with the different strains influences levels of both traits and revealed novel interactions between *Turicibacter sp.* and BAs. Together, our work provides insights into the mechanisms underlying host-gut microbe interactions and illustrates how systems genetics can be used to generate hypotheses elucidating these interactions.

INTRODUCTION

The intestinal microbiota has profound effects on host physiology and health (Le Chatelier et al., 2013; Clemente et al., 2012; Sommer and Bäckhed, 2013). The composition of the gut microbiota is governed by a combination of environmental factors including diet, drugs, maternal seeding, cohabitation, and host genetics (Lozupone et al., 2012; Rothschild et al., 2018; Zhernakova et al., 2016). Together, these environmental and genetic factors cause substantial inter-individual variation in microbiota composition and modulate disease risk (Hall et al., 2017; Ussar et al., 2016). Alterations in the composition of the microbiota are associated with a spectrum of pathologies including obesity, diabetes, metabolic syndrome, and inflammatory diseases (Clemente et al., 2018; Khan et al., 2014; Qin et al., 2012; Turnbaugh et al., 2006). A major challenge in the field is deciphering how host genetics and environmental factors interact to shape the composition of the gut microbiota and the mechanisms by which these interactions can be manipulated to improve health outcomes.

Several mouse and human studies have examined the role of host genetics in shaping the composition of the gut microbiota. The effects of genetics on the microbiome have been highlighted by composition differences among inbred mouse strains (Kreznar et al., 2017; Parks et al., 2013) and through the loss of metabolism and immune-related genes (Kurilshikov et al., 2017). Additionally, quantitative trait loci (QTL) analysis in mice have identified genetic loci that control for the abundance of different taxa (Belheouane et al., 2017; Benson et al., 2010; Leamy et al., 2014; McKnite et al., 2012). Twin studies and genome wide association studies (GWAS) have identified heritable taxa and SNPs associated with specific gut microbes. However, comparing these studies is often difficult because of differences in environmental variables among populations. Despite these confounding effects, some associations are consistently found among

geographically discrete populations, such as the association between *Bifidobacterium* and the lactase (*LCT*) gene locus (Blekhman et al., 2015; Bonder et al., 2016; Goodrich et al., 2016), indicating specific taxa are regulated by host genetics.

The host and gut microbiome interact through the production and modification of metabolites, many of which impact host physiology (Herrema et al., 2017; Krautkramer et al., 2016; Ridlon et al., 2016; Romano et al., 2017; Wang et al., 2011). Among these, bile acids (BAs) are particularly relevant for understanding the relationship between host genetic variation and gut microbiota composition. BAs are host-derived and microbial-modified metabolites that regulate both the gut microbiome and host metabolism (Kuipers et al., 2014; Ridlon et al., 2006; Wahlström et al., 2016). BAs are synthesized in the liver from cholesterol, stored in the gallbladder and are secreted in the proximal small intestine where they facilitate absorption of fat-soluble vitamins and lipids. Once in the intestine, BAs can be metabolized by gut bacteria through different reactions including deconjugation, dehyrdoxylation, epimerization, and dehydrogenation, to produce secondary BAs with differential effects on the host (Ridlon et al., 2006, 2016). In addition to their direct effects on the host, BAs shape the gut microbiota composition through antimicrobial activities (Islam et al., 2011; Zheng et al., 2017). The detergent properties of BAs cause plasma membrane damage and the bactericidal activity of a BA molecule corresponds to its hydrophobicity (Begley et al., 2005). Additionally, the microbiota regulates primary BA synthesis through regulation of the nuclear factor FXR(Sayin et al., 2013).

To investigate how genetic variation affects gut microbiota and BA profiles, we used the Diversity Outbred (DO) mouse population, which is a heterogenous population derived from eight founder strains: C57BL6/J, A/J, 1291/SvImJ, NOD/ShiLtJ, NZO/HiLtJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ (Churchill et al., 2012; Svenson et al., 2012). These eight strains capture a large breadth

of genetic diversity found in laboratory and wild mouse populations. Additionally, the founder strains harbor distinct gut microbial communities and exhibit disparate metabolic responses to diet-induced metabolic disease (Kovacs et al., 2011; Kreznar et al., 2017; O'Connor et al., 2014). The DO population is maintained by a randomized outbreeding strategy so that the genome of each DO mouse is a mosaic of the eight founder strains, making it an ideal resource for high-resolution genetic mapping of microbial and metabolic traits. Since each position of a DO mouse genome can be attributed to a founder strain, this resource also allows for subsequent validation studies in the founder strains.

We characterized the intestinal microbiota composition and plasma and cecal BA profiles in ~400 genetically district DO mice fed a high-fat/high-sucrose diet for ~22 weeks and performed quantitative trait loci (QTL) analysis to identify host genetic loci associated with these traits. Specifically, we focused our analysis on potentially pleiotropic loci, which we defined as a single genetic locus that associates with both microbial and BA traits. Our analysis revealed several instances of microbial and metabolite traits attributed to the same DO founder haplotypes mapping to the same position of the mouse genome, including a locus associated with plasma BA levels and the disease-modulating organism *Akkermansia muciniphila.* Additionally, we identified the ileal BA transporter *Slc10a2* as a candidate gene that regulates the abundance of *Turicibacter sp.* abundance and plasma cholic acid levels.

RESULTS AND DISCUSSION

We investigated the effects of genetic variation on gut microbiota composition and host BA profiles using a cohort of ~400 DO mice fed a high-fat high-sucrose diet (45% kcal fat and 34% sucrose) for 22 weeks, starting at weaning. Additionally, we incorporated in our analyses clinical weight traits collected from the same mice that were previously published (Keller et al., 2018) (Figure 3.1A). All mice were individually housed throughout the duration of the study to minimize microbial exchange by coprophagy and to monitor food intake.

Variability and associations among microbial, bile acid, and clinical traits

We found substantial variation in plasma and cecal BA profiles across the 400 mice (Table 3.1) as demonstrated by the variability seen in the levels of primary BAs in plasma and cecal contents (Figure 3.1C-D). Gut microbiota composition was profiled by 16S rRNA gene amplicon sequencing of DNA extracted from fecal samples collected the day of sacrifice (21-25 weeks-old). Within the cohort there were 907 unique Exact Sequence Variants (ESVs), (100% operational taxonomic units defined with dada2 (Callahan et al., 2016)), which were agglomerated into 151 lower taxonomic rankings (genus, family, order, class, phyla). The microbial traits represented each of the major phyla found in the intestine and the relative abundance of these phyla was highly variable among the DO mice (Figure 3.1B). For instance, the abundance of taxa classified to the Bacteroidetes phylum ranged from 1.17 – 89.28%.

For subsequent analysis, we identified a core measurable microbiota (CMM), which we defined as taxon found in at least 20% of the mice (Benson et al., 2010). This was done to remove the effects of excessive variation in the data due to bacterial taxa that were low abundance and/or sparsely distributed. In total, the CMM was comprised of 86 ESVs and 42 agglomerated taxa (Table 3.2)(ESV key Supplemental Table 3.1). The CMM traits represent a small fraction of the

total microbes detected, but account for 94.5% of the rarefied sequence reads, and therefore constitute a significant portion of the identifiable microbiota.

Since mice were received in waves of 100, we examined whether animals in each wave were more similar to each other than mice in other waves. The fecal microbiota composition significantly clustered by wave ($p < 0.001$, PERMANOVA) and sex ($p < 0.001$, PERMANOVA) (Supplemental Figure 3.1). PCA analysis of plasma and cecal bile acids showed a significant effect of sex, but not wave, on both plasma ($p < 0.01$, Kruskal Wallis) and cecal BA profiles ($p < 0.0001$, Kruskal Wallis) (Supplemental Figure 3.2)

There is substantial evidence implicating gut microbiota and BAs in metabolic disease development (Kuipers et al., 2014; Wahlström et al., 2016). To identify potential relationships among these traits, we performed correlation analysis which yielded many significant associations after FDR correction (FDR < 0.05) (Table 3.3). We found significant positive and negative associations between body weight and different Lachnospiraceae ESVs. This dual correlation is consistent with previous studies that have found this bacterial family to be positively (Org et al., 2015) and negatively (Kreznar et al., 2017; O'Connor et al., 2014) correlated with obesity and other metabolic traits. Additionally, we identified significant associations between BAs and body weight. Body weight over time was inversely correlated with plasma levels of deoxycholic acid (DCA), taurochenodeoxycholic acid (TCDCA) and taurocholic acid (TCA). Conversely, cecal levels of muricholic acid (MCA) and ursodeoxycholic acid (UDCA) were positively correlated with body weight. The unconjugated plasma BAs allo-cholic acid (ACA), UDCA, 7-dehydrocholic acid (7-dHCA), hyodeoxycholic acid (HDCA), DCA, MCA and cholic acid (CA) were all positively associated with *Turicibacter* abundance. Interestingly, the only cecal bile acid to negatively correlate with *Turicibacter* was TCA ($r = -0.2619$, $p = 0.0035$). We also found several

taxa among the Lachnospiraceae family were positively associated with conjugated secondary cecal bile acids including tauroursodeoxycholic acid (TUDCA), TCA, taurodeoxycholate (TDCA), glycodeoxycholic acid (GDCA), taurolithocholic acid (TLCA) and TCDCA. This is consistent with a previous study that found members of the Lachnospiraceae family were positively correlated with all secondary bile acids (Theriot et al., 2016).

Bacterial taxa and bile acids associate to host genome

To identify associations between regions of the mouse genome and the clinical and molecular traits discussed above, we performed QTL analysis using the r/qtl2 package (Broman, 2018). We used sex, days on the diet, and experimental cohort (wave) as covariates. We identified 459 QTL for bacterial (306), bile acid (131), and body weight (22) traits (Figure 3.2, Table 3.4).

Of the microbial QTL, we found 190 QTL for 76 distinct bacterial ESVs from four phyla that met a LOD cut-off of > 5.5. ESVs with the strongest QTL (LOD ~ 8) are classified to the Clostridiales order and map on chr 12 at ~33 Mbp, the Lachnospiraceae family on chr 2 at 164 Mbp, and the S24-7 family on chr 2 at ~115 Mbp. We also identified 116 QTL for microbial taxa collapsed by taxonomic assignment (i.e., genus to phylum). The genera *Lactococcus* and *Akkermansia* were also associated with host genetic variation, which is consistent with previous studies (Benson et al., 2010; Davenport et al., 2015; Leamy et al., 2014; Org et al., 2015).

Similarly, BA QTL mapped to multiple loci spanning the mouse genome and most BA traits mapped to multiple positions. BA synthesis and metabolism are regulated by multiple host signaling pathways: there are > 17 known host enzymes involved in the production of BAs (Wahlström et al., 2016), transporters, which are critical role for maintaining the enterohepatic circulation and BA homeostasis, and receptors that respond to BA in a variety of host tissues (de Aguiar Vallim et al., 2013; Martinot et al., 2017; Russell, 2003). Therefore, it is not surprising that

our results indicate BA levels are polygenic and shaped by multiple host factors. We observed multiple instances of related BA species associating to the same genetic locus. These overlapping QTL may indicate the presence of a pleiotropic locus. Interestingly, several of these loci associate with levels of related BA species in different stages of microbial modification. For example, cecal TCA and plasma CA QTL overlap on chr 7 at 122 Mbp. Likewise, four BA QTL that are all derivatives of the secondary BA DCA, including plasma TDCA and cecal DCA, isodeoxycholic acid (IDCA), and HDCA overlap on chr 12 at ~99 – 104 Mbp. For the cecal BA, the WSB founder haplotype was associated with higher levels of these three BA, while the NOD founder haplotype was associated with lower levels. The opposite pattern was observed for plasma TDCA, where the NOD and WSB haplotype was associated with higher and lower levels, respectively (Supplemental Figure 3.3E-H).

We also identified overlapping QTLs on chr 11 at ~71 Mbp for cecal levels of the secondary BAs lithocholic acid (LCA) and isolithocholic acid (ILCA), the isomer of LCA produced by bacterial 3α-hydroxylation (Supplemental Figure 3.3A). Higher levels of these cecal BAs are associated with the 129 founder haplotype and lower levels are associated with the A/J founder haplotype (Supplemental Figure 3.3B-C). We identified the positional candidate gene *Slc13a5*, which is a sodium-dependent transporter that mediates cellular uptake of citrate, an important precursor in the biosynthesis of fatty acids and cholesterol (Inoue et al., 2002). Recent evidence indicates *Slc13a5* influences host metabolism and energy homeostasis (Birkenfeld et al., 2011; von Loeffelholz et al., 2017; Pesta et al., 2015). *Slc13a5* is a transcriptional target of pregnane X receptor (PXR) (Li et al., 2015), which also regulates the expression of genes involved in the biosynthesis, transport, and metabolism of BAs (Staudinger et al., 2001).

Overlapping bacterial and bile acid QTL

Given the known interplay among gut microbes, BAs, and host genetics, it is reasonable to expect that some of the microbial and BAs QTLs might exhibit pleiotropic effects. To examine this possibility, we identified instances of microbial and BA QTL mapping to the same position. In total, 17 instances of overlapping microbial and BA QTL were identified on 12 chromosomes. This co-mapping indicates there are some QTLs with pleiotropic effects on BAs and the microbiota, suggesting that genetic variation influencing host BA profiles has an effect on compositional features of the gut microbiota, or genetic-driven variation in microbiota composition alters BAs. Additionally, these co-mapping traits may be evidence of interactions between the traits (Civelek and Lusis, 2014).

We found QTL for an unknown genus in the Peptostreptococcaceae family overlapping with the hotspot containing QTL for plasma levels of CA, CDCA, UDCA, MCA, 7-dHCA and glycodehydrocholic acid (G-dHCA) on chr 3 between ~40-50 Mbp. These QTL all show the same founder strain haplotype effects, where the NOD haplotype is associated with higher levels of these traits (Supplemental Figure 3.5A-F). *Peptostreptococcus productus*, a member of the Peptostreptococcaceae family, has 3α-, 3β-, and 7β-hydroxysteroid dehydrogenases and is capable of oxidation and epimerization of BAs (Edenharder et al., 1989). Several of these secondary bile acids require 7β-epimerization, including hyocholic acid (HCA) and UDCA which is produced from 7β-epimerization of CDCA (Wahlström et al., 2016), which may help explain why these BAs co-map with Peptostreptococcaceae abundance. An interesting candidate within the QTL peak region is progesterone receptor membrane component 2 (*Pgrmc2*), which is expressed in bile sensitive tissues such as intestine, liver and brown adipose (Chen et al., 2010). PGRMC2 is predicted to be a membrane receptor (Gerdes et al., 1998), which binds to P450 cytochrome proteins and has similar characteristics to PGRMC1 (Wendler and Wehling, 2013). The shared

sequence between *Pgrmc2* and *Pgrmc1* is especially interesting in the context of BAs because *Pgrmc1* directly binds to Cyp7a1 (Hughes et al., 2007), a P450 cytochrome protein responsible for the regulation of BA synthesis. These data suggest that *Pgrmc2* may be a novel gene involved in BA signaling and/or homeostasis.

On chr 1 at ~90 – 100 Mbp, we identified overlapping QTL for *Akkermansia muciniphila* and plasma levels of CA, MCA and 7-dHCA, where the NZO haplotype is positively associated, and the 129 haplotype is negatively associated with each of these traits (Supplemental Figure 3.4A-D). Significant positive correlations were also found between the abundance of *A. muciniphila* and plasma levels of CA ($r = 0.19$, $p < 0.0045$) and MCA ($r = 0.17$, $p < 0.0149$) (Supplemental Figure 3.4F-H). These observations are particularly striking given the recent studies associating the abundance of *A. muciniphila* and BAs. For example, Pierre et al. found the abundance of *A. muciniphila* was positively correlated with higher levels of circulating primary bile acids (Pierre et al., 2016) and administration of the secondary bile acid UDCA was found to increase its abundance (Van den Bossche et al., 2017). Furthermore, supplementation with up to 1% porcine bile extract increased *A. muciniphila* growth *in vivo* (van der Ark et al., 2017). In the intestine, *A. muciniphila* degrades host mucins (Derrien et al., 2004), which provide growth substrates for other intestinal commensals (Belzer and de Vos, 2012). Notably, BAs have a stimulatory effect on mucin secretion as a defense mechanism to protect the gastrointestinal epithelium against potential BA toxicity (Klinkspoor et al., 1999; Shekels et al., 1996). Therefore, the positive correlation between bile acid levels and *A. muciniphila* may be the result of this stimulatory effect where greater mucin secretion from BA stimulation can support a larger intestinal *A. muciniphila* population.

There is growing interest in the potential therapeutic role of *A. muciniphila* since it has been associated with improvements in host metabolic syndrome (Cani and de Vos, 2017; Everard

et al., 2013; Org et al., 2015; Plovier et al., 2017) and plays a key role in regulating intestinal barrier function and mucosal immunity (Derrien et al., 2011; Everard et al., 2013). Strikingly, we found several candidate genes under the QTL region on chr 1 related to host lipid metabolism and immunity (Supplemental Figure 3.4E). Top immune-related genes include *Lrrfip1*, a transcription regulator of TLR pathway signaling (Arakawa et al., 2010) and TNF expression (Shi et al., 2014), and *Gpr35*, a G protein-coupled receptor for the mucosal chemokine CXCL17 (Maravillas-Montero et al., 2015). Candidate lipid metabolism genes include *Farp2* and *Stk25*, which were previously identified as candidate genes for plasma HDL levels (Su et al., 2009a). In fact, several mouse studies using F2 crosses have identified QTL for plasma cholesterol and HDL levels at this position on chr 1 (Ishimori et al., 2004; Purcell-Huynh et al., 1995; Su et al., 2009a, 2009b) including one where the association was driven by the 129 haplotype (Ishimori et al., 2004). The plasma HDL QTL found at the position as the microbial and metabolite QTL on chr 1 is particularly interestingly because *A. muciniphila* abundance has been associated with elevated HDL levels (Fu et al., 2015) and administration of a purified protein from this microbe decreased HDL and LDL cholesterol levels, indicating it may have a regulatory impact on cholesterol metabolism (Plovier et al., 2017). Therefore, the co-mapping *A. muciniphila* and plasma bile acid traits seen in our study may be driven by another unmeasured factor or plasma lipid, which explains why they map to the same position. Future integration of additional lipid profiling may identify a causal factor that explains the relationship between these microbial and bile acid traits.

Slc10a2 is a candidate gene for Turicibacter sp. and plasma cholic acid

We focused our co-mapping analysis on chr 8 at ~ 5.5 Mbp, where *Turicibacter sp.* QTL and plasma cholic acid (CA) QTL overlap (Figure 3.3A-B). These traits were particularly interesting because both have been shown to be influenced by host genetics by previous studies.

For example, *Turicibacter* was identified as highly heritable in both mouse and human genetic studies (Benson et al., 2010; Goodrich et al., 2016; O'Connor et al., 2014; Org et al., 2015), whereas multiple studies have found differences in CA levels as a function of host genotype (Kreznar et al., 2017; Sehayek et al., 2006). Furthermore, CA levels are influenced by both host genetics and microbial metabolism since it is synthesized by host liver enzymes from cholesterol and subsequently modified by gut microbes in the intestine. Notably, these co-mapping traits also share the same allele effects pattern, where the A/J and WSB haplotypes have strong positive and negative associations, respectively (Figure 3.3C-D).

To assess whether trait patterns in the DO founder strains correspond to the observed allelic effects in the QTL mapping, we performed a separate characterization of the fecal microbiota composition and plasma bile acids in age-matched A/J and WSB animals fed the HF/HS diet. The founder strain allele patterns inferred from the QTL mapping closely resembled the observed levels of *Turicibacter sp.* (Figure 3.3E) and plasma CA in the founder strains (Figure 3.3F), where A/J animals had significantly higher levels of *Turicibacter sp.* and CA than WSB animals. However, *Turicibacter* levels in the founder strains do not completely mirror the estimated allele effects. This may be due to other genetic factors that also influence *Turicibacter* levels, as this taxa may be influenced by multiple host genes and levels of *Turicibacter* have previously been associated on chr 7 (Benson et al., 2010), 9 and 11 (Org et al., 2015). Furthermore, *Turicibacter* and plasma CA were positively correlated in the DO mice ($r = 0.43$, $p = 3.53e^{-10}$). This finding is consistent with a previous study that found positive correlations between *Turicibacter* and unconjugated cecal BAs (Theriot et al., 2016). Taken together, the overlap between the *Turicibacter sp.* QTL and plasma CA QTL, along with the similar allele effects pattern, which reflect the values observed

in the founder strains, provide strong evidence suggesting that these traits are related and they are responding to the common genetic driver.

We searched under the QTL for candidate genes via high-resolution association mapping on chr 8 and identified SNPs associated with both traits. Among these we identified SNPs upstream the candidate gene *Slc10a2*, which encodes for the apical sodium-bile transporter (Figure 3.3G). *Slc10a2* is responsible for ~95% of BA reabsorption in the distal ileum and plays a key role in BA homeostasis (Dawson et al., 2003). In humans, mutations in this gene are responsible for primary BA malabsorption, resulting in interruption of enterohepatic circulation of BAs and decreased plasma cholesterol levels (Oelkers et al., 1997). Likewise, *Slc10a2*$^{-/-}$ mice have reduced total BA pool size, increased fecal BA concentrations and reduced total plasma cholesterol in comparison to wild-type mice (Dawson et al., 2003). Additionally, a comparison between germ-free and conventionally-raised mice found that expression of *Slc10a2* is downregulated in presence of the gut microbiota, suggesting microbes may have influence the expression of the transporter (Sayin et al., 2013).

Genome analysis identified SNPs associated with levels of *Turicibacter sp.* and plasma CA at the QTL peak (Figure 3.3G). The SNPs with the strongest associations were attributed to the WSB and PWK haplotypes and fell on intergenic regions near *Slc10a2*. There is growing evidence that non-coding intergenic SNPs are often located in or closely linked to regulatory regions, suggesting they may influence host regulatory elements and alter gene expression (Chen and Tian, 2016; Maurano et al., 2012). To assess if candidate gene expression patterns in the DO founders corresponded to the estimated allelic effects in the QTL mapping, we quantified *Slc10a2* expression in distal ileum samples from AJ and WSB mice by quantitative reverse transcriptase PCR (qRT-PCR). AJ mice exhibited significantly higher expression of *Slc10a2* compared to WSB

mice (Figure 3.3H), which is consistent with estimated allele patterns for the overlapping *Turicibacter* and plasma CA QTLs on chr 8 (Figure 3.3A-B).

Mediation and causal inference testing identify a correlative relationship between Turicibacter and plasma cholic acid QTLs

The strong association between Turicibacter sp. and plasma CA levels may be due to a single shared locus (pleiotropy) or multiple closely linked loci (linkage disequilibrium). We examined whether these two traits were affected by a single locus of a pair of loci by likelihood ratio testing with a null hypothesis of pleiotropy (Boehm, 2018). Analysis of 1000 bootstrap samples resulted in a p-value of 0.531, which is consistent with the presence of a single pleiotropic locus that affects both traits (Supplemental Figure 3.6A).

We next sought to understand the causal relationships between the microbe and the BA. We asked whether the relationship between the microbe and BA was causal, reactive or independent. To establish the directionality of the relationship, we applied mediation analysis where we conditioned one trait on the other (MacKinnon et al., 2007). When we conditioned Turicibacter sp. on plasma CA (QTL → Bile Acid → Microbe), we observed a LOD drop of 3.2 (Figure 3.4A-B). Likewise, when we conditioned the plasma cholic acid on the microbe (QTL → Microbe → Bile Acid) there was a LOD drop of 3.32 (Figure 3.4C-D). The partial mediation seen in both models suggests a correlative relationship between the microbe and BA, where they exert an effect on one another and the directionality of the relationship is unclear.

Further causal model selection testing (Neto et al., 2013) found evidence that Turicibacter sp. is significantly correlated with plasma CA levels ($p < 0.05$), where the reactive SNPs occurred between ~5.39 – 7 Mbp. These reactive SNPs could be partially attributed to the WSB and PWK haplotypes (Supplemental Figure 3.6B). This reactive mediation model is consistent with the

pleiotropy analysis where the driver variant occurs between ~5.5 – 7 Mbp. There is also evidence of a causal relationship ($p < 0.005$) near ~9 Mbp where the microbe influences the abundance of the BA, which can also be attributed to the WSB and PWK alleles (Supplemental Figure 3.6C). However, this locus is in a gene desert, offering no immediate biological interpretation.

From this analysis, we can hypothesize this relationship can be explained by a pleiotropic model, where a single locus influences a microbial and BA trait, and the microbial trait is also reactive to changes in the BA trait (Figure 3.4H), with a second locus affecting the microbe, which in turn affects the BA. It is important to note that statistical inference only partially explains the relationship between the traits and there may be other hidden variables that may further explain the relationship. The complex relationship depicted by the causal inference testing is consistent with the complicated interplay between gut microbes and BAs in the intestine and their known ability to influence their other.

Bile acids inhibit Turicibacter sanguinis growth at physiologically relevant concentrations

Due to the strong correlative relationship between the QTL, we tested whether there was a direct interaction between bile acids and *Turicibacter*. *Turicibacter* inhabits the small intestine where concentrations of BAs are greater than in the cecum or colon (Li et al., 2017; Onishi et al., 2017). We screened the human isolate *Turicibacter sanguinis* for deconjugation and transformation activity *in vitro* by HPLC/MS-MS. We found that *T. sanguinis* deconjugated ~96-100% of taurocholic acid and glycochenodeoxycholic acid (Figure 3.5A) within 24 hours. It also transformed ~6 and 8 % of CA and CDCA to 7-dHCA and 7-ketolithocholic acid (7-KLCA), respectively (Figure 3.5B-C). The percent transformed did not increase after 24 hours (data not shown). Both of these transformations occur by action of the bacterial 7α-hydroxysteroid dehydrogenase.

Based on these results, we asked if conjugated and unconjugated bile acids differentially effect *T. sanguinis* growth. BA concentrations range from ~1-10 mM along the small intestine (Northfield and McColl, 1973) to ~0.2-1 mM in the cecum (Hamilton et al., 2007). Therefore, we grew *T. sanguinis* in the presence of either conjugated or unconjugated bile acids at physiologically relevant concentrations ranging from 0.1 – 1 mM. *T. sanguinis* growth decreased with increasing concentrations of conjugated bile acids and growth was completely inhibited at 1 mM (Figure 3.5D). Unconjugated bile acids affected growth rate at 1 mM (Figure 3.5E), Growth rate was significantly slower in the presence of 1 mM conjugated and unconjugated bile acids (Figure 3.5G). The slower growth rate at higher concentrations of BAs may affect intestinal abundance of *T. sanguinis*.

To compare *T. sanguinis* sensitivity to conjugated bile acids relative to other small intestine colonizers, we grew four taxa (*Bacteroides thetaiotamicron*, *Clostridium asparagaiforme*, *Lactobacillus reuteri* and *Escherichia coli* MS200-1) known to colonize this region of the intestine with or without 1 mM conjugated bile acids. Members of these genera are known to have bile salt hydrolase (BSH) activity to deconjugate bile acids (Ridlon et al., 2006). Unlike *T. sanguinis*, the addition of conjugated bile acids had little to no effect on the growth of these four gut microbes (Supplemental Figure 3.6). Based on the specific sensitivity of *T. sanguinis* to moderate-high concentrations of conjugated bile acids, *T. sanguinis* may use deconjugation as a survival mechanism to allow it to compete with other ileum colonizing organisms that can tolerate higher concentrations of conjugated bile acids. Consistent with these findings, *Turicibacter* abundance was negatively correlated with cecal TCA levels in the DO mice ($r = -0.262$, $p = 0.0035$), supporting the notion this it is sensitive to elevated conjugated bile acid levels.

Taken together, these data indicate that *T. sanguinis* is sensitive to higher concentrations of conjugated and to a lesser extent unconjugated BA compared to other small intestine colonizers and that it may use deconjugation to decrease BA toxicity. These reciprocal effects between the BA and the bacterium provide biological evidence for the correlative relationship shown by the causal model testing (Figure 3.4H). In summary, using a genetic approach, we identified and provide validation of a relationship between a genetic locus containing the BA transporter *Slc10a2*, and levels of *Turicibacter* and plasma cholic acid. Based on our findings, we hypothesize that the identified locus regulates expression of *Slc10a2*, altering active BA reabsorption in the ileum, leading to increased intestinal BA concentrations and alterations in the intestinal BA environment. Consequently, the resulting change in environmental BA concentration and/or composition provides an unfavorable habitat for *Turicibacter*. The loss of *Turicibacter's* deconjugation activity leads to a decrease in circulating free plasma cholic acid levels.

CONCLUSIONS

In this study, we performed the first known genetic mapping integration of gut microbiome and BA. We used genetics as an anchor to identify microbe-metabolite interactions and hypothesize novel host genes involved in shaping gut microbiota and bile acid profiles. Using DO mice, we identified multiple QTL for gut microbes and bile acids spanning the host genome. These included loci that associated with individual microbial and BA traits, as well as loci with potential pleiotropic effects, where a single genetic region influenced both the abundance of a gut microbe and levels of a BA.

While several studies suggest that host genetic variation has a minor impact on microbiota composition, there are overlapping findings among different studies in both human and mouse populations that indicate that specific bacterial taxa are influenced by host genetics. Our results in the DO population corroborate several of these key findings. For example, we observed the strongest associations to the host genome with members of the Firmicutes phylum, including unknown members of the Clostridiales order, the Lachnospiraceae, Christensenellacae and S24-7 families, the *Turicibacter* and *Coprococcus* genera, as well as the species *Akkermansia muciniphila* and *Ruminococcus gnavus* (Table 4), all of which have consistently been identified in multiple studies as either highly heritable or associating to positions on the host genome (Benson et al., 2010; Davenport et al., 2015; Leamy et al., 2014; McKnite et al., 2012; Org et al., 2015; Wang et al., 2016). Furthermore, our study replicated correlations between taxa including the Peptostreptococcaeae and Turicibacteriaceae families (Goodrich et al., 2016). Previous studies in humans and rats also identified a significant correlation between these taxa (Goodrich et al., 2016; Li et al., 2017), and both taxa are consistently identified as heritable in humans and mice (Benson et al., 2010; Goodrich et al., 2016; O'Connor et al., 2014). This correlation is particularly notable

since we found that these two organisms have complementary BA metabolism capabilities, where the Turicibacteriaceae family performs the deconjugation necessary for members of Peptostreptococcaceae to transformations bile acids. BAs must be deconjugated prior to epimerization, so Peptostreptococcaeae may associate with Turicibacteriaceae in order to utilize this metabolic capability. Thus, their co-occurrence may provide a fitness advantage for small intestine colonization. These findings may give insight into microbial dynamics that govern BA profiles and warrant further investigation. Given the high degree of variability in the gut microbiome across subjects and host organisms, these instances of congruence between studies argues that there are specific taxa responsive to host genotype that may warrant follow-up investigation. Our work with the DO population provides an approach to validate these associations.

The work presented here plus data from previous studies suggest that BA pool alterations driven by *Slc10a2* activity elicit an impact on gut microbiota community structure and influence the ability of *Turicibacter* to colonize and persist in the intestine. Several studies have noted concomitant changes in microbiota composition and *Slc10a2* mRNA levels (Janssen et al., 2017; Miyata et al., 2011; Out et al., 2015). Furthermore, Enterobacteraceae and enteropathogenic E. coli can modulate intestinal bile acid transport via *Slc10a2* (Annaba et al., 2012; Miyata et al., 2011). We found that *T. sanguinis* efficiently deconjugates primary BAs, which may explain the correlative relationship between the abundance of this taxa and levels of free CA in the plasma. Although this microbe deconjugates primary BAs, we also found that it is also sensitive to elevated concentrations of both conjugated and unconjugated BAs. While our data shows that higher concentrations of conjugated BAs inhibit *Turicibacter* growth, it is still unknown what intestinal environment is more favorable for *Turicibacter* growth and colonization. Future experiments are

needed to examine how a decrease in *Slc10a2* expression changes intestinal BA profiles and the consequences on *Turicibacter* colonization.

We also identified multiple host-microbe-metabolite interactions that can be validated with additional mechanistic studies. More broadly, our work demonstrates that we can identify novel interactions between microbial and metabolite traits using host genetics and provides new testable hypotheses to further dissect factors that shape gut microbiota composition. This work may provide a critical framework for future host-microbe interaction studies.

EXPERIMENTAL PROCEDURES

Animals and sample collection. Animal care and study protocols were approved by the University of Wisconsin-Madison Animal Care and Use Committee. DO mice were obtained from the Jackson Laboratories (Bar Harbor, ME, USA) at ~4 weeks of age and maintained in the Department of Biochemistry vivarium at the University of Wisconsin-Madison. Mice were housed on a 12-hour light:dark cycle under temperature- and humidity-controlled conditions. Waves of DO mice from generations 18, 19 and 21 were obtained three times per year until 500 DO mice were surveyed. Each wave was composed of equal numbers of male and female mice. All mice were fed a high-fat high-sucrose diet (TD.08811, Envigo Teklad, 44.6% kcal fat, 34% carbohydrate, and 17.3% protein) ad libitum upon arrival to the facility. Mice were kept in the same vivarium room and were individually housed to monitor food intake and prevent coprophagy between animals. DO mice were sacrificed at 22-25 weeks of age.

The eight DO founder strains (C57BL/6J, A/J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HILtJ, PWK/PhJ, WSB/EiJ and CAST/EiJ) were obtained from the Jackson Laboratories. Mice were bred at the University of Wisconsin-Madison Biochemistry Department. Mice were housed by strain and sex (2-5 mice/cage), with the exception of CAST that required individual housing. Mice were housed under the same environmental conditions as the DO animals. Like the DO mice, the eight founder strains were maintained on the HF/HS diet and were sacrificed at 22 weeks of age, except for NZO males that were sacrificed at 14 weeks, due to high mortality attributable to severe disease.

For both DO and founder mice, fecal samples for 16S rRNA sequencing were collected immediately before sacrifice after a 4 hour fast. Cecal contents, plasma, and additional tissues were harvested promptly after sacrifice and all samples were immediately flash frozen in liquid nitrogen and stored at -80°C until further processing.

DNA extraction. DNA was isolated from feces using a bead-beating protocol {Turnbaugh:2009ei}{Turnbaugh:2009ei}. Mouse feces (~1 pellet per animal) 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. Contaminants were removed using QIAquick 96-well PCR Purification Kit (Qiagen, Germantown, MD, USA). Isolated DNA was eluted in 5 mM Tris/HCL (pH 8.5) and was stored at -80°C until further use.

16S rRNA Sequencing. PCR was performed using universal primers flanking the variable 4 (V4) region of the bacterial 16S rRNA gene (Kozich et al., 2013). Genomic DNA samples were amplified in duplicate. Each reaction contained 10-30 ng genomic DNA, 10 µM each primer, 12.5 µl 2x HiFi HotStart ReadyMix (KAPA Biosystems, Wilmington, MA, USA), and water to a final reaction volume of 25 µl. 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 QIAquick 96-well PCR Purification Kit (Qiagen, Germantown, MD, USA) and quantified using Qubit dsDNA HS Assay kit (Invitrogen, Oregon, USA). Samples were equimolar pooled and sequenced by the University of Wisconsin – Madison Biotechnology Center with the MiSeq 2x250 v2 kit (Illumina, San Diego, CA, USA) using custom sequencing primers.

16S analysis. Demultiplexed paired end fastq files generated by CASAVA (Illumina) and a mapping file were used as input files. Sequences were processed, quality filtered and analyzed with QIIME2 (version 2018.4) (https://qiime2.org), a plugin-based microbiome analysis platform (Caporaso et al., 2010). DADA2 (Callahan et al., 2016) was used to denoise sequencing reads with the q2-dada2 plugin for quality filtering and identification of de novo exact sequence variants (ESVs) (i.e. 100% exact sequence match). This resulted in 20,831,573 total sequences with an average of 52,078 sequences per sample for the DO mice, and 2,128,796 total sequences with an average of 34,335.4 sequences per sample for the eight DO founder strains. Sequence variants were aligned with mafft (Katoh and Standley, 2013) with the q2-alignment plugin. The q2-phylogeny plugin was used for phylogenetic reconstruction via FastTree (Price et al., 2010). Taxonomic classification was assigned using classify-sklearn (Bokulich et al., 2018) against the Greengenes 13_8 99% reference sequences (McDonald et al., 2012). Alpha- and beta-diversity (weighted and unweighted UniFrac (Lozupone and Knight, 2005) analyses were performed using q2-diversity plugin at a rarefaction depth of 10000 sequences per sample. For the DO mice, one sample (DO071) was removed from subsequent analysis because it did not reach this sequencing depth. For analysis of the eight DO founder strains, one sample (NOD5) was removed because it did not reach this sequencing depth. Subsequent processing and analysis were performed in R (v.3.5.1), and data generated in QIIME2 was imported into R using Phyloseq (McMurdie and Holmes, 2013). Sequencing data was normalized by cumulative sum scaling (CSS) using MetagenomeSeq (Paulson et al., 2013). Summaries of the taxonomic distributions were generated by collapsing normalized ESV counts into higher taxonomic levels (genus to phylum) by phylogeny. We defined a core measurable microbiota (CMM) (Benson et al., 2010) to include only

microbial traits present in 20% of individuals in the QTL mapping. In total, 86 ESVs and 42 collapsed microbial taxonomies comprised the CMM.

Sample preparation for plasma bile acid analysis. 40 µL of DO plasma collected at sacrifice (30 µL used for founder strains) were aliquoted into a tube with 10 µL SPLASH Lipidomix internal standard mixture (Avanti Polar Lipids, Inc.). Protein was precipitated by addition of 215 µL MeOH. After the mixture was vortexed for 10 s, 750 µL methyl tert-butyl ether (MTBE) were added as extraction solvent and the mixture was vortexed for 10 s and mixed on an orbital shaker for 6 min. Phase separation was induced by adding 187.5 µL of water followed by 20 s of vortexing. All steps were performed at 4 °C on ice. Finally, the mixture was centrifuged for 4 min at 14,000 x g at 4 °C and stored at -80 °C. For targeted bile acids analysis, samples were thawed on ice. 400 µL of ethanol were added to further precipitate protein, as well as 15 µL of isotope-labeled internal standard mix (12.5 µM d4-TαMCA, 10 µM d4-CDCA). The samples were vortexed for 20 s and centrifuged for 4 min at 14,000 g at 4 °C after which the supernatant (ca. 1000 µL) was taken out and dried down. Dried supernatants were resuspended in 60 µL mobile phase (50 %B), vortexed for 20 s, centrifuged for 4 min at 14,000 g and then 50 µL were transferred to vials with glass inserts for MS analysis.

Sample preparation for cecal bile acid analysis. 30 (± 7.5) mg cecal contents along with 10 µL SPLASH Lipidomix internal standard mixture were aliquoted into a tube with a metal bead and 270 µL MeOH were added for protein precipitation. To each tube, 900 µL MTBE and 225 µL of water were added as extraction solvents. All steps were performed at 4 °C on ice. The mixture was homogenized by bead beating for 8 min at 25 Hz. Finally, the mixture was centrifuged for 4-8 min

at 11,000 x g at 4 °C. Subsequent processing for the DO mice and eight DO founder strains differed due to other analysis performed on the samples that is not presented in this paper. For DO samples, 100 µL of the aqueous and 720 µL of organic layer were combined and stored at -80 °C. For analysis, these were thawed on ice and 400 µL of ethanol were added to further precipitate protein, as well as 15 µL of isotope-labeled internal standard mix (12.5 µM d4-TαMCA, 10 µM d4-CDCA). The samples were vortexed for 20 s and centrifuged for 4 min at 14,000 g at 4 °C after which the supernatant (ca. 1000 µL) was taken out and dried down. Dried supernatants were resuspended in 100 µL mobile phase (50 %B), vortexed for 20 s, centrifuged for 8 min at 14,000 g and then 50 µL were transferred to vials with glass inserts for MS analysis. For the eight DO founder strains, the mixture was dried down including all solid parts and stored dried at -80 °C. For targeted bile acid analysis, these dried down samples were then thawed on ice and reconstituted in 270 µL of methanol, 900 µL of MTBE, and 225 µL of water. 400 µL of ethanol were added to further precipitate protein, as well as 15 µL of isotope-labeled internal standard mix (12.5 µM d4-TαMCA, 10 µM d4-CDCA). The mixture was bead beat for 8 min at 25 Hz and centrifuged at 14,000 g for 8 minutes after which the supernatant (ca. 1500 µL) was taken out and dried down. Dried supernatants were resuspended in 100 µL mobile phase (50 %B), vortexed for 20 s, centrifuged for 4 min at 14,000 g and then 90 µL were transferred to vials with glass inserts for MS analysis.

Measurement and analysis of mouse bile acids. LC-MS analysis was performed in randomized order using an Acquity CSH C18 column held at 50 °C (100 mm × 2.1 mm × 1.7 µm particle size; Waters) connected to an Ultimate 3000 Binary Pump (400 µL/min flow rate; Thermo Scientific). Mobile phase A consisted of 10 mM ammonium acetate containing 1 mL/L ammonium hydroxide. Mobile phase B consisted of MeOH with the same additives. Mobile phase B was initially held at

50% for 1.5 min and then increased to 70% over 13.5 min. Mobile phase B was further increased to 99% over 0.5 min and held for 2.5 min. The column was reequilibrated for 5.5 min before the next injection. Twenty microliters of plasma sample or ten microliters of cecum sample were injected by an Ultimate 3000 autosampler (Thermo Scientific). The LC system was coupled to a TSQ Quantiva Triple Quadrupole mass spectrometer (Thermo Scientific) by a heated ESI source kept at 325°C (Thermo Scientific). The inlet capillary was kept at 350 °C, sheath gas was set to 15 units, auxiliary gas to 10 units, and the negative spray voltage was set to 2,500 V. For targeted analysis the MS was operated in negative single reaction monitoring (SRM) mode acquiring scheduled, targeted scans to quantify selected bile acid transitions, with two transitions for each species' precursor and 3 min retention time windows. Collision energies were optimized for each species and ranging from 20-55 V. Due to insufficient fragmentation for unconjugated bile acids, the precursor was monitored as one transition with a CE of 20 V. MS acquisition parameters were 0.7 FWHM resolution for Q1 and Q3, 1 s cycle time, 1.5 mTorr CID gas and 3 s Chrom filter. In total, 27 bile acids, including 14 unconjugated, 9 tauro- and 4 glycine-conjugated species, were measured. The resulting bile acid data were processed using Skyline 3.6.0.10493 (University of Washington). For each species, one transition was picked for quantitation, while the other was used for retention time confirmation. Normalization of the quantitative data was performed to the internal standard d4-CDCA as indicated in **Equation 1**.

Equation 1: (Peak Area / d4-CDCA Peak Area) · Average of d4-CDCA Peak Area

Genotyping. Genotyping was performed on tail biopsies as previously described (Svenson et al., 2012) using the Mouse Universal Genotyping Array (GigaMUGA) [143,259 markers] (Morgan et al., 2015) at Neogen (Lincoln, NE). Genotypes were converted to founder strain-haplotype

reconstructions using a hidden Markov model (HMM) implemented in the R/DOQTL package (Gatti et al., 2014). We interpolated the GigaMUGA markers onto an evenly spaced grid with 0.02-cM spacing and added markers to fill in regions with sparse physical representation, which resulted in 69,005 pseudomarkers.

QTL mapping. We performed QTL mapping using the R package R/qtl2 (Broman, 2018). QTL mapping was done through a regression of the phenotype on the founder haplotype probabilities estimated with a HMM designed for multi-parental populations. Genome scans were performed for each phenotype with sex, cohort (wave), and days on diet were included as additive covariates for the trait mapping. Genetic similarity between mice was accounted for using a kinship matrix based on the leave-one-chromosome-out (LOCO) methods. For microbial QTL mapping, normalized gut microbiota abundance data was nqrank transformed. For bile acid QTL mapping, normalized plasma and cecal bile acid levels were log2 transformed. The mapping statistic reported is log of the odds ratio (LOD). The significance thresholds were determined by performing 1000 permutations of genome-wide scans by shuffling phenotypic data in relation to individual genotypes. QTL reaching a LOD score > 5.5 were considered of interest and the QTL support interval was defined using the 95% Bayesian credible interval.

Mediation and pleiotropy analysis. To assess whether two co-mapping traits were caused by a pleiotropic locus, we used a likelihood ratio test implemented with the open source R package R/qtl2pleio (Boehm, 2018). Here, we compared the alternative hypothesis of two distinct loci with the null hypothesis of pleiotropy for two traits that map to the same genetic region. Parametric bootstrapping was used to determine statistical significance ($p < 0.05$). Mediation analysis was

applied to identify whether a microbe or bile acid were likely to be a causal mediator of the QTL as presented in Li et al. (Li et al., 2010). This analysis was adapted from a general approach previously described to differentiate target from mediator variables (Baron and Kenny, 1986). The effect of a mediator on a target was evaluated by performing an allele scan or SNP scan using the target adjusted by mediator. Only individuals with both values for both traits were considered for mediation analysis. Traits with a LOD drop >2 after controlling for the mediator were considered for further causality testing. To statistically assess causality between microbial and bile acid trait sets (causal, reactive, independent, undecided), a causal model selection test (Neto et al., 2013) was applied using the R packages R/intermediate and R/qtl2. Causal model selection tests were evaluated on both alleles and SNPs in peak region.

RNA extraction. Total RNA was extracted from flash-frozen distal ileum tissues by TRIzol extraction and further cleaned using the RNeasy Mini Kit (Qiagen, Germantown, MD, USA). DNA was removed by on-column DNase digestion (Qiagen). Purified RNA was quantified using a Nanodrop 2000 spectrophotometer.

Quantitative Real-Time PCR. SuperScript II Reverse Transcriptase with oligo(dT) primer (all from Invitrogen, Carlsbad, CA, USA) was used to synthesize 20 µl cDNA templates from 1 µg purified RNA. cDNA was diluted 2X before use and qRT-PCR reactions were prepared in a 10 µl volume using SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, USA) and 400 nM specific primers targeting the gene of interest (SLC10A2-F [5'-TGGGTTTCTTCCTGGCTAGACT-3']; SLC10A2-R [5'- TGTTCTGCATTCCAGTTTCCAA-3'] (Rao et al., 2016)). All reactions were performed in triplicate. Reactions were run on a CFX96

Real-Time PCR System (Bio-Rad, Hercules, CA, USA). The 2-ΔΔCt method (Livak and Schmittgen, 2001) was used to calculate relative changes in gene expression and all results were normalized to GAPDH.

Bacterial culturing. Bacterial strains were obtained from DSMZ and ATCC. All strains were cultured at 37°C under anaerobic conditions using an anaerobic chamber (Coy Laboratory Products) with a gas mix of 5% hydrogen, 20% carbon dioxide and 75% nitrogen. Strains were grown in rich medium (Supplemental Table 2) that was filter sterilized and stored in the anaerobic chamber at least 24 hours prior to use. L. reuteri was grown in medium supplemented with 20 mM glucose. For all in vitro assays, cultures used for inoculation were grown overnight at 37°C in 10 mL 14b medium in anaerobic Hungate tubes. Stock solutions of conjugated bile acids (TCA, GCDCA) and unconjugated bile acids (CA, CDCA, DCA) were prepared to a final concentration of 100 mM and used for all in vitro assays. All bile acids used were soluble in methanol.

Microbial bile acid metabolism screen. Stock solutions of conjugated and unconjugated bile acids (100 mM) were added to 3 ml 14b medium to obtain a final concentration of 100 μM total bile acid. Tubes were inoculated with a T. sanguinis cultured overnight, then incubated in the anaerobic chamber at 37°C for 48 hours. At the 24- and 48-hour timepoints, 1 mL of each culture was removed and the supernant was collected after brief centrifugation. Each culture supernant was diluted 10x in initial running solvent (30:70 MeOH:10 mM ammonium acetate). Samples were spun at max speed for 3 minutes to remove suspended particles prior to loading on the uHPLC. Samples were analyzed using a uHPLC coupled with a high-resolution mass spectrometer.

Microbial bile acid screen uHPLC-MS/MS parameters. 10 µL aliquots of diluted supernatant samples were analyzed using a uHPLC-MS/MS system consisting of a Vanquish uHPLC coupled by electrospray ionization (ESI) (negative mode) to a hybrid quadrupole-high-resolution mass spectrometer (Q Exactive Orbitrap; Thermo Scientific). Liquid chromatography separation was achieved on an Acquity UPLC BEH C18 column (2.1-by 100-mm column, 1.7-µm particle size) heated to 50˚C. Solvent A was 10 mM Ammonium acetate, pH 6; solvent B was 100% methanol. The total run time was 31.5 minutes with the following gradient: 0 min, 30% B; 0.5 min, 30% B; 24 min, 100% B; 29 min, 100% B; 29 min, 30% B; 31.5 min, 30% B. Bile acid peaks were identified using the Metabolomics Analysis and Visualization Engine (MAVEN) (Clasquin et al., 2012).

Growth curves. Bacterial growth rate was measured in medium 14b supplemented with either 100 µM, 300 µM, 1 mM bile acids or methanol control. Medium was dispensed inside an anerobic chamber into Hungate tubes. Tubes containing 10 mL of medium were inoculated with 30 µL of an overnight culture and incubated at 37°C for 24 hours. T. sanguinis was grown with shaking to disrupt the formation of flocculent colonies. Growth was monitored as the increase in absorbance at 600 nm in a Spectronic 20D+ spectrophotometer (Thermo Scientific, Waltham, MA, USA). Growth rate was determined as $\mu = \ln(X/Xo)/T$, where X is the OD600 value during the linear portion of growth and T is time in hours. Values given are the mean µ values from two independent cultures done in triplicate.

Statistical analysis. All statistical analyses were performed in R (v.3.5.1) (Team). Unless otherwise indicated in the figure legends, differences between groups were evaluated using

unpaired two-tailed Welch's t-test. For multiple comparisons, Krustkal-Wallis test was used if ANOVA conditions were not met, followed by Mann-Whitney/Wilcoxon rank-sum for multiple comparisons and adjusted for multiple testing using the Benjamini-Hochberg FDR procedure. The correlation between the abundance of microbial taxa was performed using Spearman's correlation in the "Hmisc" (v.4.1-1) R package (Harrell Jr and others, 2018). The p-values were adjusted using the Benjamini and Hochberg method, and correlation coefficients were visualized using the "pheatmap" (v.1.0.10) (Kolde, 2018). Multiple groups were compared by Kruskal-Wallis test and adjusted for multiple testing using the Benjamini-Hochberg FDR procedure. Significance was determined as p-value < 0.05. To assess magnitude of variability of the CMMs, summary statistics were calculated on each CMM (taxa and ESVs). Non-parametric-based PERMANOVA statistical test (McArdle and Anderson, 2001) with 999 Monte Carlo permutations was used to compare microbiota compositions among groups using the Vegan R package (Oksanen et al., 2018).

ACCESSION NUMBERS

The data reported in this paper are accessible in the NCBI Short Read Archive (SRP) under accession ID PRJNA492330.

CONTRIBUTIONS

Project was conceived by Federico Rey, Alan Attie, Mark Keller, and Julia Kemis. Diversity Outbred (DO) animals were purchased from Jackson Laboratories. Kathryn Schueler, Mary Rabaglia, and Donald Stapleton maintained the DO animals, performed clinical trait measurements, and collected tissues and fecal samples. Vanessa Linke extracted and analyzed plasma and cecal bile acids. Karl Broman, Brian Yandell, Dan Gatti and Gary Churchill developed methods for QTL analysis and calculated the genotype probabilities and kinship matrix. Fred Boehm conducted pleiotropy analysis. Brian Yandell performed causal inference testing. Julia Kemis sequenced and analyzed fecal microbiota composition data, performed QTL mapping of microbial and bile acid traits, designed and executed *in vitro* experiments with *Turicibacter*, analyzed and interpreted data. The manuscript was written by Julia Kemis and Federico Rey.

ACKNOWLEDGEMENTS

The authors thank the University of Wisconsin Biotechnology Center DNA Sequencing Facility for providing sequencing and support services, and the University of Wisconsin Center for High Throughput Computing (CHTC) in the Department of Computer Sciences for providing computational resources, support, and assistance.

This work was supported in part by grants from the National Institutes of Health (NIH) to F.E.R. (DK-1018259-01) and A.D.A. (DK-101573 and DK-102948). L.L.T. is supported by the NLM Computation and Informatics in Biology and Medicine Postdoctoral Fellowship (5T15LM007359). J.H.K. is supported by the NIH National Institute of Allergy and Infectious Diseases grant T32AI55397.

REFERENCES

de Aguiar Vallim, T.Q., Tarling, E.J., and Edwards, P.A. (2013). Pleiotropic roles of bile acids in metabolism. Cell Metab. *17*, 657–669.

Annaba, F., Sarwar, Z., Gill, R.K., Ghosh, A., Saksena, S., Borthakur, A., Hecht, G.A., Dudeja, P.K., and Alrefai, W.A. (2012). Enteropathogenic Escherichia coli inhibits ileal sodium-dependent bile acid transporter ASBT. Am. J. Physiol. Gastrointest. Liver Physiol. *302*, G1216-22.

Arakawa, R., Bagashev, A., Song, L., Maurer, K., and Sullivan, K.E. (2010). Characterization of LRRFIP1. Biochem. Cell Biol. *88*, 899–906.

van der Ark, K.C.H., Nugroho, A.D.W., Berton-Carabin, C., Wang, C., Belzer, C., de Vos, W.M., and Schroen, K. (2017). Encapsulation of the therapeutic microbe Akkermansia muciniphila in a double emulsion enhances survival in simulated gastric conditions. Food Res. Int. *102*, 372–379.

Baron, R.M., and Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J. Pers. Soc. Psychol. *51*, 1173–1182.

Begley, M., Gahan, C.G.M., and Hill, C. (2005). The interaction between bacteria and bile. FEMS Microbiol. Rev. *29*, 625–651.

Belheouane, M., Gupta, Y., Künzel, S., Ibrahim, S., and Baines, J.F. (2017). Improved detection of gene-microbe interactions in the mouse skin microbiota using high-resolution QTL mapping of 16S rRNA transcripts. Microbiome *5*, 59.

Belzer, C., and de Vos, W.M. (2012). Microbes inside--from diversity to function: the case of Akkermansia. ISME J. *6*, 1449–1458.

Benson, A.K., Kelly, S.A., Legge, R., Ma, F., Low, S.J., Kim, J., Zhang, M., Oh, P.L., Nehrenberg, D., Hua, K., et al. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. Proc. Natl. Acad. Sci. U. S. A. *107*, 18933–18938.

Birkenfeld, A.L., Lee, H.-Y., Guebre-Egziabher, F., Alves, T.C., Jurczak, M.J., Jornayvaz, F.R., Zhang, D., Hsiao, J.J., Martin-Montalvo, A., Fischer-Rosinsky, A., et al. (2011). Deletion of the mammalian INDY homolog mimics aspects of dietary restriction and protects against adiposity and insulin resistance in mice. Cell Metab. *14*, 184–195.

Blekhman, R., Goodrich, J.K., Huang, K., Sun, Q., Bukowski, R., Bell, J.T., Spector, T.D., Keinan, A., Ley, R.E., Gevers, D., et al. (2015). Host genetic variation impacts microbiome composition across human body sites. Genome Biol. *16*, 191.

Boehm, F. (2018). qtl2pleio: Hypothesis test of close linkage vs pleiotropy in multiparental populations.

Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., Huttley, G.A., and Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome *6*, 90.

Bonder, M.J., Kurilshikov, A., Tigchelaar, E.F., Mujagic, Z., Imhann, F., Vila, A.V., Deelen, P., Vatanen, T., Schirmer, M., Smeekens, S.P., et al. (2016). The effect of host genetics on the gut microbiome. Nat. Genet. *48*, 1407–1412.

Van den Bossche, L., Hindryckx, P., Devisscher, L., Devriese, S., Van Welden, S., Holvoet, T., Vilchez-Vargas, R., Vital, M., Pieper, D.H., Vanden Bussche, J., et al. (2017). Ursodeoxycholic Acid and Its Taurine- or Glycine-Conjugated Species Reduce Colitogenic Dysbiosis and Equally Suppress Experimental Colitis in Mice. Appl. Environ. Microbiol. *83*, e02766-16.

Broman, K.W. (2018). qtl2: Quantitative Trait Locus Mapping in Experimental Crosses.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods *13*, 581–583.

Cani, P.D., and de Vos, W.M. (2017). Next-Generation Beneficial Microbes: The Case of Akkermansia muciniphila. Front. Microbiol. *8*, 1765.

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., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods *7*, 335–336.

Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.-M., Kennedy, S., et al. (2013). Richness of human gut microbiome correlates with metabolic markers. Nature *500*, 541–546.

Chen, J., and Tian, W. (2016). Explaining the disease phenotype of intergenic SNP through predicted long range regulation. Nucleic Acids Res. *44*, 8641–8654.

Chen, C., Sargent, C., Quilter, C., Yang, Z., Ren, J., Affara, N., Brenig, B., and Huang, L. (2010). Cloning, mapping and molecular characterization of porcine progesterone receptor membrane component 2 (PGRMC2) gene. Genet. Mol. Biol. *33*, 471–474.

Churchill, G.A., Gatti, D.M., Munger, S.C., and Svenson, K.L. (2012). The Diversity Outbred mouse population. Mamm. Genome *23*, 713–718.

Civelek, M., and Lusis, A.J. (2014). Systems genetics approaches to understand complex traits. Nat. Rev. Genet. *15*, 34–48.

Clasquin, M.F., Melamud, E., and Rabinowitz, J.D. (2012). LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. Curr. Protoc. Bioinforma. *Chapter 14*, Unit14.11-14.11.23.

Clemente, J.C., Ursell, L.K., Parfrey, L.W., and Knight, R. (2012). The impact of the gut microbiota on human health: an integrative view. Cell *148*, 1258–1270.

Clemente, J.C., Manasson, J., and Scher, J.U. (2018). The role of the gut microbiome in systemic inflammatory disease. BMJ *360*, j5145.

Davenport, E.R., Cusanovich, D.A., Michelini, K., Barreiro, L.B., Ober, C., and Gilad, Y. (2015). Genome-Wide Association Studies of the Human Gut Microbiota. PLoS One *10*, e0140301.

Dawson, P.A., Haywood, J., Craddock, A.L., Wilson, M., Tietjen, M., Kluckman, K., Maeda, N., and Parks, J.S. (2003). Targeted deletion of the ileal bile acid transporter eliminates enterohepatic cycling of bile acids in mice. J. Biol. Chem. *278*, 33920–33927.

Derrien, M., Vaughan, E.E., Plugge, C.M., and de Vos, W.M. (2004). Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int. J. Syst. Evol. Microbiol. *54*, 1469–1476.

Derrien, M., Van Baarlen, P., Hooiveld, G., Norin, E., Müller, M., and de Vos, W.M. (2011). Modulation of Mucosal Immune Response, Tolerance, and Proliferation in Mice Colonized by the Mucin-Degrader Akkermansia muciniphila. Front. Microbiol. *2*, 166.

Edenharder, R., Pfützner, A., and Hammann, R. (1989). Characterization of NAD-dependent 3 alpha- and 3 beta-hydroxysteroid dehydrogenase and of NADP-dependent 7 beta-hydroxysteroid dehydrogenase from Peptostreptococcus productus. Biochim. Biophys. Acta *1004*, 230–238.

Everard, A., Belzer, C., Geurts, L., Ouwerkerk, J.P., Druart, C., Bindels, L.B., Guiot, Y., Derrien, M., Muccioli, G.G., Delzenne, N.M., et al. (2013). Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. Proc. Natl. Acad. Sci. U. S. A. *110*, 9066–9071.

Fu, J., Bonder, M.J., Cenit, M.C.M.C., Tigchelaar, E.F., Maatman, A., Dekens, J.A.M.M., Brandsma, E., Marczynska, J., Imhann, F., Weersma, R.K., et al. (2015). The Gut Microbiome Contributes to a Substantial Proportion of the Variation in Blood Lipids. Circ. Res. *117*, 817–824.

Gatti, D.M., Svenson, K.L., Shabalin, A., Wu, L.-Y., Valdar, W., Simecek, P., Goodwin, N., Cheng, R., Pomp, D., Palmer, A., et al. (2014). Quantitative trait locus mapping methods for diversity outbred mice. G3 (Bethesda). *4*, 1623–1633.

Gerdes, D., Wehling, M., Leube, B., and Falkenstein, E. (1998). Cloning and tissue expression of two putative steroid membrane receptors. Biol. Chem. *379*, 907–911.

Goodrich, J.K., Davenport, E.R., Beaumont, M., Jackson, M.A., Knight, R., Ober, C., Spector, T.D., Bell, J.T., Clark, A.G., and Ley, R.E. (2016). Genetic Determinants of the Gut Microbiome in UK Twins. Cell Host Microbe *19*, 731–743.

Hall, A.B., Tolonen, A.C., and Xavier, R.J. (2017). Human genetic variation and the gut microbiome in disease. Nat. Rev. Genet. *18*, 690–699.

Hamilton, J.P., Xie, G., Raufman, J.-P., Hogan, S., Griffin, T.L., Packard, C.A., Chatfield, D.A., Hagey, L.R., Steinbach, J.H., and Hofmann, A.F. (2007). Human cecal bile acids: concentration and spectrum. Am. J. Physiol. Gastrointest. Liver Physiol. *293*, G256-63.

Harrell Jr, F.E., and others, with contributions from C.D. and many (2018). Hmisc: Harrell Miscellaneous.

Herrema, H., IJzerman, R.G., and Nieuwdorp, M. (2017). Emerging role of intestinal microbiota and microbial metabolites in metabolic control. Diabetologia *60*, 613–617.

Hughes, A.L., Powell, D.W., Bard, M., Eckstein, J., Barbuch, R., Link, A.J., and Espenshade, P.J. (2007). Dap1/PGRMC1 binds and regulates cytochrome P450 enzymes. Cell Metab. *5*, 143–149.

Inoue, K., Zhuang, L., Maddox, D.M., Smith, S.B., and Ganapathy, V. (2002). Structure, function, and expression pattern of a novel sodium-coupled citrate transporter (NaCT) cloned from mammalian brain. J. Biol. Chem. *277*, 39469–39476.

Ishimori, N., Li, R., Kelmenson, P.M., Korstanje, R., Walsh, K.A., Churchill, G.A., Forsman-Semb, K., and Paigen, B. (2004). Quantitative trait loci analysis for plasma HDL-cholesterol concentrations and atherosclerosis susceptibility between inbred mouse strains C57BL/6J and 129S1/SvImJ. Arterioscler. Thromb. Vasc. Biol. *24*, 161–166.

Islam, K.B.M.S., Fukiya, S., Hagio, M., Fujii, N., Ishizuka, S., Ooka, T., Ogura, Y., Hayashi, T., and Yokota, A. (2011). Bile acid is a host factor that regulates the composition of the cecal microbiota in rats. Gastroenterology *141*, 1773–1781.

Janssen, A.W.F., Dijk, W., Boekhorst, J., Kuipers, F., Groen, A.K., Lukovac, S., Hooiveld, G.J.E.J., and Kersten, S. (2017). ANGPTL4 promotes bile acid absorption during taurocholic acid supplementation via a mechanism dependent on the gut microbiota. Biochim. Biophys. Acta *1862*, 1056–1067.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Keller, M.P., Gatti, D.M., Schueler, K.L., Rabaglia, M.E., Stapleton, D.S., Simecek, P., Vincent, M., Allen, S., Broman, A.T., Bacher, R., et al. (2018). Genetic Drivers of Pancreatic Islet Function. Genetics *209*, 335–356.

Khan, M.T., Nieuwdorp, M., and Bäckhed, F. (2014). Microbial modulation of insulin sensitivity. Cell Metab. *20*, 753–760.

Klinkspoor, J.H., Mok, K.S., Van Klinken, B.J.W., Tytgat, G.N.J., Lee, S.P., and Groen, A.K. (1999). Mucin secretion by the human colon cell line LS174T is regulated by bile salts. Glycobiology *9*, 13–19.

Kolde, R. (2018). pheatmap: Pretty Heatmaps.

Kovacs, A., Ben-Jacob, N., Tayem, H., Halperin, E., Iraqi, F.A., and Gophna, U. (2011). Genotype is a stronger determinant than sex of the mouse gut microbiota. Microb. Ecol. *61*, 423–428.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and 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. Appl. Environ. Microbiol. *79*, 5112–5120.

Krautkramer, K.A., Kreznar, J.H., Romano, K.A., Vivas, E.I., Barrett-Wilt, G.A., Rabaglia, M.E., Keller, M.P., Attie, A.D., Rey, F.E., and Denu, J.M. (2016). Diet-Microbiota Interactions Mediate Global Epigenetic Programming in Multiple Host Tissues. Mol. Cell *64*, 982–992.

Kreznar, J.H., Keller, M.P., Traeger, L.L., Rabaglia, M.E., Schueler, K.L., Stapleton, D.S., Zhao, W., Vivas, E.I., Yandell, B.S., Broman, A.T., et al. (2017). Host Genotype and Gut Microbiome Modulate Insulin Secretion and Diet-Induced Metabolic Phenotypes. Cell Rep. *18*, 1739–1750.

Kuipers, F., Bloks, V.W., and Groen, A.K. (2014). Beyond intestinal soap--bile acids in metabolic control. Nat. Rev. Endocrinol. *10*, 488–498.

Kurilshikov, A., Wijmenga, C., Fu, J., and Zhernakova, A. (2017). Host Genetics and Gut Microbiome: Challenges and Perspectives. Trends Immunol. *38*, 633–647.

Leamy, L.J., Kelly, S.A., Nietfeldt, J., Legge, R.M., Ma, F., Hua, K., Sinha, R., Peterson, D.A., Walter, J., Benson, A.K., et al. (2014). Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. Genome Biol. *15*, 552.

Li, D., Chen, H., Mao, B., Yang, Q., Zhao, J., Gu, Z., Zhang, H., Chen, Y.Q., and Chen, W. (2017). Microbial Biogeography and Core Microbiota of the Rat Digestive Tract. Sci. Rep. *8*, 45840.

Li, L., Li, H., Garzel, B., Yang, H., Sueyoshi, T., Li, Q., Shu, Y., Zhang, J., Hu, B., Heyward, S., et al. (2015). SLC13A5 is a novel transcriptional target of the pregnane X receptor and sensitizes drug-induced steatosis in human liver. Mol. Pharmacol. *87*, 674–682.

Li, Y., Tesson, B.M., Churchill, G.A., and Jansen, R.C. (2010). Critical reasoning on causal inference in genome-wide linkage and association studies. Trends Genet. *26*, 493–498.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402–408.

von Loeffelholz, C., Lieske, S., Neuschäfer-Rube, F., Willmes, D.M., Raschzok, N., Sauer, I.M., König, J., Fromm, M.F., Horn, P., Chatzigeorgiou, A., et al. (2017). The human longevity gene homolog INDY and interleukin-6 interact in hepatic lipid metabolism. Hepatology *66*, 616–630.

Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. Appl. Environ. Microbiol. *71*, 8228–8235.

Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., and Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. Nature *489*, 220–230.

MacKinnon, D.P., Fairchild, A.J., and Fritz, M.S. (2007). Mediation analysis. Annu. Rev. Psychol. *58*, 593–614.

Maravillas-Montero, J.L., Burkhardt, A.M., Hevezi, P.A., Carnevale, C.D., Smit, M.J., and Zlotnik, A. (2015). Cutting edge: GPR35/CXCR8 is the receptor of the mucosal chemokine CXCL17. J. Immunol. *194*, 29–33.

Martinot, E., Sèdes, L., Baptissart, M., Lobaccaro, J.-M., Caira, F., Beaudoin, C., and Volle, D.H. (2017). Bile acids and their receptors. Mol. Aspects Med. *56*, 2–9.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

McArdle, B.H., and Anderson, M.J. (2001). Fitting Multivariate Models to Community Data: A Comment on Distance-Based Redundancy Analysis. Ecology *82*, 290–297.

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

McKnite, A.M., Perez-Munoz, M.E., Lu, L., Williams, E.G., Brewer, S., Andreux, P.A., Bastiaansen, J.W.M., Wang, X., Kachman, S.D., Auwerx, J., et al. (2012). Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. PLoS One *7*, e39191.

McMurdie, P.J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One *8*, e61217.

Miyata, M., Yamakawa, H., Hamatsu, M., Kuribayashi, H., Takamatsu, Y., and Yamazoe, Y. (2011). Enterobacteria modulate intestinal bile acid transport and homeostasis through apical sodium-dependent bile acid transporter (SLC10A2) expression. J. Pharmacol. Exp. Ther. *336*, 188–196.

Morgan, A.P., Fu, C.-P., Kao, C.-Y., Welsh, C.E., Didion, J.P., Yadgary, L., Hyacinth, L., Ferris, M.T., Bell, T.A., Miller, D.R., et al. (2015). The Mouse Universal Genotyping Array: From Substrains to Subspecies. G3 (Bethesda). *6*, 263–279.

Neto, E.C., Broman, A.T., Keller, M.P., Attie, A.D., Zhang, B., Zhu, J., and Yandell, B.S. (2013). Modeling causality for pairs of phenotypes in system genetics. Genetics *193*, 1003–1013.

Northfield, T.C., and McColl, I. (1973). Postprandial concentrations of free and conjugated bile acids down the length of the normal human small intestine. Gut *14*, 513–518.

O'Connor, A., Quizon, P.M., Albright, J.E., Lin, F.T., and Bennett, B.J. (2014). Responsiveness of cardiometabolic-related microbiota to diet is influenced by host genetics. Mamm. Genome *25*, 583–599.

Oelkers, P., Kirby, L.C., Heubi, J.E., and Dawson, P.A. (1997). Primary bile acid malabsorption caused by mutations in the ileal sodium-dependent bile acid transporter gene (SLC10A2). J. Clin. Invest. *99*, 1880–1887.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., et al. (2018). vegan: Community Ecology Package.

Onishi, J.C., Campbell, S., Moreau, M., Patel, F., Brooks, A.I., Zhou, Y.X., Häggblom, M.M., and Storch, J. (2017). Bacterial communities in the small intestine respond differently to those in the caecum and colon in mice fed low- and high-fat diets. Microbiology *163*, 1189–1197.

Org, E., Parks, B.W., Joo, J.W.J., Emert, B., Schwartzman, W., Kang, E.Y., Mehrabian, M., Pan, C., Knight, R., Gunsalus, R., et al. (2015). Genetic and environmental control of host-gut microbiota interactions. Genome Res. *25*, 1558–1569.

Out, C., Patankar, J. V, Doktorova, M., Boesjes, M., Bos, T., de Boer, S., Havinga, R., Wolters, H., Boverhof, R., van Dijk, T.H., et al. (2015). Gut microbiota inhibit Asbt-dependent intestinal bile acid reabsorption via Gata4. J. Hepatol. *63*, 697–704.

Parks, B.W., Nam, E., Org, E., Kostem, E., Norheim, F., Hui, S.T., Pan, C., Civelek, M., Rau, C.D., Bennett, B.J., et al. (2013). Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. Cell Metab. *17*, 141–152.

Paulson, J.N., Stine, O.C., Bravo, H.C., and Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. Nat. Methods *10*, 1200–1202.

Pesta, D.H., Perry, R.J., Guebre-Egziabher, F., Zhang, D., Jurczak, M., Fischer-Rosinsky, A., Daniels, M.A., Willmes, D.M., Bhanot, S., Bornstein, S.R., et al. (2015). Prevention of diet-induced hepatic steatosis and hepatic insulin resistance by second generation antisense oligonucleotides targeted to the longevity gene mIndy (Slc13a5). Aging (Albany. NY). *7*, 1086–1093.

Pierre, J.F., Martinez, K.B., Ye, H., Nadimpalli, A., Morton, T.C., Yang, J., Wang, Q., Patno, N., Chang, E.B., and Yin, D.P. (2016). Activation of bile acid signaling improves metabolic phenotypes in high-fat diet-induced obese mice. Am. J. Physiol. Gastrointest. Liver Physiol. *311*, G286-304.

Plovier, H., Everard, A., Druart, C., Depommier, C., Van Hul, M., Geurts, L., Chilloux, J., Ottman, N., Duparc, T., Lichtenstein, L., et al. (2017). A purified membrane protein from Akkermansia muciniphila or the pasteurized bacterium improves metabolism in obese and diabetic mice. Nat. Med. *23*, 107–113.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One *5*, e9490.

Purcell-Huynh, D.A., Weinreb, A., Castellani, L.W., Mehrabian, M., Doolittle, M.H., and Lusis, A.J. (1995). Genetic factors in lipoprotein metabolism. Analysis of a genetic cross between inbred mouse strains NZB/BINJ and SM/J using a complete linkage map approach. J. Clin. Invest. *96*, 1845–1858.

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature *490*, 55–60.

Rao, A., Kosters, A., Mells, J.E., Zhang, W., Setchell, K.D.R., Amanso, A.M., Wynn, G.M., Xu, T., Keller, B.T., Yin, H., et al. (2016). Inhibition of ileal bile acid uptake protects against nonalcoholic fatty liver disease in high-fat diet-fed mice. Sci. Transl. Med. *8*, 357ra122-357ra122.

Ridlon, J.M., Kang, D.-J., and Hylemon, P.B. (2006). Bile salt biotransformations by human intestinal bacteria. J. Lipid Res. *47*, 241–259.

Ridlon, J.M., Harris, S.C., Bhowmik, S., Kang, D.-J., and Hylemon, P.B. (2016). Consequences of bile salt biotransformations by intestinal bacteria. Gut Microbes *7*, 22–39.

Romano, K.A., Martinez-Del Campo, A., Kasahara, K., Chittim, C.L., Vivas, E.I., Amador-Noguez, D., Balskus, E.P., and Rey, F.E. (2017). Metabolic, Epigenetic, and Transgenerational Effects of Gut Bacterial Choline Consumption. Cell Host Microbe *22*, 279–290.e7.

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P.I., Godneva, A., Kalka, I.N., Bar, N., et al. (2018). Environment dominates over host genetics in shaping human gut microbiota. Nature *555*, 210–215.

Russell, D.W. (2003). The enzymes, regulation, and genetics of bile acid synthesis. Annu. Rev. Biochem. *72*, 137–174.

Sayin, S.I., Wahlström, A., Felin, J., Jäntti, S., Marschall, H.-U., Bamberg, K., Angelin, B., Hyötyläinen, T., Orešič, M., and Bäckhed, F. (2013). Gut microbiota regulates bile acid metabolism by reducing the levels of tauro-beta-muricholic acid, a naturally occurring FXR antagonist. Cell Metab. *17*, 225–235.

Sehayek, E., Hagey, L.R., Fung, Y.-Y., Duncan, E.M., Yu, H.J., Eggertsen, G., Björkhem, I., Hofmann, A.F., and Breslow, J.L. (2006). Two loci on chromosome 9 control bile acid composition: evidence that a strong candidate gene, Cyp8b1, is not the culprit. J. Lipid Res. *47*, 2020–2027.

Shekels, L.L., Lyftogt, C.T., and Ho, S.B. (1996). Bile acid-induced alterations of mucin production in differentiated human colon cancer cell lines. Int. J. Biochem. Cell Biol. *28*, 193–201.

Shi, L., Song, L., Fitzgerald, M., Maurer, K., Bagashev, A., and Sullivan, K.E. (2014). Noncoding RNAs and LRRFIP1 regulate TNF expression. J. Immunol. *192*, 3057–3067.

Sommer, F., and Bäckhed, F. (2013). The gut microbiota--masters of host development and physiology. Nat. Rev. Microbiol. *11*, 227–238.

Staudinger, J.L., Goodwin, B., Jones, S.A., Hawkins-Brown, D., MacKenzie, K.I., LaTour, A., Liu, Y., Klaassen, C.D., Brown, K.K., Reinhard, J., et al. (2001). The nuclear receptor PXR is a lithocholic acid sensor that protects against liver toxicity. Proc. Natl. Acad. Sci. U. S. A. *98*, 3369–3374.

Su, Z., Cox, A., Shen, Y., Stylianou, I.M., and Paigen, B. (2009a). Farp2 and Stk25 are candidate genes for the HDL cholesterol locus on mouse chromosome 1. Arterioscler. Thromb. Vasc. Biol. *29*, 107–113.

Su, Z., Ishimori, N., Chen, Y., Leiter, E.H., Churchill, G.A., Paigen, B., and Stylianou, I.M. (2009b). Four additional mouse crosses improve the lipid QTL landscape and identify Lipg as a QTL gene. J. Lipid Res. *50*, 2083–2094.

Svenson, K.L., Gatti, D.M., Valdar, W., Welsh, C.E., Cheng, R., Chesler, E.J., Palmer, A.A., McMillan, L., and Churchill, G.A. (2012). High-resolution genetic mapping using the Mouse Diversity outbred population. Genetics *190*, 437–447.

Team, R.C. R: A Language and environment for statistical computing.

Theriot, C.M., Bowman, A.A., and Young, V.B. (2016). Antibiotic-Induced Alterations of the Gut Microbiota Alter Secondary Bile Acid Production and Allow for Clostridium difficile Spore Germination and Outgrowth in the Large Intestine. MSphere *1*, e00045-15.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. Nature *444*, 1027–1031.

Ussar, S., Fujisaka, S., and Kahn, C.R. (2016). Interactions between host genetics and gut microbiome in diabetes and metabolic syndrome. Mol. Metab. *5*, 795–803.

Wahlström, A., Sayin, S.I., Marschall, H.-U., and Bäckhed, F. (2016). Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism. Cell Metab. *24*, 41–50.

Wang, J., Thingholm, L.B., Skiecevičienė, J., Rausch, P., Kummen, M., Hov, J.R., Degenhardt, F., Heinsen, F.-A., Rühlemann, M.C., Szymczak, S., et al. (2016). Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. Nat. Genet. *48*, 1396–1406.

Wang, Z., Klipfell, E., Bennett, B.J., Koeth, R., Levison, B.S., Dugar, B., Feldstein, A.E., Britt, E.B., Fu, X., Chung, Y.-M., et al. (2011). Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature *472*, 57–63.

Wendler, A., and Wehling, M. (2013). PGRMC2, a yet uncharacterized protein with potential as tumor suppressor, migration inhibitor, and regulator of cytochrome P450 enzyme activity. Steroids *78*, 555–558.

Zheng, X., Huang, F., Zhao, A., Lei, S., Zhang, Y., Xie, G., Chen, T., Qu, C., Rajani, C., Dong, B., et al. (2017). Bile acid is a significant host factor shaping the gut microbiome of diet-induced obese mice. BMC Biol. *15*, 120.

Zhernakova, A., Kurilshikov, A., Bonder, M.J., Tigchelaar, E.F., Schirmer, M., Vatanen, T., Mujagic, Z., Vila, A.V., Falony, G., Vieira-Silva, S., et al. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science *352*, 565–569.

FIGURES AND TABLES

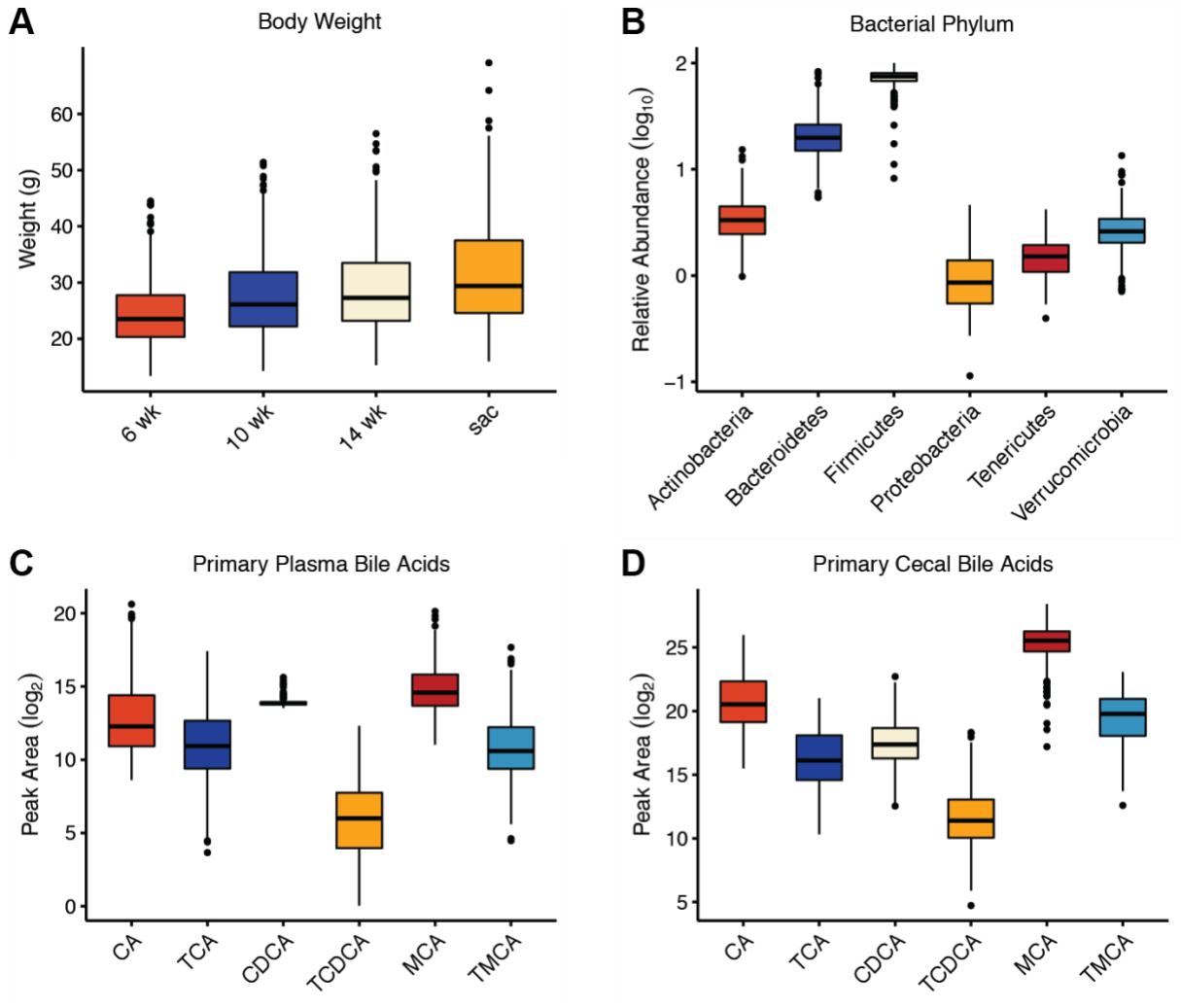


Figure 3.1. Phenotypic variation among Diversity Outbred (DO) mice fed high-fat and high-sucrose diet. (A) Body weight at 6, 10, 14, and 21-25 (sacrifice) weeks in DO mice fed high-fat and high-sucrose diet (n = 500) (Adapted from Keller et al. (Keller et al., 2018)) (B) Distributions of the normalized relative abundance of bacterial phyla identified in DO fecal microbiota (n = 399). (C) Abundance (peak area) of primary bile acids detected in plasma and (D) cecal contents (n = 384).

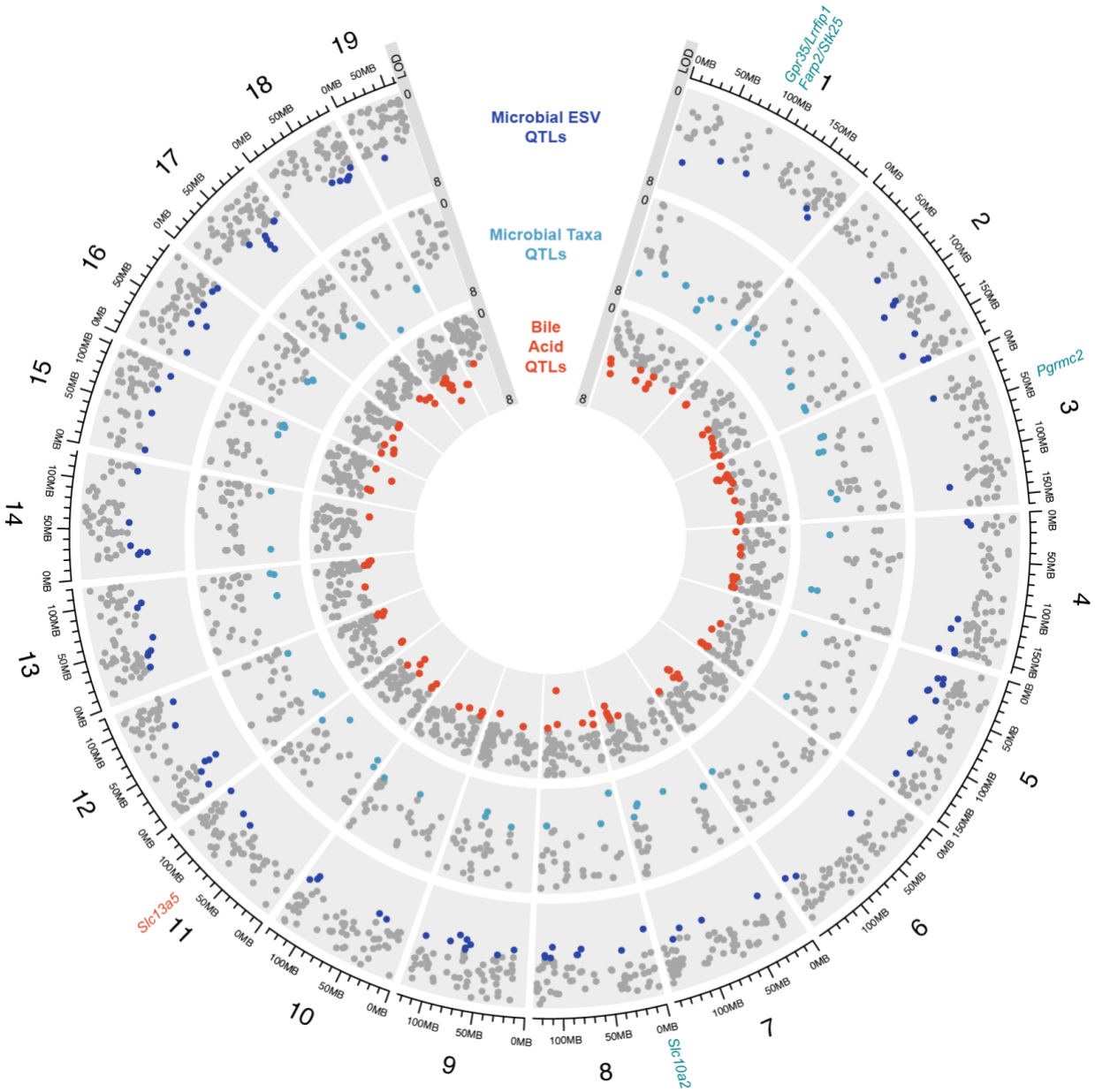


Figure 3.2. Genetic architecture of quantitative trait loci (QTL) for microbial exact sequence variants (ESVs) and taxa abundance, and plasma and cecal bile acids in Diversity Outbred (DO) mice. The outer layer shows the chromosome location where major tick marks correspond to 25 Mbp. Logarithm of the odds (LOD) range is shown for each track. Each dot represents a QTL on each chromosome of the mouse genome for a given trait. Grey dots denote QTLs with LOD < 5.5. Colored dots correspond to QTL with LOD > 5.5.

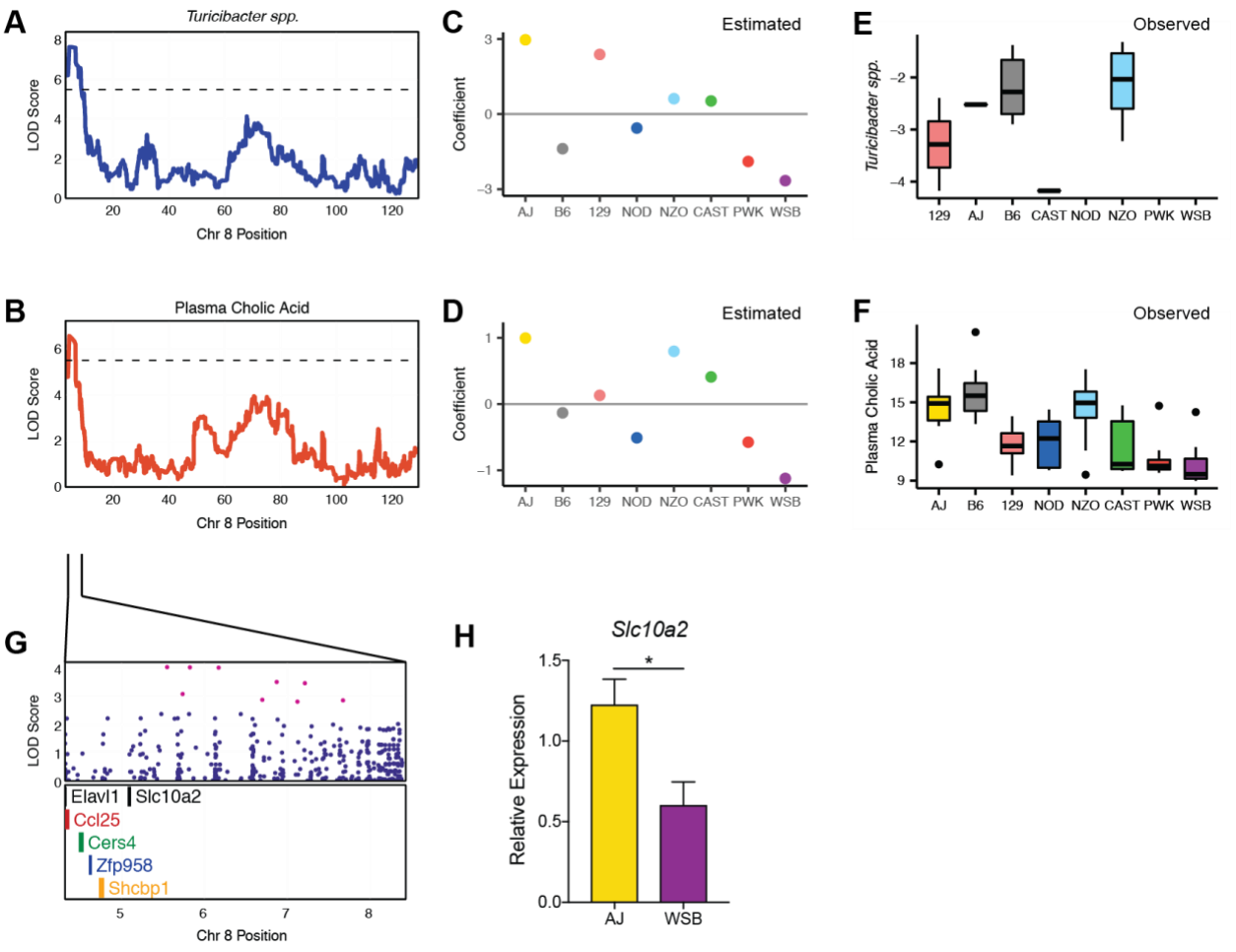


Figure 3.3. Co-mapping of *Turicibacter sp.* and plasma cholic acid (CA) QTL on chromosome 8. Haplotype effects of the eight DO founder strains on the (A) fecal abundance of *Turicibacter sp.* and (B) plasma CA levels. The x-axis indicates the position in Mb along chromosome (chr) 8. The y-axis for the top panel indicates the effect coefficient depicting the estimated contributions of each founder allele, and the y-axis in the bottom panel is the LOD score. A/J and WSB founder alleles are associated with higher and lower levels of *Turicibacter* and plasma CA levels, respectively. The estimated founder strain abundance of (C) *Turicibacter* and (D) levels of plasma CA in the DO population reflects measured values observed in founder strains for (E) the abundance of Turicibacter sp. and (F) plasma cholic acid levels (n = 8 mice/genotype, 4 male and 4 female). (G) SNPs (top panel) and protein coding genes (bottom panel) under the QTL interval.

Magenta dots correspond to SNPs with the strongest association where the LOD drop < 1.5 from the top SNP. (H) Relative expression of Slc10a2 measured in the distal ileum by qRT-PCR in AJ and WSB parental strains (n = 6, 3 male and 3 female). Data are presented as mean ± SEM; Welch's t test; * $p < 0.05$. Correlation p-values adjusted for multiple tests using Benjamini and Hochberg correction.

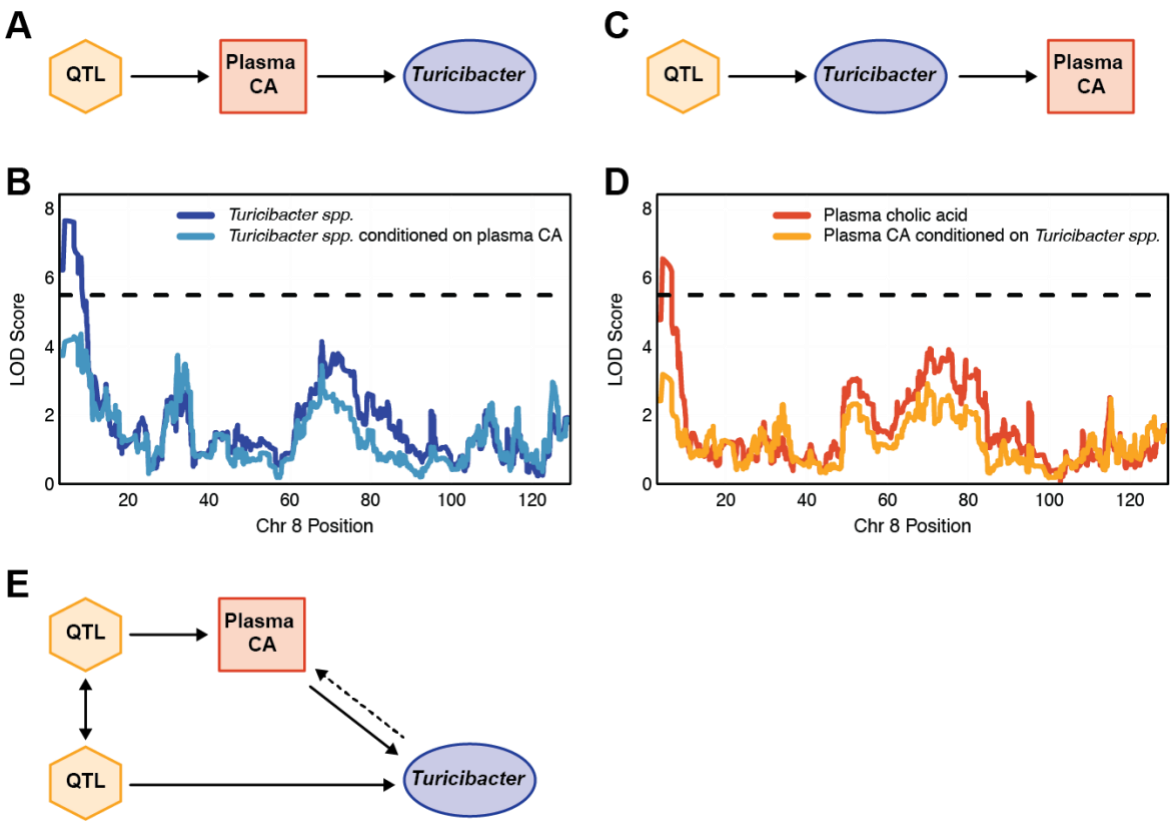


Figure 3.4. Mediation analysis and causal inference testing suggest causal relationship between *Turicibacter sp*. abundance and plasma cholic acid (CA) levels. (A) Hypothetical causal model that proposes that cholic acid (CA) mediates the changes in *Turicibacter sp.* abundance. (B) Change in LOD score of plasma CA when adjusting for *Turicibacter sp*. abundance. (C) Hypothetical causal model that proposes that *Turicibacter sp*. mediates changes in abundance of plasma CA levels. (D) Change in LOD score of *Turicibacter sp.* when controlling for plasma CA levels. (E) Predicted model based on pleiotropy and causal model hypothesis testing.

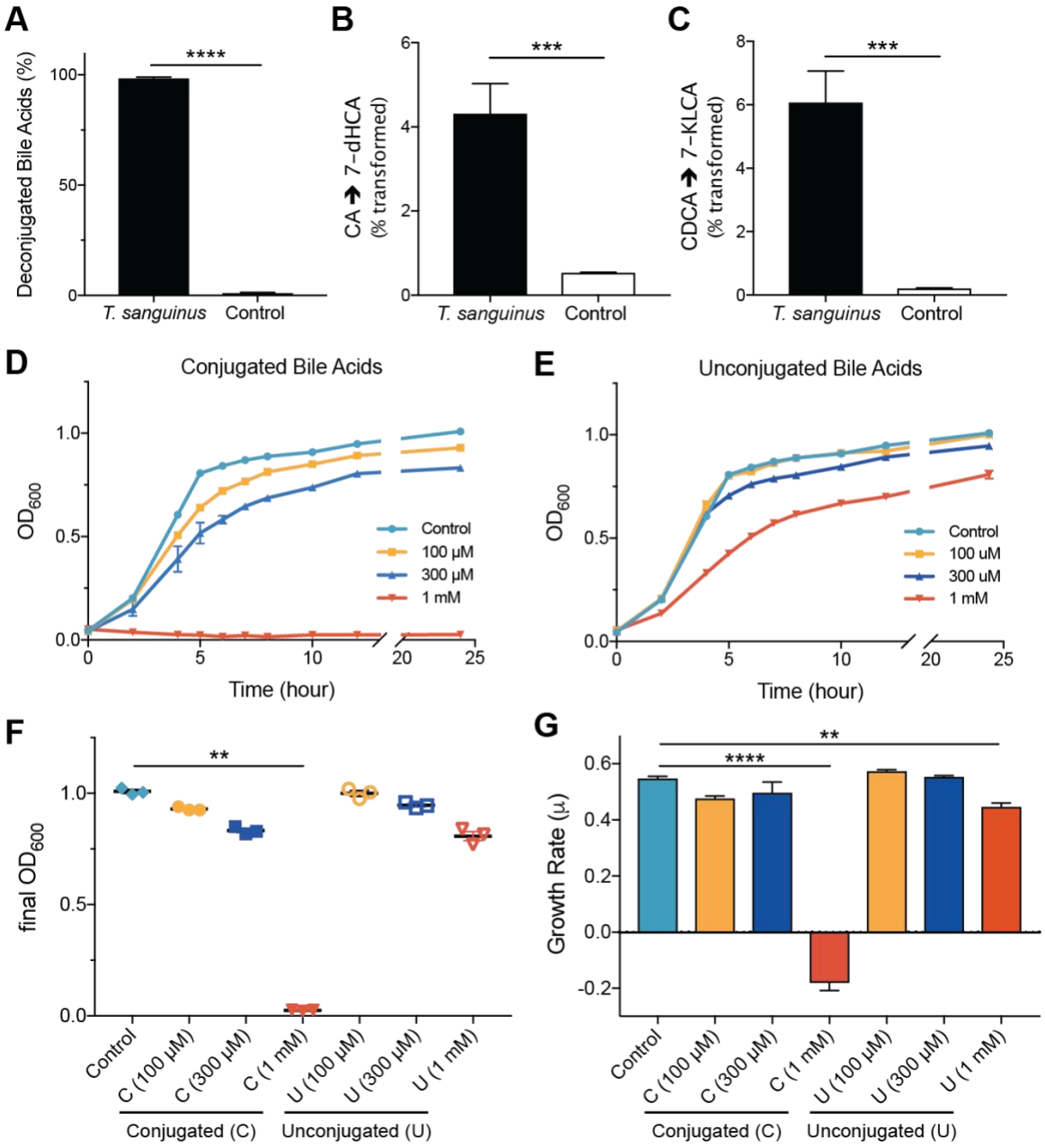
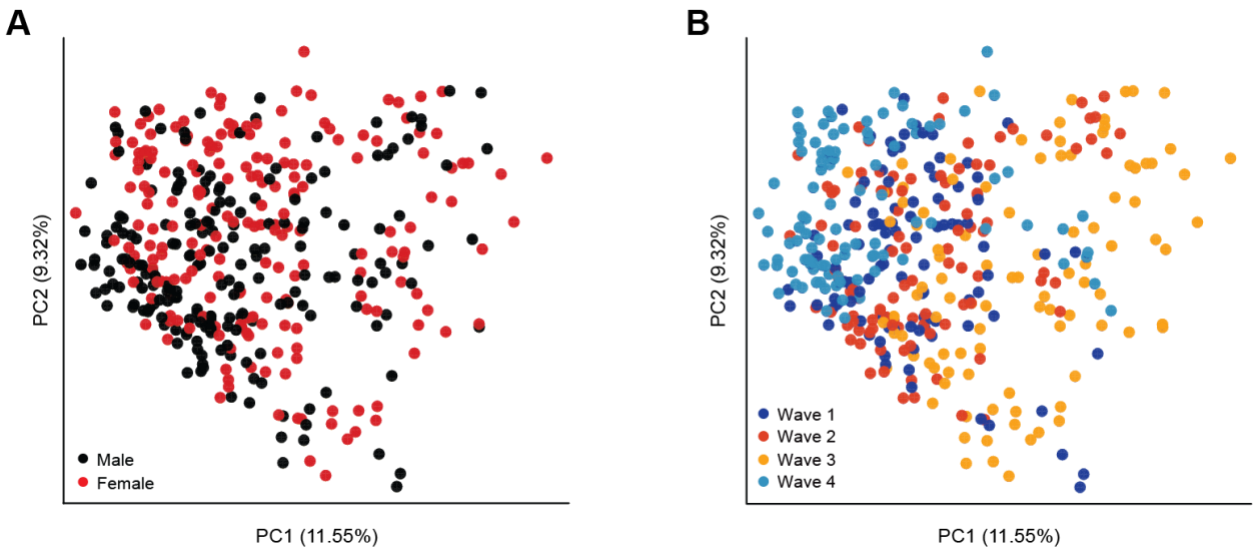
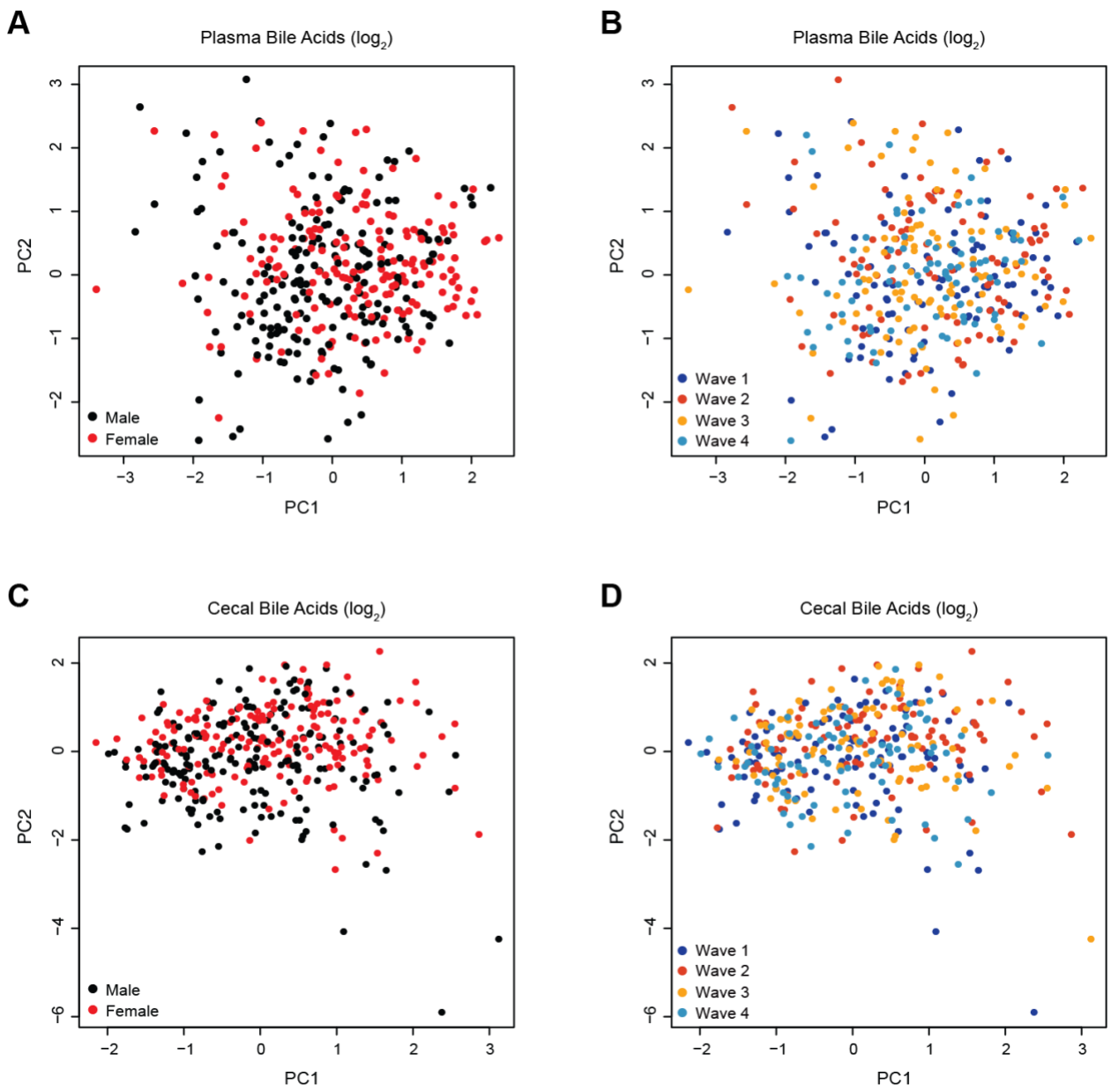


Figure 3.5. Turicibacter sanguinis and bile acid interactions. (A) Percent of conjugated bile acids detected after 24-hour incubation with or without the presence of T. sanguinis. (B) Transformation of cholic acid (CA) to 7-dehydrocholic acid (7-dHCA), and (C) chenodeoxycholic acid (CDCA) to 7-ketolithocholic acid (7-KLCA) by T. sanguinis after 24 hours. Growth rate of T. sanguinis in the presence of 100uM, 300uM and 1mM (D) conjugated (equimolar pool of taurocholic acid

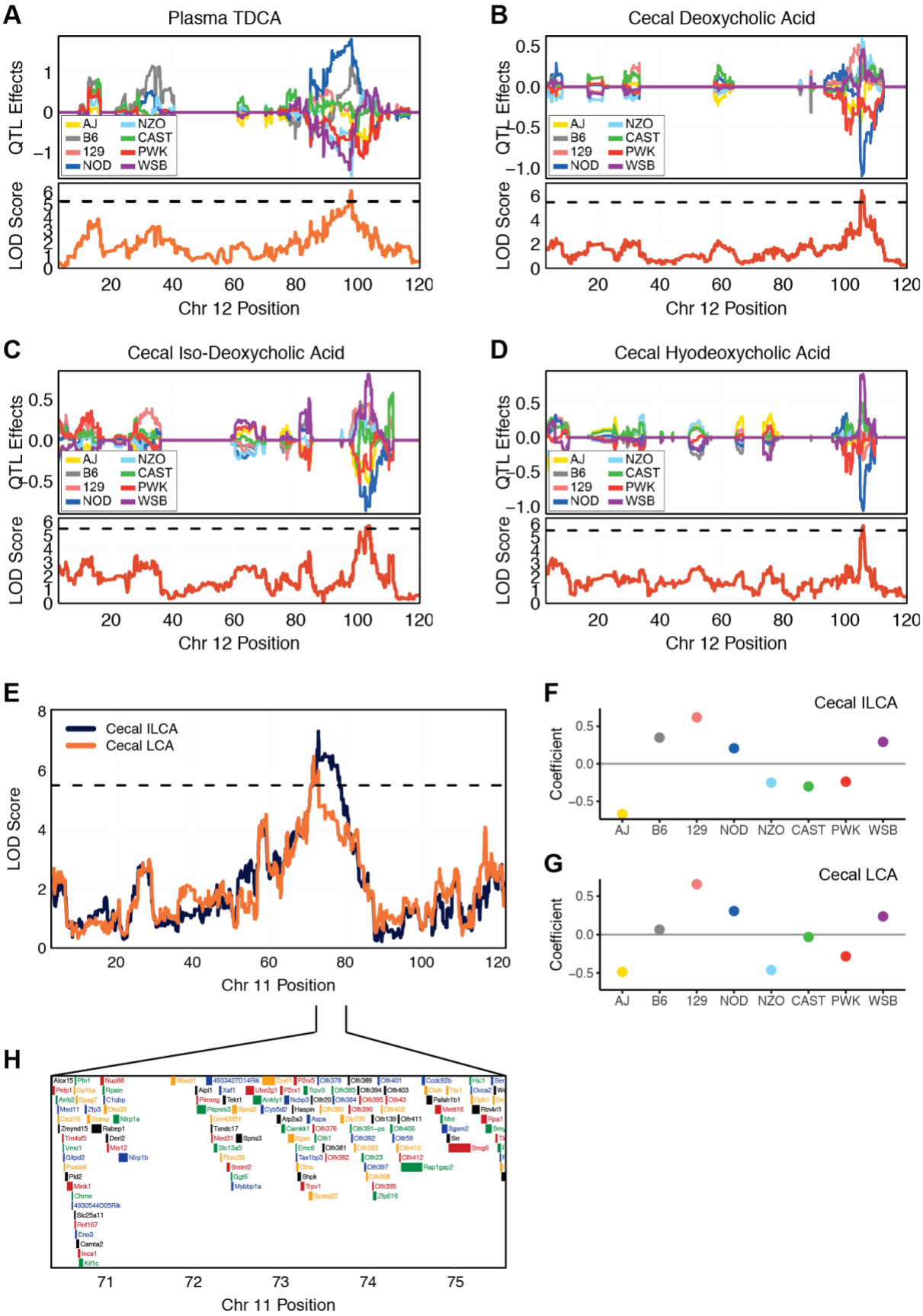
(TCA) and glycochenodeoxycholic acid (GCDCA)), and (E) unconjugated (equimolar pool of cholic CA, CDCA, and deoxycholic acid (DCA)) bile acids over 24 hours. (F) T. sanguinis yield after 24 hours of incubation with varying concentrations of conjugated (c) and unconjugated (u) bile acids, as determined by optical density at 600nm. (G) Growth rate of T. sanguinis in medium supplemented with varying concentrations of conjugated and unconjugated bile acids. Data shown are from one experiment with three technical replicates. Data are presented as mean ± SEM; one-way ANOVA followed by Tukey's multiple comparisons test; ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.



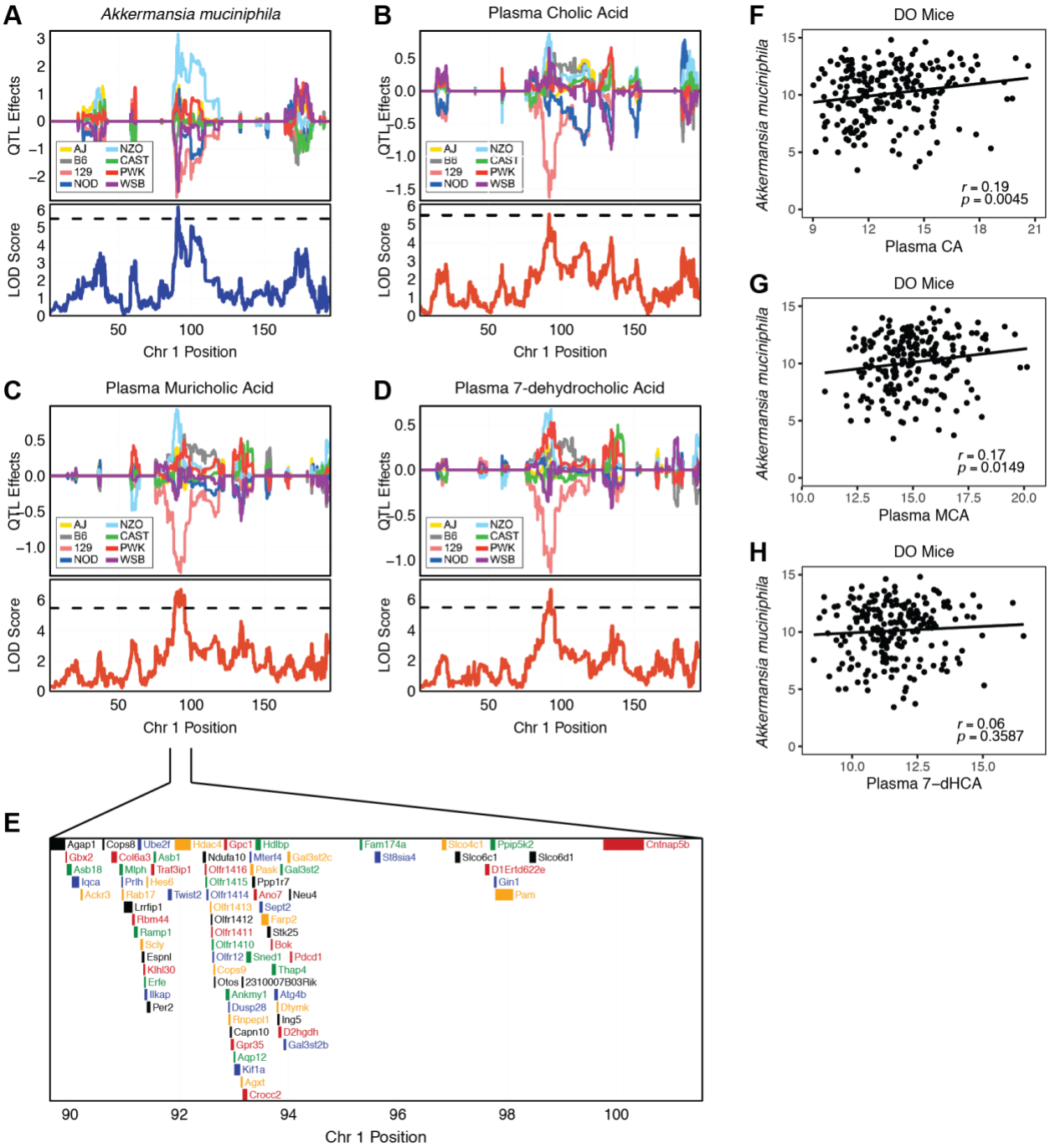
Supplemental Figure 3.1. Principal coordinate analysis (PCoA) of unweighted UniFrac distances for fecal samples shows significant clustering by (A) wave (F = 16.9535, p = 0.001) and (B) sex (F = 5.57169, p = 0.001). Clustering by treatment evaluated by PERMANOVA.



Supplemental Figure 3.2. Plasma and cecal bile acids do not group by batch or sex. PCAs of plasma bile acid profiles colored by (A) sex (PC1, p = 2.2e-16; PC2, p = 0.001696) and (B) batch (PC1, p = 0.5937; PC2, p = 0.4588), and PCAs of cecal bile acid profiles colored by (C) sex (PC1, p = 0.011; PC2, p = 8.4e-05) and (D) batch (PC1, p = 0.2072; PC2, p = 0.009). Kruskal Wallis one-way test followed by Wilcoxon pair-wise multiple comparisons with Benjamini and Hochberg correction.

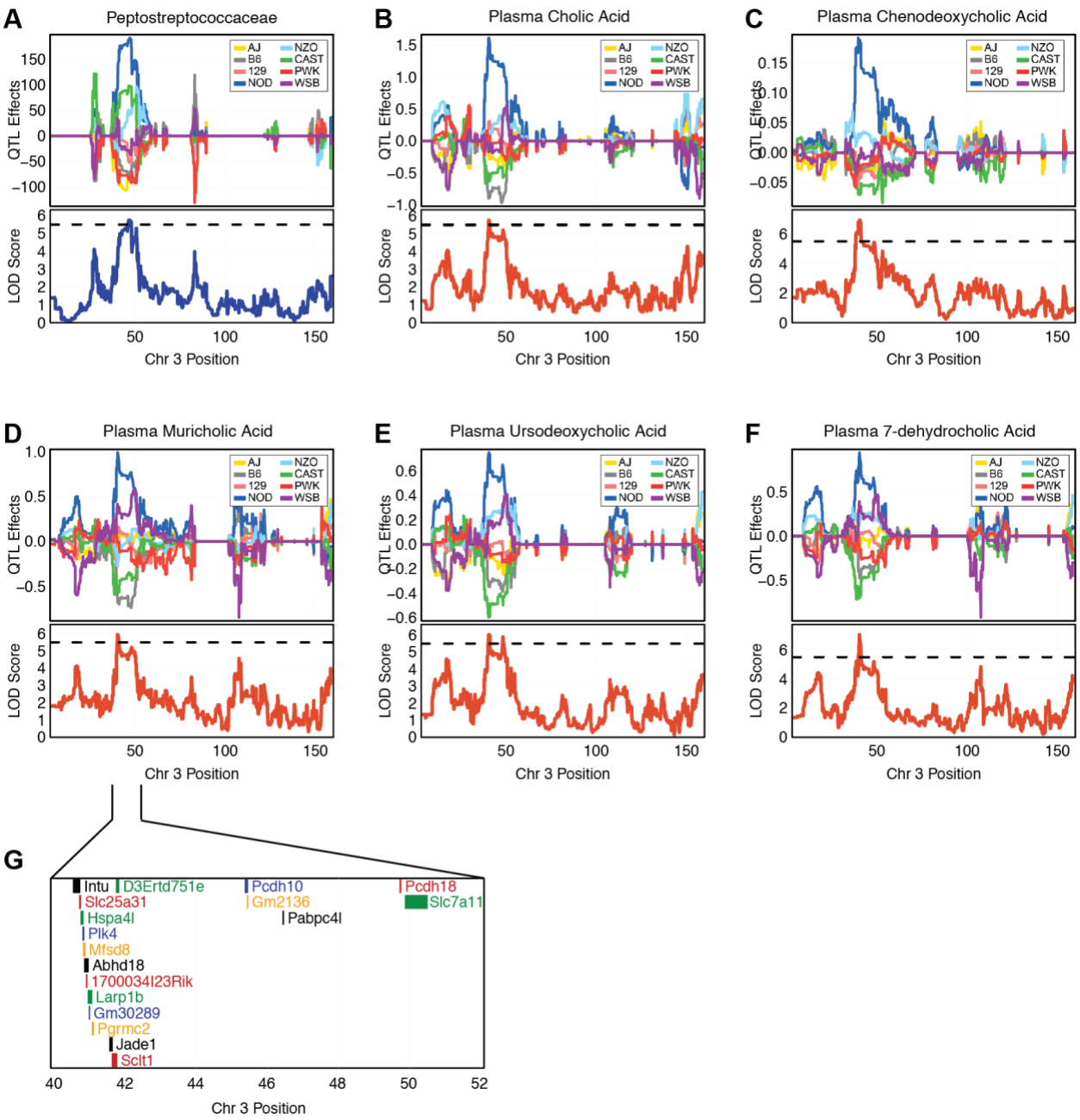

A
Plasma TDCA
B
Cecal Deoxycholic Acid
C
Cecal Iso-Deoxycholic Acid
D
Cecal Hyodeoxycholic Acid
QTL Effects
LOD Score
AJ
B6
129
NOD
NZO
CAST
PWK
WSB
Chr 12 Position
E
Cecal ILCA
Cecal LCA
Chr 11 Position
F
Cecal ILCA
G
Cecal LCA
Coefficient
H
Chr 11 Position

Supplemental Figure 3.3. Related bile acid species map associate to same locus. (A) Cecal levels of isolithocholic acid (ILCA) and lithocholic acid (LCA) associate to same locus on chr 11. (B) Estimated founder allele effects for cecal ILCA and (C) LCA. (D) Genes under cecal LCA and ILCA QTL interval. (E) Haplotype effects and LOD scores of plasma taurodeoxycholic acid (TDCA), (F) cecal deoxycholic acid (DCA), (G) cecal isodeoxycholic acid (IDCA) and (H) cecal hyodeoxycholic acid (HDCA). For each plot, the x-axis is the physical position in Mb along chr 12. The y-axis for the top panel is the effect coefficient depicting the estimated contributions of each founder allele, and the y-axis in the bottom panel is the LOD score.



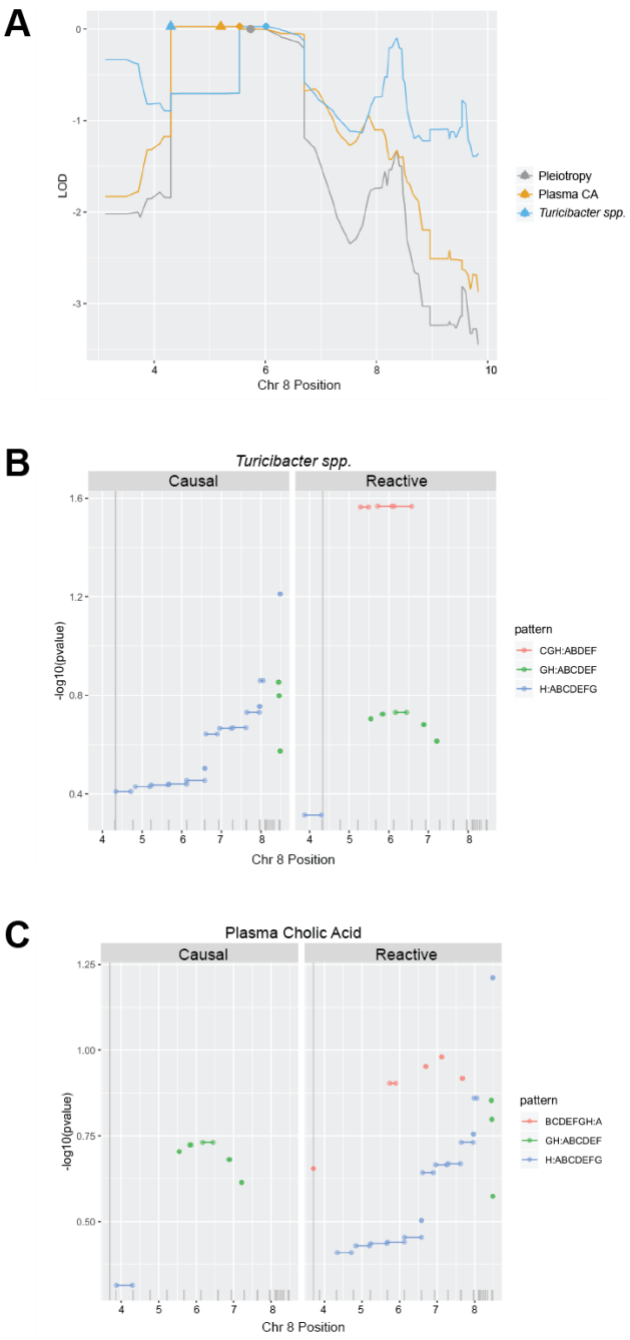
Supplemental Figure 3.4. Exact sequence variant of *Akkermansia muciniphila* and plasma bile acid QTL overlap on chr 1. Haplotype effects and LOD scores of (A) *A. muciniphila* (B) plasma cholic acid (CA), (C) plasma muricholic acid (MCA), and (D) plasma 7-dehydrocholic acid (7-dHCA). For each plot, the x-axis is the physical position in Mb along chr 1. The y-axis for the top

panel is the effect coefficient depicting the estimated contributions of each founder allele, and the y-axis in the bottom panel is the LOD score. (E) Protein coding genes under 10 Mbp QTL interval. Spearman correlations in the DO mice between *A. muiniphila* and (F) plasma CA, (G) plasma MCA, and (H) plasma 7-dHCA levels. Correlation p-values adjusted for multiple tests using Benjamini and Hochberg correction. Higher levels of these microbial and bile acid traits were associated with the NZO haplotype and lower levels were associated with the 129 haplotype. (E) Protein coding genes under 10 Mbp QTL interval. Spearman correlations in the DO mice between *A. muiniphila* and (F) plasma CA, (G) plasma MCA, and (H) plasma 7-dHCA levels. Correlation p-values adjusted for multiple tests using Benjamini and Hochberg correction



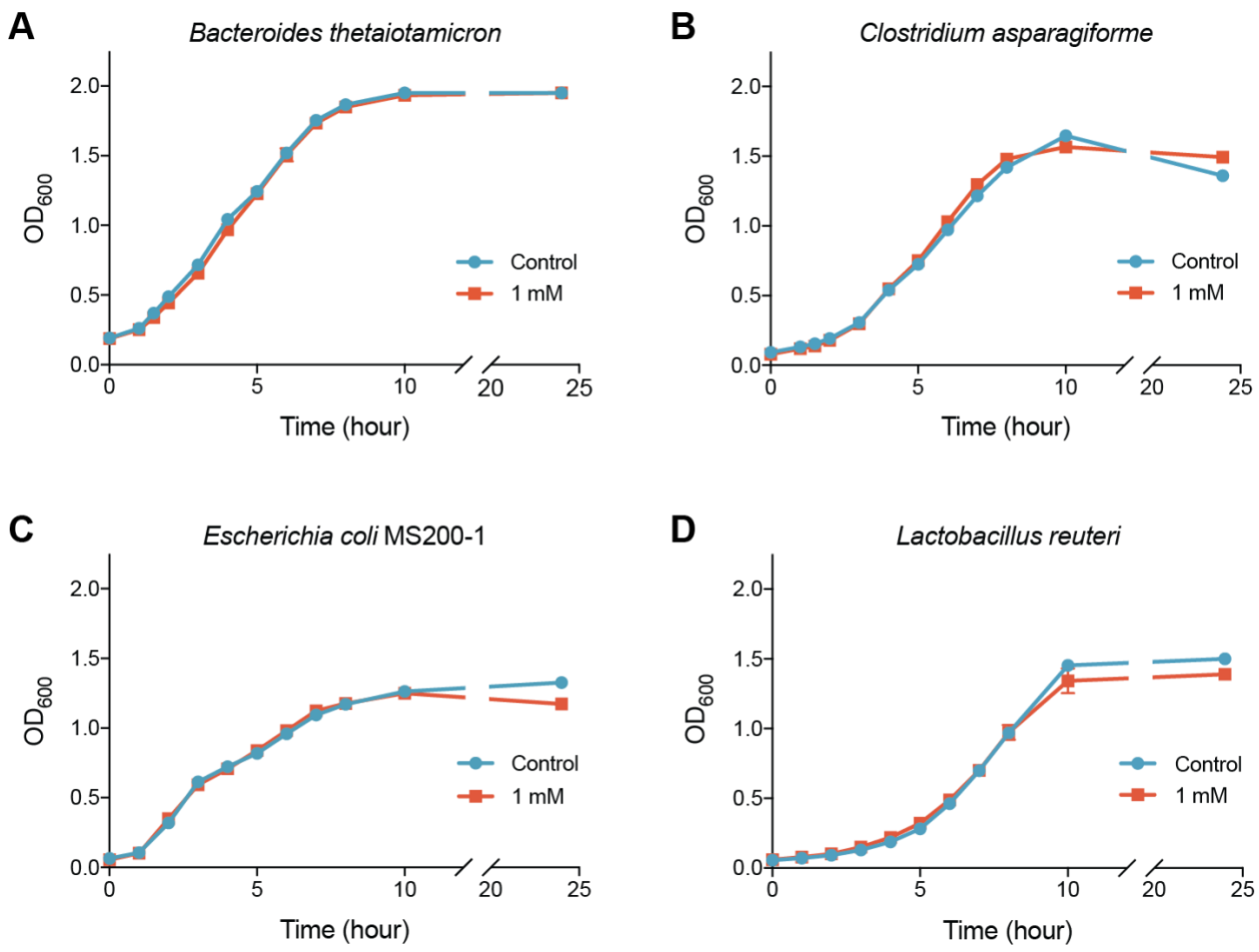
Supplemental Figure 3.5. Peptostreptococcaceae and plasma bile acids co-map on chr 3. Haplotype effects and LOD scores of (A) Peptostreptococcaceae family, (B) plasma cholic acid (CA), (C) plasma chenodeoxycholic acid (CDCA), (D) plasma muricholic acid (MCA), (E) plasma ursodeoxycholic acid (UDCA), and (F) plasma 7-dehydrocholic acid (7-dHCA). For each plot, the x-axis is the physical position in Mb along chr 3. The y-axis for the top panel is the effect

coefficient depicting the estimated contributions of each founder allele, and the y-axis in the bottom panel is the LOD score. All overlapping QTL have positive association with the NOD allele. (G) Protein coding genes under QTL interval.



Supplemental Figure 3.6. Pleiotropy and causal inference testing describe overlapping QTL for *Turicibacter sp*. and plasma cholic acid (CA) on chromosome 8. (A) Profile logarithm of the odds (LOD) curves for close linkage vs pleiotropy hypothesis test for *Turicibacter sp*. abundance and plasma CA levels. Gray trace denotes pleiotropy. Triangles indicate the univariate LOD maxima and diamonds indicate the profile LOD maxima. Pleiotropy analysis performed using 1000

bootstrap samples. (B) *Turicibacter sp*. is reactive as shown by causal model hypothesis testing. Plot indicates SNPs associated with Turicibacter sp. as either causal or reactive. Allele pattern of SNPs denoted by different pattern colors. (G) Plasma CA is reactive to *Turicibacter sp*. as shown by causal model hypothesis testing.



Supplemental Figure 3.7. Gut associated bacteria have differential growth responses to conjugated bile acids. Growth rate in the presence of 1 mM conjugated bile acids or methanol control for (A) *Bacteroides thetaiotamicron*, (B) *Clostridium asparagiforme*, (C) *Escherichia coli* MS200-1, and (D) *Lactobacillus reuteri*. Data shown are from one experiment with three technical replicates. Data are presented as mean ± SEM; Welch's t test; no significant differences were observed between growth conditions for any of the tested organisms.

Table 3.1. Measures of variability of microbial exact sequence variants (ESVs) or taxon (phylum, class, order, family, genus) in DO mice, n = 399. Data presented as normalized read counts. SD, standard deviation.

Rank	Trait	Samples	Mean	Median	SD	Min	Max
ESV	00cd2f6860312475904748780758 9f27	174	8.069	8.051	1.254	2.129	11.147
ESV	02408cd609a6b7f8134da2f8955136ac	200	10.011	10.131	1.712	2.377	13.427
ESV	029ac1f87abeae7dbac1ababc489cfec	328	10.624	10.776	1.495	4.262	13.917
ESV	06538b77450f85903962437f449fa60a	381	7.521	7.606	1.447	3.550	11.297
ESV	0ad13b6d1cd98ad7e8f6ee9909d33114	357	8.552	8.479	1.377	5.111	12.663
ESV	0e064a94474c099cd422c1ef160ab28c	295	8.343	8.270	1.412	4.399	12.758
ESV	14c78619269b8c45554ab4a8ddf2f22c	237	5.901	5.996	0.913	2.699	8.319
ESV	1930d2ae4018583d606e705beea31bd1	135	5.010	4.947	1.764	0.655	13.218
ESV	1c281deaf71c6d702d1ddadfa953eac6	139	6.001	6.071	1.224	2.687	9.741
ESV	20ac30224f0d6923615fbc03ac84563f	142	7.557	7.655	1.190	3.896	10.924
ESV	270a745647e1f4f2daefddcae78b43c7	230	5.854	6.092	1.505	0.999	10.190
ESV	31217980cfdf305668ba9c8482a34e81	125	4.003	3.946	0.873	2.019	7.147
ESV	3192f0892c08ec282493637f4fa28d60	149	3.635	3.687	0.813	1.548	5.673
ESV	324a68faca87d16c8c64abc6856c10c5	261	5.163	5.158	0.936	2.877	8.093
ESV	3562d3a0374b9f2ed190c1a7aa7dedb7	239	5.248	5.092	1.117	2.832	8.896
ESV	4598e7db6dff8f019ff4f6c50dc05df4	241	5.618	5.447	1.697	1.860	10.179
ESV	50313555c1a4385e25c85b06a5c8ad42	244	5.639	5.730	0.956	3.147	8.628
ESV	5284495840f983847ec0d0b741a9a471	228	8.121	8.508	2.265	1.761	12.195
ESV	538add871396bc31fcc52b4c6c73542a	226	8.794	9.089	1.471	2.464	11.474
ESV	54dcb911e3ab04e9d5b30c3912100b41	249	6.296	6.186	1.588	2.190	11.048
ESV	62dc76a09326c359be5946507e82e9fa	148	4.544	4.491	1.144	1.347	7.206
ESV	6536a7bc84c95500ef05185628c9f407	177	5.586	5.702	0.875	3.273	7.574
ESV	682dde043b6d08cf20021dc1e423446d	156	4.893	4.695	1.435	2.386	11.980
ESV	6a8118b17d1d40cf877db24449ceb616	249	6.282	6.321	1.440	2.153	9.439
ESV	6e77545756ddc39c81c925b6fb1c8b17	142	5.671	5.647	1.114	2.478	8.587
ESV	727992952afbbf55406b97c29855c9c2	316	6.260	6.318	1.523	2.644	10.691
ESV	72e530a4206f2e4804d502f4dfca8387	391	9.963	9.973	1.276	5.571	13.596
ESV	7b28c20e72c6c95b3e604f0849245770	222	10.107	10.529	2.368	3.433	14.816
ESV	80dd71f0135ba7b0b937f699b90d18d8	145	5.074	5.136	0.993	1.662	7.946
ESV	811f358e4e89eecbd75298ffce878cba	157	4.634	4.334	1.324	2.626	11.324
ESV	82b6b13c800633859b139cf0569b4cda	373	8.555	8.477	1.412	4.755	12.066
ESV	86555947ddfc576c1bb9c42deb3998df	268	5.286	5.333	0.909	1.769	7.744
ESV	89cb9aae28895b309fa5446ed4fa0817	278	9.636	9.940	1.635	2.968	12.951
ESV	8c65d91e013525d1c8606fab40e0a88b	163	6.014	6.064	1.422	1.541	9.712
ESV	8da18b476b64f6f053647455b6890f80	376	7.459	7.523	0.940	1.339	10.025
ESV	8e74f6f63a5f9a304aeb8284be71fd23	318	6.552	6.590	0.888	0.999	9.417
ESV	901c9973bde010c1a5c49023e63cc252	247	5.387	5.431	1.060	1.713	8.810
ESV	912487364b4fda5993a03f47d60111c9	189	4.345	4.362	0.839	2.475	6.882
ESV	924668f542df9de35c2226e4a6cf09bd	148	7.688	7.701	1.198	5.003	10.887
ESV	958d78a02bef69a795806f97adb117ea	241	8.671	8.831	1.010	4.916	10.688
ESV	9b4036adb487adfe4ca484346b5b9a7b	237	7.535	7.359	1.165	4.988	10.487
ESV	9b89b811be9292552e327f16a217ee6f	127	3.854	3.924	0.549	2.468	5.371
ESV	9c3b0cc9bb690b78f97f20c102089c55	297	7.108	7.041	1.130	3.782	9.967
ESV	9f3ae9f6096976f00cd7c13a4717ff1c	196	5.381	5.439	1.237	1.881	10.722
ESV	a728a1dce17d5afb4677e6cc919c2391	266	5.549	5.625	1.183	2.385	9.428
ESV	b0f809cbabf3ac99182581cc095868b2	289	8.817	9.063	1.368	2.342	11.901
ESV	b155d97b79acc3086ce118451afa0124	301	9.973	10.247	1.460	3.272	12.508
ESV	b1652dc664d4c8cc2d123d26687d1204	144	4.834	4.848	1.324	1.955	7.356
ESV	b65410081c87ed629211d7fb136ec45f	288	6.988	7.096	1.315	1.210	10.000

Table 3.1, ***Continued.***

Rank	Trait	Samples	Mean	Median	SD	Min	Max
ESV	ba19cb8193a715cbfc3d44d03cc4a608	191	4.780	4.843	0.846	2.104	7.838
ESV	bbee9f3d2ff7eae779b70c7ac2c971e4	291	7.736	7.778	1.297	3.114	10.694
ESV	beb1407c19f7fa27ce3bbc0aa7f36157	144	5.896	6.054	1.237	2.420	9.322
ESV	bfa3561ca0d31bca54be08e2d9f7937f	196	5.422	5.453	1.222	1.667	9.675
ESV	c6f4aa25b9a6b316e50c3f0308c05ab9	201	5.513	5.601	1.094	0.580	8.491
ESV	c75ae6008025d80134bd14e9712f9d5c	230	6.733	6.489	1.390	3.874	11.287
ESV	c977f14c76481a059ef7ba611bb8b1d0	161	5.741	5.787	1.169	2.036	9.124
ESV	cb8b8a6cc6bcfa115f8b24a45fa1c12a	154	6.439	6.493	1.741	2.390	11.974
ESV	cbdc0f4fcc7b1e56aacc302c2b9b870c	268	7.413	7.657	1.310	2.716	9.935
ESV	cc82a5a0b284dea47583c8afa5a99755	135	6.909	6.973	0.933	2.490	8.858
ESV	ce78414357068359ddae6a47b6592a08	345	7.095	7.230	1.324	1.842	10.582
ESV	d114fb4c335125128be28401522dd41a	392	11.908	12.113	1.726	5.894	15.891
ESV	d6cda88bd8370a52076b0dafdc12249a	205	5.898	5.957	1.401	2.197	9.675
ESV	daae43be6cf06991f62a085ba8bff3b6	212	8.201	8.090	1.051	5.307	11.200
ESV	df0e3d38eec730326754d8c17a8b8efe	198	10.407	11.004	2.540	0.996	14.375
ESV	e04cb8c96d35fee0e181a15fc4511d0c	204	4.519	4.481	0.864	1.062	6.937
ESV	ef3a40c4f26b5019887d73ceaaab84f2	301	8.791	8.815	1.558	4.100	12.852
ESV	f3e9d78daeea42d345080748e31ae3dd	153	6.547	6.412	1.290	3.984	9.978
ESV	f9c752d5c510e07c457e029af2d6d31b	156	3.869	3.676	1.072	1.843	10.521
class	Bacilli	397	23.471	21.338	12.350	0.914	120.627
class	Bacteroidia	396	77.834	76.991	23.967	14.216	169.558
class	Clostridia	397	265.702	263.955	108.537	5.016	569.749
class	Coriobacteriia	388	12.729	13.050	3.450	5.317	27.067
class	Erysipelotrichi	110	11.718	10.887	7.634	1.381	32.040
class	Gammaproteobacteria	103	4.487	3.721	2.855	0.484	16.522
class	Mollicutes	68	6.363	6.223	2.227	2.584	16.331
class	Verrucomicrobiae	222	10.094	10.387	2.507	3.708	18.635
family	[Mogibacteriaceae]	238	5.660	4.923	2.655	0.471	14.681
family	Anaeroplasmataceae	60	5.993	6.207	1.656	2.584	9.610
family	Bacteroidaceae	68	14.074	12.400	5.540	2.967	26.964
family	Christensenellaceae	222	5.465	5.515	1.757	0.464	12.713
family	Clostridiaceae	99	13.876	10.487	9.267	0.948	44.111
family	Coriobacteriaceae	388	12.729	13.050	3.450	5.317	27.067
family	Dehalobacteriaceae	75	5.124	5.172	0.957	1.713	7.058
family	Enterobacteriaceae	99	4.553	3.979	2.744	0.484	13.477
family	Erysipelotrichaceae	110	11.718	10.887	7.634	1.381	32.040
family	Lachnospiraceae	395	126.030	130.745	53.326	9.158	267.013
family	Peptostreptococcaceae	90	16.686	19.425	8.210	1.264	30.271
family	Ruminococcaceae	393	54.676	49.730	28.003	4.799	132.370
family	S24-7	395	74.584	73.785	23.926	20.087	165.716
family	Staphylococcaceae	41	8.025	6.524	3.910	4.133	21.743
family	Streptococcaceae	392	13.727	12.520	4.463	5.867	51.742
family	Turicibacteraceae	199	12.040	10.960	6.301	0.948	36.953
family	Verrucomicrobiaceae	222	10.094	10.387	2.507	3.708	18.635
genus	[Ruminococcus]	284	8.639	7.457	4.537	2.476	26.215
genus	Adlercreutzia	388	12.621	12.935	3.365	5.317	27.067
genus	Akkermansia	222	10.094	10.387	2.507	3.708	18.635
genus	Anaeroplasma	60	5.993	6.207	1.656	2.584	9.610
genus	Bacteroides	68	14.074	12.400	5.540	2.967	26.964
genus	Coprococcus	354	26.617	26.441	9.796	3.643	51.791
genus	Dehalobacterium	75	5.124	5.172	0.957	1.713	7.058

Table 3.1, ***Continued.***

Rank	Trait	Samples	Mean	Median	SD	Min	Max
genus	Dorea	330	13.421	13.680	5.626	3.098	28.636
genus	Escherichia	75	4.088	3.640	1.900	0.484	9.719
genus	Lactococcus	392	13.683	12.473	4.431	5.867	51.742
genus	Oscillospira	387	52.154	48.195	26.497	5.927	130.493
genus	Ruminococcus	76	7.871	6.315	4.220	2.697	26.398
genus	Staphylococcus	41	8.025	6.524	3.910	4.133	21.743
genus	Turicibacter	199	12.040	10.960	6.301	0.948	36.953
order	Anaeroplasmatales	60	5.993	6.207	1.656	2.584	9.610
order	Bacillales	93	7.414	5.798	4.063	1.502	21.743
order	Bacteroidales	396	77.834	76.991	23.967	14.216	169.558
order	Clostridiales	397	265.695	263.955	108.519	5.016	568.751
order	Coriobacteriales	388	12.729	13.050	3.450	5.317	27.067
order	Enterobacteriales	99	4.553	3.979	2.744	0.484	13.477
order	Erysipelotrichales	110	11.718	10.887	7.634	1.381	32.040
order	Lactobacillales	393	15.710	13.373	8.180	0.914	104.225
order	Turicibacterales	199	12.040	10.960	6.301	0.948	36.953
order	Verrucomicrobiales	222	10.094	10.387	2.507	3.708	18.635
phyla	Actinobacteria	388	12.880	13.070	3.706	5.317	33.054
phyla	Bacteroidetes	396	77.841	76.991	23.958	14.216	169.558
phyla	Firmicutes	399	290.969	288.123	109.443	8.676	607.951
phyla	Proteobacteria	118	4.915	4.019	3.446	0.484	19.723
phyla	Tenericutes	68	6.363	6.223	2.227	2.584	16.331
phyla	Verrucomicrobia	222	10.094	10.387	2.507	3.708	18.635

Table 3.2. Measures of variability of cecal and plasma bile acids in DO mice. Bile acid levels are presented as log2(peak area), n = 384. SD, standard deviation

Tissue	Bile Acid Species	Samples	Mean	Median	SD	Min	Max
cecal	ACA	383	19.962	20.209	1.514	10.857	22.968
cecal	CA	383	19.618	20.394	5.488	-3.921	25.981
cecal	CDCA	383	17.473	17.389	1.728	12.546	22.717
cecal	DCA	383	25.229	25.457	1.665	16.053	28.378
cecal	GDCA	383	5.789	7.268	4.410	-3.921	11.607
cecal	GdHCA	383	3.037	4.614	4.079	-3.921	8.444
cecal	GLCA	383	0.361	0.613	4.312	-3.921	8.764
cecal	GUDCA	383	0.561	-3.921	4.786	-3.921	10.534
cecal	HDCA	383	20.919	21.270	1.878	10.889	24.416
cecal	IDCA	383	13.537	13.614	1.596	7.700	17.503
cecal	ILCA	383	16.041	16.077	1.344	9.208	19.180
cecal	LCA	383	21.001	21.114	1.439	13.285	24.408
cecal	MCA	383	25.336	25.547	1.469	17.210	28.406
cecal	TaMCA.TbMCA	383	19.446	19.786	1.963	12.598	23.068
cecal	TCA	383	15.265	16.001	4.958	-3.921	21.023
cecal	TCDCA	383	11.439	11.381	2.798	-3.921	18.332
cecal	TDCA	383	14.633	14.941	3.718	-3.921	20.342
cecal	TdHCA	383	6.022	6.839	2.991	-3.921	10.172
cecal	THDCA	383	13.118	13.628	3.709	-3.921	18.871
cecal	TLCA	383	11.382	11.525	2.173	3.003	16.133
cecal	TUDCA	383	13.467	13.378	2.104	7.925	18.858
cecal	TwMCA	383	16.709	16.976	2.634	-3.921	21.807
cecal	UCA	383	8.645	8.786	2.088	-0.352	15.132
cecal	UDCA	383	19.629	19.751	1.672	13.495	24.426
cecal	12-KLCA	383	18.557	18.795	1.391	9.358	21.101
cecal	3-dHCA	383	15.231	15.206	1.722	10.893	20.904
cecal	7-dHCA	383	20.650	20.830	1.624	13.316	23.731
plasma	ACA	384	10.041	9.835	1.639	6.101	15.155
plasma	CA	384	12.749	12.285	2.366	8.623	20.625
plasma	CDCA	384	13.894	13.852	0.223	13.547	15.630
plasma	DCA	384	14.091	14.087	1.699	10.426	19.192
plasma	GDCA	384	-3.891	-4.133	1.297	-4.133	7.043
plasma	GdHCA	384	-4.079	-4.133	0.475	-4.133	1.464
plasma	GLCA	384	-3.856	-4.133	1.255	-4.133	5.558
plasma	GUDCA	384	-4.047	-4.133	0.828	-4.133	8.323
plasma	HDCA	384	10.919	10.787	1.055	7.935	15.759
plasma	IDCA	384	9.082	9.043	0.959	5.885	14.023
plasma	ILCA	384	9.644	9.668	0.733	6.478	11.572
plasma	LCA	384	10.229	10.229	0.800	7.798	12.932
plasma	MCA	384	14.777	14.593	1.602	11.038	20.138
plasma	TaMCA.TbMCA	384	10.800	10.602	2.256	-4.133	17.677
plasma	TCA	384	10.942	10.930	2.268	3.667	17.417
plasma	TCDCA	384	4.015	5.184	4.462	-4.133	12.310
plasma	TDCA	384	7.253	8.309	3.860	-4.133	12.795
plasma	TdHCA	384	-3.506	-4.133	1.950	-4.133	5.701
plasma	THDCA	384	5.184	6.682	4.627	-4.133	13.338

Table 3.2, ***Continued.***

Tissue	Bile Acid Species	Samples	Mean	Median	SD	Min	Max
plasma	TLCA	384	-3.650	-4.133	1.761	-4.133	5.891
plasma	TUDCA	384	6.359	7.242	3.649	-4.133	14.142
plasma	TwMCA	384	9.271	9.061	1.260	6.274	15.385
plasma	UCA	384	10.661	10.593	0.780	7.571	14.592
plasma	UDCA	384	11.066	10.924	1.235	8.208	15.599
plasma	12-KLCA	384	9.470	9.468	0.839	7.303	12.933
plasma	3-dHCA	384	9.119	9.017	1.056	6.536	13.789
plasma	7-dHCA	384	11.389	11.258	1.416	8.215	16.559

Table 3.3. QTL peaks for gut microbiota, plasma and cecal bile acid, and weight traits in the Diversity Outbred mice. Only QTL with LOD > 6 shown. "Pos" is peak position is Mbp. "ci_lo" and "ci_hi" correspond to the positions for the 95% bayesian confidence interval.

Group	Trait	Chr	Peak Position	LOD	CI_lo	CI_hi
16S_ESV	811f358e4e89eecbd75298ffce878cba	1	19.371181	8.2610	18.500121	21.127272
16S_ESV	029ac1f87abeae7dbac1ababc489cfec	1	39.156427	6.3519	38.757802	151.593486
16S_ESV	c6f4aa25b9a6b316e50c3f0308c05ab9	1	59.248326	6.1808	56.267238	61.79937
16S_ESV	7b28c20e72c6c95b3e604f0849245770	1	90.979223	6.1609	89.630159	101.5690017
16S_ESV	b155d97b79acc3086ce118451afa0124	1	172.39215	6.3697	40.840301	172.713578
16S_ESV	86555947ddfc576c1bb9c42deb3998df	1	175.199659	6.4246	172.713578	176.044287
16S_ESV	bfa3561ca0d31bca54be08e2d9f7937f	2	108.929956	6.0659	106.511715	113.191065
16S_ESV	cb8b8a6cc6bcfa115f8b24a45fa1c12a	2	115.150381	7.8351	106.499471	115.557252
16S_ESV	270a745647e1f4f2daefddcae78b43c7	2	164.314068	8.4422	164.255184	165.143532
16S_ESV	ba19cb8193a715cbfc3d44d03cc4a608	2	180.240944	6.0044	179.813971	181.334954
16S_ESV	d114fb4c335125128bc28401522dd41a	3	32.655432	6.7619	31.324428	33.888837
16S_ESV	8c65d91c013525d1c8606fab40c0a88b	3	135.232817	7.4494	133.651296	136.132389
16S_ESV	0e064a94474c099cd422c1ef160ab28c	4	53.588189	6.5113	15.725362	54.842888
16S_ESV	5284495840f983847ec0d0b741a9a471	4	127.442688	6.4200	123.134205	127.531669
16S_ESV	324a68faea87d16c8e64abc6856e10e5	4	154.897633	7.2597	154.793084	155.016685
16S_ESV	ce78414357068359ddae6a47b6592a08	5	20.619491	6.2762	17.33613	24.584674
16S_ESV	f3e9d78daeea42d345080748e31ae3dd	5	36.060028	6.7424	32.432453	41.026864
16S_ESV	5284495840f983847ec0d0b741a9a471	5	37.317278	6.8845	37.151983	38.235269
16S_ESV	86555947ddfc576c1bb9c42deb3998df	5	72.347411	6.9511	71.78836	74.125376
16S_ESV	d114fb4c335125128bc28401522dd41a	5	75.965721	7.1440	53.955615	80.679204
16S_ESV	682dde043b6d08cf20021dc1e423446d	6	41.816538	6.7972	17.806817	44.015743
16S_ESV	31217980cfdf305668ba9c8482a34c81	6	148.878394	6.0052	88.150735	149.721874
16S_ESV	958d78a02bcf69a795806f97adb117ca	7	24.672478	6.1850	19.586003	24.785743
16S_ESV	324a68faea87d16c8c64abc6856c10e5	7	92.518876	6.4509	90.225765	94.342611
16S_ESV	6a8118b17d1d40cf877db24449ceb616	7	119.033971	6.4851	118.459217	120.12465
16S_ESV	df0e3d38eec730326754d8c17a8b8efe	8	5.222420143	7.6229	3	8.532219
16S_ESV	cbdc0f4fcc7b1e56aacc302c2b9b870c	8	78.245881	6.3266	76.590861	81.855579
16S_ESV	f9c752d5c510e07c457e029af2d6d31b	8	110.268158	6.6600	108.809724	112.109201
16S_ESV	b1652dc664d4e8ce2d123d26687d1204	9	15.699149	6.1945	3	28.023622
16S_ESV	c977f14c76481a059ef7ba611bb8b1d0	9	64.462733	6.1142	63.59107	66.396123
16S_ESV	e04cb8c96d35fee0e181a15fc4511d0c	9	69.495874	6.5651	69.401241	70.941615
16S_ESV	0c064a94474c099cd422c1cf160ab28c	9	77.797741	6.9339	74.161547	78.384728
16S_ESV	f3e9d78daeea42d345080748e31ae3dd	9	87.121725	6.0478	85.976427	88.333365
16S_ESV	029ac1f87abcac7dbac1ababc489cfec	10	29.556788	6.0706	28.736097	30.789154
16S_ESV	b0f809cbabf3ac99182581cc095868b2	10	39.288014	6.3124	37.041839	40.711299
16S_ESV	06538b77450f85903962437f449fa60a	10	117.96262	6.6468	116.367023	117.977894
16S_ESV	727992952afbbf55406b97e29855c9c2	10	118.543274	6.2653	116.408963	119.582858
16S_ESV	1930d2ae4018583d606e705beea31bd1	11	75.888435	6.0064	74.732783	93.123381
16S_ESV	ba19eb8193a715ebfc3d44d03ec4a608	11	90.410989	6.3010	89.481658	95.109999
16S_ESV	62dc76a09326c359be5946507e82e9fa	11	121.047954	6.7698	119.004814	122.07865
16S_ESV	9b4036adb487adfe4ca484346b5b9a7b	12	32.113752	6.9593	31.439036	33.372128
16S_ESV	ef3a40c4f26b5019887d73ceaaab84f2	12	33.226705	8.2429	31.287554	33.38232
16S_ESV	4598e7db6dff8f019ff4f6c50dc05df4	12	33.314826	6.4604	31.363295	33.80949
16S_ESV	bcb1407c19f7fa27cc3bbc0aa7f36157	12	108.922998	6.4726	107.834116	110.712645
16S_ESV	daae43be6cf06991f62a085ba8bff3b6	13	42.992646	6.0656	42.542597	43.838146
16S_ESV	b1652dc664d4e8ce2d123d26687d1204	13	56.792444	6.6271	53.609063	57.665623
16S_ESV	8da18b476b64f6f053647455b6890f80	13	97.066118	6.1521	92.434065	109.268801
16S_ESV	727992952afbbf55406b97e29855c9c2	13	111.35717	6.4748	12.728539	112.251959
16S_ESV	3562d3a0374b9f2ed190c1a7aa7dedb7	14	22.565214	6.3396	22.486446	23.929316
16S_ESV	9b4036adb487adfe4ca484346b5b9a7b	14	24.952981	7.4766	24.579786	25.415171

Table 3.3, ***Continued.***

Group	Trait	Chr	Peak Position	LOD	CI_lo	CI_hi
16S_ESV	ef3a40c4f26b5019887d73ceaaab84f2	14	24.952981	6.5684	18.51212117	25.421781
16S_ESV	cb8b8a6cc6bcfa115f8b24a45fa1c12a	14	28.136816	6.0222	26.576227	73.423611
16S_ESV	a728a1dce17d5afb4677e6cc919c2391	14	117.862217	6.5029	46.503464	119.656779
16S_ESV	beb1407c19f7fa27ce3bbc0aa7f36157	15	6.822614	6.0913	4.064824	8.303673
16S_ESV	86555947ddfc576c1bb9c42deb3998df	15	40.218796	6.2829	24.059061	93.879197
16S_ESV	02408cd609a6b7f8134da2f8955136ac	15	88.426196	6.3903	41.698616	96.172368
16S_ESV	8e74f6f63a5f9a304aeb8284be71fd23	15	93.728085	6.3090	93.602473	93.879197
16S_ESV	c75ae6008025d80134bd14e9712f9d5c	16	13.601789	6.8280	10.487495	14.95168
16S_ESV	9b89b811be9292552e327f16a217ee6f	16	50.209134	7.4044	45.655104	52.408954
16S_ESV	8da18b476b64f6f053647455b6890f80	16	69.663826	6.1794	66.905616	97.635634
16S_ESV	270a745647e1f4f2daefddcae78b43c7	17	62.42707	7.1618	60.5743	64.831601
16S_ESV	8c65d91c013525d1c8606fab40c0a88b	17	62.864536	8.0634	62.839716	62.994328
16S_ESV	3192f0892c08cc282493637f4fa28d60	17	63.681254	6.4057	62.850598	73.72302
16S_ESV	02408cd609a6b7f8134da2f8955136ac	17	65.118631	6.0124	55.336739	66.728536
16S_ESV	0e064a94474c099cd422c1ef160ab28c	17	80.523135	6.3224	44.607302	80.896107
16S_ESV	b155d97b79acc3086ce118451afa0124	18	65.528436	6.1314	65.362441	71.610719
16S_ESV	c977f14c76481a059ef7ba611bb8b1d0	18	72.645875	6.3878	71.966545	73.209461
16S_ESV	c75ae6008025d80134bd14e9712f9d5c	18	73.241283	6.4451	71.753426	75.444163
16S_ESV	89cb9aae28895b309fa5446ed4fa0817	18	75.475846	6.2185	75.140066	85.516641
16S_ESV	1c281deaf71c6d702d1ddadfa953eac6	18	76.82683	6.1614	72.861939	77.299959
16S_ESV	82b6b13c800633859b139cf0569b4cda	19	22.651972	6.1118	20.801032	25.101665
16S_ESV	682dde043b6d08cf20021dc1e423446d	X	56.428014	7.3023	53.366141	57.77883
16S_ESV	c75ae6008025d80134bd14e9712f9d5c	X	163.245884	6.8336	161.393378	163.874991
16S_taxon	g_[Ruminococcus]	1	12.732461	6.3062	11.572899	13.275729
16S_taxon	p_Proteobacteria	1	124.238885	6.1384	122.328968	128.840159
16S_taxon	c_Gammaproteobacteria	1	126.492682	7.0377	122.328968	135.088167
16S_taxon	o_Enterobacteriales	1	126.492682	6.9091	122.328968	135.128708
16S_taxon	f_Enterobacteriaceae	1	126.492682	6.9091	122.328968	135.128708
16S_taxon	f_Clostridiaceae	1	174.605368	7.0998	172.481584	189.196043
16S_taxon	f_Christensenellaceae	2	31.001321	6.1271	30.745909	168.905507
16S_taxon	g_Ruminococcus	3	51.8356	6.1115	51.275005	52.486241
16S_taxon	f_Bacteroidaceae	4	19.978905	6.2606	19.724371	20.226775
16S_taxon	g_Bacteroides	4	19.978905	6.2606	19.724371	20.226775
16S_taxon	f_Streptococcaceae	4	114.63299	6.8545	110.195087	117.431836
16S_taxon	g_Lactococcus	4	114.63299	7.4226	110.195087	117.431836
16S_taxon	g_Oscillospira	5	26.233194	6.9093	25.708607	28.187788
16S_taxon	g_Coprococcus	7	73.317519	6.4531	72.879931	73.792766
16S_taxon	p_Proteobacteria	7	120.166967	6.2931	119.174583	121.265902
16S_taxon	f_Ruminococcaceae	7	122.764241	6.1346	6.886117	123.769458
16S_taxon	o_Turicibacterales	8	4.316585571	8.0438	3	8.457798
16S_taxon	f_Turicibacteraceae	8	4.316585571	8.0438	3	8.457798
16S_taxon	g_Turicibacter	8	4.316585571	8.0438	3	8.457798
16S_taxon	f_[Mogibacteriaceae]	9	78.915194	6.2119	75.331319	81.097855
16S_taxon	g_Coprococcus	11	8.167195	6.1775	6.846503	8.65504
16S_taxon	f_Clostridiaceae	11	87.759758	7.0462	87.321774	92.645335
16S_taxon	c_Bacilli	12	19.867285	6.5144	16.402611	24.633072
16S_taxon	g_Dorea	12	29.335653	6.2333	26.703078	33.380272
16S_taxon	c_Erysipelotrichi	13	73.735345	6.3341	72.552722	76.790659
16S_taxon	o_Erysipelotrichales	13	73.735345	6.3341	72.552722	76.790659
16S_taxon	f_Erysipelotrichaceae	13	73.735345	6.3341	72.552722	76.790659
16S_taxon	f_[Mogibacteriaceae]	13	108.902354	6.3898	106.514566	116.777127
16S_taxon	p_Actinobacteria	13	111.35717	6.1075	109.591937	111.83217

Table 3.3, ***Continued.***

Group	Trait	Chr	Peak Position	LOD	CI_lo	CI_hi
16S_taxon	g_[Ruminococcus]	14	22.531425	6.3042	22.262318	23.291899
16S_taxon	g_Dorea	14	117.892982	6.0684	112.765789	124.867725
16S_taxon	p_Bacteroidetes	17	78.535623	6.0403	78.360998	79.442392
16S_taxon	c_Bacteroidia	17	78.535623	6.0312	78.360998	79.442392
16S_taxon	o_Bacteroidales	17	78.535623	6.0312	78.360998	79.442392
16S_taxon	p_Proteobacteria	18	23.734117	8.6701	23.103739	23.787565
16S_taxon	f_Streptococcaceae	18	83.941055	6.1560	83.314147	84.209135
16S_taxon	g_Lactococcus	18	83.941055	6.0076	83.080879	84.209135
16S_taxon	p_Tenericutes	X	108.1758067	6.1138	106.795755	113.454148
16S_taxon	c_Mollicutes	X	108.1758067	6.1138	106.795755	113.454148
16S_taxon	o_Anaeroplasmatales	X	108.1758067	6.1575	105.509305	113.454148
16S_taxon	f_Anaeroplasmataceae	X	108.1758067	6.1575	105.509305	113.454148
16S_taxon	g_Anaeroplasma	X	108.1758067	6.1575	105.509305	113.454148
bile_acid	cecal_HDCA	1	18.180457	6.1190	16.219251	124.761482
bile_acid	cecal_TCA	1	25.409657	7.2389	16.219251	25.648999
bile_acid	plasma_X7dHCA	1	92.814235	6.6724	89.99822	95.084861
bile_acid	plasma_MCA	1	92.877907	6.7112	88.592613	95.552177
bile_acid	cecal_UDCA	1	125.76851	6.9489	119.608537	128.520288
bile_acid	cecal_DCA	1	125.76851	6.2822	121.317748	128.520288
bile_acid	cecal_UDCA	2	109.540781	6.1553	26.29342	128.664063
bile_acid	cecal_TdHCA	2	169.623584	6.2600	169.541177	169.631838
bile_acid	cecal_TUDCA	3	9.242439	6.0983	9.100785	15.797831
bile_acid	plasma_GdHCA	3	39.635478	6.9090	37.37487	50.421529
bile_acid	plasma_X7dHCA	3	40.321565	7.0417	39.653779	41.176845
bile_acid	plasma_UDCA	3	40.530684	6.0251	16.86208	50.719235
bile_acid	plasma_CDCA	3	40.530684	6.9568	38.268133	42.537396
bile_acid	plasma_TLCA	3	42.802943	6.7119	41.082964	47.037036
bile_acid	cecal_ACA	5	96.525149	6.0102	92.605094	100.397159
bile_acid	cecal_GUDCA	5	143.920616	6.1562	105.844204	144.727745
bile_acid	cecal_TaMCA,TbMCA	5	147.944084	6.0672	146.734615	148.213321
bile_acid	cecal_GUDCA	6	85.546851	6.7921	72.801787	86.505014
bile_acid	cecal_TUDCA	6	88.138776	7.0356	81.150653	90.891845
bile_acid	plasma_CA	7	122.188931	7.0122	120.871261	122.264107
bile_acid	cecal_TCA	7	122.283363	6.1908	119.171629	122.618828
bile_acid	plasma_CA	8	4.316585571	6.5561	3.863668286	6.588675
bile_acid	plasma_TUDCA	8	94.619354	12.4973	92.258469	94.93712
bile_acid	cecal_TUDCA	9	103.011666	6.3463	102.22695	103.631026
bile_acid	plasma_GUDCA	11	8.562401	6.1073	5.590572	33.693427
bile_acid	cecal_LCA	11	71.782692	6.4792	70.661194	73.111862
bile_acid	cecal_ILCA	11	72.974476	7.3233	71.436346	78.151256
bile_acid	plasma_TCDCA	12	15.804608	6.4534	14.081757	15.859788
bile_acid	plasma_X7dHCA	12	16.484186	6.6199	15.856173	16.640678
bile_acid	plasma_HDCA	12	97.887484	6.1323	93.463652	98.294021
bile_acid	cecal_DCA	12	105.488733	6.4398	104.995986	106.319764
bile_acid	cecal_TUDCA	13	101.781132	6.1881	101.296192	103.001155
bile_acid	cecal_X12KLCA	13	108.844173	6.3889	107.895874	114.297395
bile_acid	cecal_DCA	13	108.844173	6.3664	107.96907	108.951876
bile_acid	cecal_ACA	13	108.848034	6.3020	107.895874	109.317842
bile_acid	plasma_UDCA	13	117.098378	6.3666	116.843339	120.387272
bile_acid	cecal_TwMCA	13	118.273532	6.4503	115.755144	119.024281
bile_acid	plasma_GLCA	14	94.567949	6.4134	94.205664	94.7929
bile_acid	plasma_CDCA	15	27.22439	6.0282	26.581796	30.307434

Table 3.3, ***Continued.***

Group	Trait	Chr	Peak Position	LOD	CI_lo	CI_hi
bile_acid	plasma_MCA	15	28.317531	6.0230	25.826221	28.487178
bile_acid	plasma_GUDCA	15	66.904956	8.3159	65.597543	88.097035
bile_acid	cecal_LCA	16	24.227344	6.8885	23.663421	26.087386
bile_acid	cecal_THDCA	16	32.8523235	6.7072	32.539119	33.350802
bile_acid	plasma_GLCA	17	75.69297	6.0280	57.071792	76.391316
bile_acid	cecal_CA	17	87.333217	6.9694	86.941496	88.845846
bile_acid	cecal_HDCA	17	87.482811	6.1049	28.673505	91.030924
bile_acid	plasma_TCA	18	39.473025	6.5987	34.860924	44.53206
bile_acid	plasma_HDCA	18	40.497868	6.0634	34.798997	43.062062
bile_acid	cecal_UDCA	18	42.162702	6.2414	41.751758	49.591482
bile_acid	cecal_DCA	18	42.277142	6.9832	41.750711	43.062062
bile_acid	cecal_TaMCA.TbMCA	18	46.410922	6.4716	41.751758	82.492788
bile_acid	cecal_TwMCA	18	46.410922	9.2661	44.53206	47.404138
bile_acid	cecal_GUDCA	18	82.346839	7.2851	81.832735	82.492788
bile_acid	plasma_GDCA	18	86.129718	7.4071	85.394126	90.672596
bile_acid	plasma_GUDCA	X	107.447444	7.4236	95.048232	145.180627
tissue_weight	liver_weight	9	66.709912	9.3300	62.273806	67.850302
tissue_weight	heart_weight	12	97.14182	7.1416	92.527894	97.887484
tissue_weight	fat_pad_weight	17	45.97704	6.6703	45.301577	46.576551
tissue_weight	heart_weight	19	3	7.5563	3	4.267368
tissue_weight	liver_weight	X	62.509448	6.0148	57.312056	70.827748
weight	weight_sac	1	55.765244	6.7403	53.963403	135.682079
weight	weight_14wk	2	135.298676	6.4476	28.187457	136.018886
weight	weight_sac	3	151.191183	6.2401	150.317473	151.942191
weight	weight_6wk	9	44.26364	6.2283	44.128836	45.224039
weight	weight_6wk	17	34.180423	6.7973	31.458566	44.522035
weight	weight_10wk	17	34.660915	6.9212	31.365705	44.067339

Table 3.4. Correlations among microbial taxa, bile acid and weight traits. Spearman's rank correlation. Only microbial exact sequence variants, genera and familiy included in figure. Correlations shown passed FDR < 0.01 cut-off and correlation coefficient either < -0.35 or > 0.35. Correlating bile acids from same tissue removed from table for brevity.

Trait.1	Trait.2	cor	p	p.fdr
f44404ca806c18f6b05b7cbb4f3e5fd2	af26ba64d6f2fab3985dbc4fb2a8241f	-0.60847	0.00028	0.00366
811f358e4e89eecbd75298ffce878cba	f_Clostridiaceae	-0.57988	0.00040	0.00498
1c281deaf71c6d702d1ddadfa953eac6	f_S24-7	-0.49540	0.00000	0.00000
cbdc0f4fcc7b1e56aacc302c2b9b870c	f_Clostridiaceae	-0.49054	0.00004	0.00072
d114fb4c335125128be28401522dd41a	f_Lachnospiraceae	-0.46932	0.00000	0.00000
d6eda88bd8370a52076b0daf'de12249a	f_Clostridiaceae	-0.46453	0.00052	0.00611
f44404ea806c18f6b05b7cbb4f3e5fd2	b155d97b79acc3086ce118451afa0124	-0.46066	0.00019	0.00257
00cd2f6860312475904748780758 9f27	cecal_HDCA	-0.45171	0.00000	0.00000
heart_norm_weightSac	plasma_DCA	-0.45001	0.00000	0.00000
89cb9aae28895b309fa5446ed4fa0817	cc82a5a0b284dea47583c8afa5a99755	-0.44526	0.00000	0.00004
00cd2f6860312475904748780758 9f27	f_S24-7	-0.44119	0.00000	0.00000
6c77545756ddc39c81c925b6fb1c8b17	fat_norm_weightSac	-0.43699	0.00000	0.00003
9c3b0cc9bb690b78f97f20c102089c55	f_Clostridiaceae	-0.43679	0.00006	0.00099
heart_norm_weightSac	plasma_ACA	-0.43498	0.00000	0.00000
270a745647e1f4f2daefddcae78b43c7	f_Clostridiaceae	-0.43337	0.00076	0.00833
ef3a40c4f26b5019887d73ceaaab84f2	f_Clostridiaceae	-0.43164	0.00021	0.00287
weight_14wk	plasma_TDCA	-0.42913	0.00000	0.00000
weight_14wk	plasma_DCA	-0.42393	0.00000	0.00000
weight_14wk	plasma_ACA	-0.42160	0.00000	0.00000
721bbde09abf2f51fc7d8caeab5b22f1	ef3a40c4f26b5019887d73ceaaab84f2	-0.41801	0.00089	0.00956
b155d97b79acc3086ce118451afa0124	f_Peptostreptococcaceae	-0.41530	0.00058	0.00667
82b36f3b6dc8ec399c84ff5907c36ead	cecal_TCA	-0.41407	0.00015	0.00211
86555947ddfc576c1bb9c42dcb3998df	f_Clostridiaceae	-0.41272	0.00038	0.00476
weight_10wk	plasma_DCA	-0.41061	0.00000	0.00000
weight_6wk	plasma_DCA	-0.41036	0.00000	0.00000
8e74f6f63a5f9a304aeb8284be71fd23	f_S24-7	-0.40979	0.00000	0.00000
heart_norm_weightSac	plasma_TDCA	-0.40941	0.00000	0.00000
weight_sac	plasma_DCA	-0.40923	0.00000	0.00000
weight_6wk	plasma_ACA	-0.40838	0.00000	0.00000
weight_sac	plasma_ACA	-0.40612	0.00000	0.00000
weight_10wk	plasma_TDCA	-0.40564	0.00000	0.00000
811f358e4e89eecbd75298ffce878cba	f_S24-7	-0.40437	0.00000	0.00001
d114fb4c335125128be28401522dd41a	g_Oscillospira	-0.40287	0.00000	0.00000
weight_sac	plasma_TDCA	-0.40233	0.00000	0.00000
weight_6wk	plasma_TDCA	-0.40184	0.00000	0.00000
f_Erysipelotrichaceae	f_Lachnospiraceae	-0.40046	0.00003	0.00043
weight_10wk	plasma_ACA	-0.39841	0.00000	0.00000
bbee9f3d2ff7eae779b70c7ac2c971e4	cc82a5a0b284dea47583c8afa5a99755	-0.39701	0.00011	0.00157
82b36f3b6dc8ec399c84ff5907c36cad	cecal_TUDCA	-0.39587	0.00028	0.00363
cb8b8a6ee6bcfa115f8b24a45fa1c12a	50313555c1a4385e25c85b06a5c8ad42	-0.39421	0.00013	0.00190
6e77545756ddc39c81c925b6fb1c8b17	heart_norm_weightSac	-0.39067	0.00002	0.00039
9b4036adb487adfe4ca484346b5b9a7b	f_S24-7	-0.38797	0.00000	0.00000
b155d97b79acc3086ce118451afa0124	g_Akkermansia	-0.38633	0.00000	0.00003
weight_14wk	plasma_TCA	-0.37669	0.00000	0.00000

Table 3.4, ***Continued.***

Trait.1	Trait.2	cor	p	p.fdr
cb8b8a6ee6bcfa115f8b24a45fa1c12a	cbdc0f4fcc7b1e56aacc302c2b9b870c	-0.37587	0.00012	0.00168
heart_norm_weightSac	plasma_TCA	-0.37317	0.00000	0.00000
7b28c20e72c6c95b3e604f0849245770	b155d97b79acc3086ce118451afa0124	-0.37302	0.00000	0.00006
6e77545756ddc39c81c925b6fb1c8b17	weight_sac	-0.37178	0.00001	0.00014
4ad0fb5fe1bf702d971cd3c3fe0ad2d3	cecal_TaMCA.TbMCA	-0.36841	0.00045	0.00539
00cd2f6860312475904748780758 9f27	cecal_UDCA	-0.36799	0.00000	0.00002
4598e7db6dff8f019ff4f6c50dc05df4	f_S24-7	-0.36778	0.00000	0.00000
weight_6wk	plasma_TCA	-0.36607	0.00000	0.00000
f_Lachnospiraceae	f_Streptococcaceae	-0.36492	0.00000	0.00000
b155d97b79acc3086ce118451afa0124	f_Erysipelotrichaceae	-0.36374	0.00058	0.00663
f_Lachnospiraceae	g_Lactococcus	-0.36292	0.00000	0.00000
ecb3759e804e8c86d9fb0de817207fe9	cecal_MCA	-0.36104	0.00026	0.00343
3562d3a0374b9f2ed190c1a7aa7dedb7	f_S24-7	-0.36034	0.00000	0.00000
d114fb4c335125128be28401522dd41a	f_Ruminococcaceae	-0.35985	0.00000	0.00000
cb8b8a6ee6bcfa115f8b24a45fa1c12a	538add871396bc31fee52b4c6e73542a	-0.35709	0.00041	0.00503
4e0a65b0a2fab7bf0f307d360b679d87	f_Ruminococcaceae	-0.35687	0.00033	0.00423
af26ba64d6f2fab3985dbc4fb2a8241f	f_S24-7	-0.35616	0.00022	0.00297
weight_10wk	plasma_TCA	-0.35584	0.00000	0.00000
00cd2f6860312475904748780758 9f27	cecal_LCA	-0.35539	0.00000	0.00004
89cb9aae28895b309fa5446cd4fa0817	4598e7db6dff8f019ff4f6c50dc05df4	-0.35524	0.00000	0.00008
6e77545756ddc39c81c925b6fb1c8b17	weight_10wk	-0.35267	0.00002	0.00038
00cd2f6860312475904748780758 9f27	cecal_MCA	-0.35262	0.00000	0.00005
4ad0fb5fe1bf702d971cd3c3fe0ad2d3	cecal_TUDCA	-0.35254	0.00081	0.00887
25f537c4856de1617e682de90caba215	plasma_HDCA	-0.35063	0.00049	0.00584
029ac1f87abeae7dbac1ababc489cfec	62dc76a09326c359be5946507e82e9fa	0.35029	0.00004	0.00067
0e064a94474c099cd422c1ef160ab28c	a728a1dce17d5afb4677e6cc919c2391	0.35073	0.00000	0.00001
06538b77450f85903962437f449fa60a	g_Adlercreutzia	0.35150	0.00000	0.00000
f_Turicibacteraceae	plasma_DCA	0.35254	0.00000	0.00001
g_Turicibacter	plasma_DCA	0.35254	0.00000	0.00001
plasma_MCA	cecal_MCA	0.35259	0.00000	0.00000
plasma_ACA	cecal_TDCA	0.35279	0.00000	0.00000
f_Streptococcaceae	g_Adlercreutzia	0.35291	0.00000	0.00000
f_Coriobacteriaceae	f_Streptococcaceae	0.35348	0.00000	0.00000
00cd2f6860312475904748780758 9f27	c04cb8c96d35fcc0c181a15fc4511d0c	0.35385	0.00058	0.00666
0c064a94474c099cd422c1ef160ab28c	g_Coprococcus	0.35452	0.00000	0.00000
20ac30224f0d6923615fbc03ac84563f	g_Dorea	0.35467	0.00003	0.00054
270a745647e1f4f2daefddcae78b43c7	1c281deaf71c6d702d1ddadfa953eac6	0.35757	0.00024	0.00321
g_Turicibacter	plasma_MCA	0.35848	0.00000	0.00001
270a745647e1f4f2daefddcae78b43c7	a728a1dce17d5afb4677e6cc919c2391	0.35894	0.00000	0.00003
a8220bca75e4d05eb793d6e501236264	g_Coprococcus	0.35915	0.00016	0.00219
plasma_HDCA	cecal_HDCA	0.36041	0.00000	0.00000
924ab5cb47309a83e71d60320378d194	8da18b476b64f6f053647455b6890f80	0.36125	0.00008	0.00128
811f358e4e89eecbd75298ffce878cba	9e3b0ce9bb690b78f97f20c102089c55	0.36266	0.00002	0.00027
af26ba64d6f2fab3985dbc4fb2a8241f	f_Lachnospiraceae	0.36499	0.00015	0.00213
912487364b4fda5993a03f47d60111c9	811f358c4c89ccebd75298ffce878cba	0.36599	0.00020	0.00265
f44404ea806c18f6b05b7cbb4f3e5fd2	plasma_X7dHCA	0.36678	0.00076	0.00832
8c74f6f63a5f9a304acb8284bc71fd23	g_Coprococcus	0.36798	0.00000	0.00000

Table 3.4, ***Continued.***

Trait.1	Trait.2	cor	p	p.fdr
3562d3a0374b9f2ed190c1a7aa7dedb7	a728a1dce17d5afb4677e6cc919c2391	0.36836	0.00000	0.00000
6d4ab2107d92228c6751838e7dbbe697	8e74f6f63a5f9a304aeb8284be71fd23	0.36848	0.00033	0.00416
df0e3d38eec730326754d8c17a8b8efe	721bbde09abf2f51fc7d8caeab5b22f1	0.36882	0.00056	0.00644
8e74f6f63a5f9a304aeb8284be71fd23	9b4036adb487adfe4ca484346b5b9a7b	0.36892	0.00000	0.00000
029ac1f87abeae7dbac1ababc489cfec	72e530a4206f2e4804d502f4dfca8387	0.36930	0.00000	0.00000
20ac30224f0d6923615fbc03ac84563f	6536a7bc84c95500ef05185628c9f407	0.36945	0.00014	0.00204
8e74f6f63a5f9a304aeb8284be71fd23	31217980cfdf305668ba9c8482a34e81	0.36946	0.00004	0.00067
270a745647c1f4f2dacfddcac78b43c7	811f358c4c89ccebd75298ffce878cba	0.37125	0.00006	0.00094
54dcb911c3ab04c9d5b30c3912100b41	9b4036adb487adfe4ca484346b5b9a7b	0.37135	0.00000	0.00002
924668f542df9dc35c2226c4a6cf09bd	ef3a40c4f26b5019887d73ceaaab84f2	0.37153	0.00003	0.00046
d114fb4c335125128be28401522dd41a	06538b77450f85903962437f449fa60a	0.37224	0.00000	0.00000
cb8b8a6ee6bcfa115f8b24a45fa1c12a	bbee9f3d2ff7eae779b70c7ac2c971e4	0.37224	0.00001	0.00019
1c281deaf71c6d702d1ddadfa953eac6	g_Dorea	0.37270	0.00001	0.00018
86555947ddfc576c1bb9c42deb3998df	00cd2f6860312475904748780758 9f27	0.37361	0.00004	0.00068
g_Coprococcus	g_Dorea	0.37364	0.00000	0.00000
00cd2f6860312475904748780758 9f27	4598e7db6dff8f019ff4f6c50dc05df4	0.37436	0.00006	0.00088
8e74f6f63a5f9a304aeb8284be71fd23	1c281deaf71c6d702d1ddadfa953eac6	0.37461	0.00002	0.00030
3e064c3f883fdd051884a58537182eef	ef3a40c4f26b5019887d73ceaaab84f2	0.37488	0.00038	0.00468
df0e3d38eec730326754d8c17a8b8efe	plasma_MCA	0.37506	0.00000	0.00000
1930d2ac4018583d606c705bcca31bd1	89cb9aac28895b309fa5446ed4fa0817	0.37611	0.00005	0.00081
9b4036adb487adfe4ca484346b5b9a7b	9c3b0ce9bb690b78f97f20c102089c55	0.37770	0.00000	0.00000
958d78a02bcf69a795806f97adb117ca	3562d3a0374b9f2ed190c1a7aa7dedb7	0.37916	0.00000	0.00001
50313555c1a4385c25c85b06a5c8ad42	ef3a40c4f26b5019887d73ceaaab84f2	0.37929	0.00000	0.00000
3562d3a0374b9f2ed190c1a7aa7dedb7	4598e7db6dff8f019ff4f6c50dc05df4	0.37981	0.00000	0.00001
811f358e4e89eecbd75298ffce878cba	ef3a40c4f26b5019887d73ceaaab84f2	0.38028	0.00001	0.00021
324a68faea87d16c8e64abc6856e10e5	g_Coprococcus	0.38137	0.00000	0.00000
6536a7bc84c95500ef05185628c9f407	00cd2f6860312475904748780758 9f27	0.38146	0.00048	0.00570
958d78a02bcf69a795806f97adb117ea	0e064a94474c099cd422c1ef160ab28c	0.38153	0.00000	0.00000
6536a7bc84c95500ef05185628c9f407	g_Dorea	0.38158	0.00000	0.00002
f_Lachnospiraceae	g_Dorea	0.38196	0.00000	0.00000
82b6b13c800633859b139cf0569b4cda	72e530a4206f2e4804d502f4dfca8387	0.38326	0.00000	0.00000
plasma_DCA	cecal_CA	0.38358	0.00000	0.00000
df0e3d38eec730326754d8c17a8b8efe	f_Peptostreptococcaceae	0.38375	0.00037	0.00465
721bbde09abf2f51fc7d8caeab5b22f1	g_Turicibacter	0.38408	0.00031	0.00397
50313555c1a4385c25c85b06a5c8ad42	g_Dorea	0.38428	0.00000	0.00000
df0e3d38eec730326754d8c17a8b8efe	f_Clostridiaceae	0.38545	0.00018	0.00243
ecb3759e804e8c86d9fb0de817207fe9	f_Lachnospiraceae	0.38860	0.00007	0.00108
924ab5cb47309a83e71d60320378d194	912487364b4fda5993a03f47d60111e9	0.38890	0.00047	0.00563
912487364b4fda5993a03f47d60111e9	00cd2f6860312475904748780758 9f27	0.38994	0.00023	0.00301
f44404ea806c18f6b05b7cbb4f3e5fd2	plasma_MCA	0.39092	0.00031	0.00396
g_Turicibacter	plasma_CA	0.39154	0.00000	0.00000
b0f809cbabf3ac99182581ec095868b2	af26ba64d6f2fab3985dbc4fb2a8241f	0.39188	0.00046	0.00555
cb8b8a6ee6bcfa115f8b24a45fa1c12a	89cb9aae28895b309fa5446ed4fa0817	0.39357	0.00000	0.00007
ec82a5a0b284dea47583c8afa5a99755	g_Dorea	0.39395	0.00001	0.00015
8c74f6f63a5f9a304acb8284bc71fd23	3562d3a0374b9f2ed190c1a7aa7dedb7	0.39433	0.00000	0.00000
cb8b8a6ee6bcfa115f8b24a45fa1c12a	f_S24-7	0.39469	0.00000	0.00001
4ad0fb5fc1bf702d971ed3c3fc0ad2d3	f_Streptococcaceae	0.40252	0.00013	0.00192

Table 3.4, ***Continued.***

Trait.1	Trait.2	cor	p	p.fdr
4ad0fb5fe1bf702d971ed3c3fe0ad2d3	g_Lactococcus	0.40252	0.00013	0.00192
ec82a5a0b284dea47583c8afa5a99755	af26ba64d6f2fab3985dbc4fb2a8241f	0.40410	0.00029	0.00380
82b36f3b6dc8ec399c84ff5907c36ead	bbee9f3d2ff7eae779b70c7ac2c971e4	0.40499	0.00034	0.00435
811f358e4e89eecbd75298ffce878cba	cbdc0f4fcc7b1e56aacc302c2b9b870c	0.40538	0.00002	0.00034
270a745647e1f4f2daefddcae78b43c7	3e064c3f883fdd051884a58537182eef	0.40710	0.00039	0.00479
3e064c3f883fdd051884a58537182eef	cbdc0f4fcc7b1e56aacc302c2b9b870c	0.40896	0.00030	0.00383
20ac30224f0d6923615fbc03ac84563f	682dde043b6d08cf20021dc1e423446d	0.41132	0.00018	0.00253
029ac1f87abcac7dbac1ababc489cfcc	g_Akkermansia	0.41172	0.00000	0.00000
a728a1dce17d5afb4677e6cc919c2391	1c281deaf71c6d702d1ddadfa953eac6	0.41191	0.00001	0.00015
029ac1f87abcac7dbac1ababc489cfcc	7b28c20c72c6c95b3c604f0849245770	0.41583	0.00000	0.00000
89cb9aae28895b309fa5446ed4fa0817	02408cd609a6b7f8134da2f8955136ac	0.41983	0.00000	0.00000
df0e3d38eec730326754d8c17a8b8efe	plasma_CA	0.42128	0.00000	0.00000
0e064a94474c099cd422c1ef160ab28c	25f537c4856de1617e682de90caba215	0.42202	0.00012	0.00172
plasma_DCA	cecal_TDCA	0.42300	0.00000	0.00000
00cd2f6860312475904748780758 9f27	a728a1dce17d5afb4677e6cc919c2391	0.42360	0.00000	0.00004
1c281deaf71c6d702d1ddadfa953eac6	31217980cfdf305668ba9c8482a34e81	0.42618	0.00011	0.00162
811f358e4e89eecbd75298ffce878cba	31217980cfdf305668ba9c8482a34e81	0.42749	0.00018	0.00249
a728a1dce17d5afb4677e6cc919c2391	af26ba64d6f2fab3985dbc4fb2a8241f	0.42832	0.00005	0.00084
f_Ruminococcaceae	g_Coprococcus	0.43117	0.00000	0.00000
plasma_ACA	cecal_X12KLCA	0.43141	0.00000	0.00000
bbee9f3d2ff7eae779b70c7ac2c971e4	1930d2ae4018583d606e705beea31bd1	0.43211	0.00000	0.00001
af26ba64d6f2fab3985dbc4fb2a8241f	g_Coprococcus	0.43286	0.00001	0.00011
811f358e4e89eecbd75298ffce878cba	af26ba64d6f2fab3985dbc4fb2a8241f	0.43550	0.00089	0.00956
df0e3d38eec730326754d8c17a8b8efe	f44404ea806c18f6b05b7cbb4f3e5fd2	0.43635	0.00007	0.00111
02408cd609a6b7f8134da2f8955136ac	f_S24-7	0.43814	0.00000	0.00000
811f358e4e89eecbd75298ffce878cba	00cd2f6860312475904748780758 9f27	0.43829	0.00017	0.00231
682dde043b6d08cf20021dc1e423446d	g_[Ruminococcus]	0.43895	0.00000	0.00000
811f358e4e89eecbd75298ffce878cba	1c281deaf71c6d702d1ddadfa953eac6	0.43919	0.00004	0.00061
f_Clostridiaceae	g_Turicibacter	0.44624	0.00001	0.00019
1930d2ae4018583d606e705beea31bd1	02408cd609a6b7f8134da2f8955136ac	0.44746	0.00001	0.00020
0e064a94474c099cd422c1ef160ab28c	00cd2f6860312475904748780758 9f27	0.44877	0.00000	0.00000
e04cb8c96d35fee0e181a15fc4511d0c	d6eda88bd8370a52076b0dafde12249a	0.45268	0.00000	0.00000
924668f542df9dc35c2226c4a6cf09bd	00cd2f6860312475904748780758 9f27	0.45440	0.00016	0.00227
plasma_ACA	cecal_LCA	0.45547	0.00000	0.00000
811f358e4e89eecbd75298ffce878cba	e04cb8c96d35fee0e181a15fc4511d0c	0.45564	0.00000	0.00005
00cd2f6860312475904748780758 9f27	9e3b0ce9bb690b78f97f20c102089c55	0.45713	0.00000	0.00001
029ac1f87abeae7dbac1ababc489cfcc	f_S24-7	0.45738	0.00000	0.00000
b155d97b79acc3086ce118451afa0124	af26ba64d6f2fab3985dbc4fb2a8241f	0.45745	0.00000	0.00004
958d78a02bef69a795806f97adb117ea	af26ba64d6f2fab3985dbc4fb2a8241f	0.45863	0.00001	0.00012
e04cb8c96d35fee0e181a15fc4511d0c	31217980cfdf305668ba9c8482a34e81	0.45928	0.00001	0.00013
g_Coprococcus	g_Oscillospira	0.46109	0.00000	0.00000
d114fb4c335125128be28401522dd41a	4ad0fb5fe1bf702d971ed3c3fe0ad2d3	0.46355	0.00001	0.00015
03d1db1ccc574034f8242e94d57eca3b	270a745647e1f4f2daefddcae78b43c7	0.46777	0.00001	0.00014
25f537c4856de1617e682de90caba215	g_Coprococcus	0.46816	0.00000	0.00006
cbdc0f4fcc7b1e56aacc302c2b9b870c	ef3a40c4f26b5019887d73ccaaab84f2	0.46912	0.00000	0.00000
6536a7bc84c95500ef05185628c9f407	f_Lachnospiraceae	0.47045	0.00000	0.00000
3562d3a0374b9f2cd190c1a7aa7dcdb7	g_[Ruminococcus]	0.47338	0.00000	0.00000

Table 3.4, ***Continued.***

Trait.1	Trait.2	cor	p	p.fdr
4598e7db6dff8f019ff4f6c50dc05df4	g_Dorea	0.47340	0.00000	0.00000
82b36f3b6dc8ec399c84ff5907c36ead	02408cd609a6b7f8134da2f8955136ac	0.47560	0.00007	0.00110
af26ba64d6f2fab3985dbc4fb2a8241f	g_Dorea	0.47728	0.00000	0.00002
00cd2f6860312475904748780758 9f27	g_Dorea	0.47916	0.00000	0.00000
25f537c4856de1617e682de90caba215	f_Lachnospiraceae	0.48109	0.00000	0.00002
c2cd6f99655915682fbe061583f8b578	6a8118b17d1d40cf877db24449ceb616	0.48292	0.00001	0.00021
6536a7bc84c95500ef05185628c9f407	g_Coprococcus	0.48365	0.00000	0.00000
bbcc9f3d2ff7cac779b70c7ac2c971c4	f_S24-7	0.48954	0.00000	0.00000
ef3a40c4f26b5019887d73ceaaab84f2	9c3b0ce9bb690b78f97f20c102089c55	0.49046	0.00000	0.00000
727992952afbbf55406b97e29855c9c2	f_Coriobacteriaceae	0.49092	0.00000	0.00000
ba19eb8193a715ebfc3d44d03ec4a608	1c281deaf71c6d702d1ddadfa953eac6	0.49217	0.00000	0.00005
plasma_DCA	cecal_X12KLCA	0.49244	0.00000	0.00000
6536a7bc84c95500ef05185628c9f407	25f537c4856de1617e682de90caba215	0.49384	0.00002	0.00033
958d78a02bef69a795806f97adb117ea	g_Coprococcus	0.49649	0.00000	0.00000
1c281deaf71c6d702d1ddadfa953eac6	4598e7db6dff8f019ff4f6c50dc05df4	0.49997	0.00000	0.00000
727992952afbbf55406b97e29855c9c2	g_Adlercreutzia	0.50007	0.00000	0.00000
plasma_ACA	cecal_ACA	0.50106	0.00000	0.00000
0e064a94474c099cd422c1ef160ab28c	af26ba64d6f2fab3985dbc4fb2a8241f	0.50140	0.00000	0.00001
9b4036adb487adfe4ca484346b5b9a7b	ef3a40c4f26b5019887d73ceaaab84f2	0.50260	0.00000	0.00000
82b6b13c800633859b139cf0569b4cda	0ad13b6d1cd98ad7c8f6cc9909d33114	0.50314	0.00000	0.00000
00cd2f6860312475904748780758 9f27	af26ba64d6f2fab3985dbc4fb2a8241f	0.50669	0.00011	0.00158
1930d2ac4018583d606c705bcca31bd1	f_S24-7	0.50859	0.00000	0.00000
plasma_DCA	cecal_LCA	0.51120	0.00000	0.00000
a728a1dce17d5afb4677e6cc919c2391	g_Dorea	0.51433	0.00000	0.00000
cc82a5a0b284dea47583c8afa5a99755	1c281deaf71c6d702d1ddadfa953eac6	0.51546	0.00000	0.00001
89cb9aae28895b309fa5446ed4fa0817	f_S24-7	0.52047	0.00000	0.00000
plasma_ACA	cecal_DCA	0.52616	0.00000	0.00000
682dde043b6d08cf20021dc1e423446d	3562d3a0374b9f2ed190c1a7aa7dedb7	0.52785	0.00000	0.00000
bbee9f3d2ff7eae779b70c7ac2c971e4	89cb9aae28895b309fa5446ed4fa0817	0.53028	0.00000	0.00000
811f358e4e89eecbd75298ffce878cba	924668f542df9de35c2226e4a6cf09bd	0.55812	0.00000	0.00000
06538b77450f85903962437f449fa60a	plasma_TdHCA	0.56422	0.00034	0.00428
plasma_DCA	cecal_ACA	0.57220	0.00000	0.00000
f_Clostridiaceae	f_Peptostreptococcaceae	0.57766	0.00000	0.00009
bbcc9f3d2ff7cac779b70c7ac2c971c4	02408cd609a6b7f8134da2f8955136ac	0.58293	0.00000	0.00000
270a745647c1f4f2dacfddcac78b43c7	af26ba64d6f2fab3985dbc4fb2a8241f	0.58744	0.00000	0.00003
029ac1f87abeae7dbac1ababc489cfec	89cb9aae28895b309fa5446ed4fa0817	0.61084	0.00000	0.00000
8e74f6f63a5f9a304aeb8284be71fd23	a8220bca75e4d05eb793d6e501236264	0.62016	0.00000	0.00000
1c281deaf71c6d702d1ddadfa953eac6	af26ba64d6f2fab3985dbc4fb2a8241f	0.62942	0.00000	0.00000
f_Lachnospiraceae	f_Ruminococcaceae	0.63405	0.00000	0.00000
plasma_DCA	cecal_DCA	0.64315	0.00000	0.00000
liver_norm_weightSac	fat_norm_weightSac	0.64683	0.00000	0.00000
f_Lachnospiraceae	g_Oscillospira	0.66551	0.00000	0.00000
weight_6wk	fat_norm_weightSac	0.67503	0.00000	0.00000
8e74f6f63a5f9a304aeb8284be71fd23	ba19eb8193a715ebfc3d44d03ec4a608	0.67696	0.00000	0.00000
c75ac6008025d80134bd14c9712f9d5c	a8220bca75e4d05eb793d6e501236264	0.68693	0.00000	0.00000
f_Lachnospiraceae	g_Coprococcus	0.71349	0.00000	0.00000
6d4ab2107d92228c6751838c7dbbc697	ba19eb8193a715ebfc3d44d03ec4a608	0.71429	0.00040	0.00496

Table 3.4, ***Continued.***

Trait.1	Trait.2	cor	p	p.fdr
heart_norm_weightSac	fat_norm_weightSac	0.72055	0.00000	0.00000
cb8b8a6ee6bcfa115f8b24a45fa1c12a	cfdf39cd6545fb6a7ddb747c3393c225	0.72611	0.00001	0.00015
721bbde09abf2f51fc7d8caeab5b22f1	f_Clostridiaceae	0.75054	0.00000	0.00000
weight_10wk	fat_norm_weightSac	0.75479	0.00000	0.00000
weight_6wk	liver_norm_weightSac	0.75548	0.00000	0.00000
f44404ea806c18f6b05b7cbb4f3e5fd2	f_Peptostreptococcaceae	0.80481	0.00000	0.00000
liver_norm_weightSac	heart_norm_weightSac	0.80513	0.00000	0.00000
weight_14wk	fat_norm_weightSac	0.81024	0.00000	0.00000
weight_10wk	liver_norm_weightSac	0.81065	0.00000	0.00000
weight_6wk	heart_norm_weightSac	0.81159	0.00000	0.00000
weight_14wk	liver_norm_weightSac	0.82436	0.00000	0.00000
weight_sac	liver_norm_weightSac	0.83997	0.00000	0.00000
d114fb4c335125128be28401522dd41a	g_Lactococcus	0.84263	0.00000	0.00000
6e77545756ddc39c81c925b6fb1c8b17	c977f14c76481a059ef7ba611bb8b1d0	0.85520	0.00000	0.00000
weight_sac	fat_norm_weightSac	0.86281	0.00000	0.00000
4e0a65b0a2fab7bf0f307d360b679d87	538add871396bc31fee52b4c6e73542a	0.86488	0.00000	0.00000
c6f4aa25b9a6b316e50c3f0308c05ab9	f_Christensenellaceae	0.86929	0.00000	0.00000
weight_10wk	heart_norm_weightSac	0.87118	0.00000	0.00000
weight_6wk	weight_sac	0.87354	0.00000	0.00000
weight_14wk	heart_norm_weightSac	0.89685	0.00000	0.00000
weight_sac	heart_norm_weightSac	0.91275	0.00000	0.00000
weight_6wk	weight_14wk	0.91552	0.00000	0.00000
weight_10wk	weight_sac	0.93265	0.00000	0.00000
df0e3d38eec730326754d8c17a8b8efe	g_Turicibacter	0.93580	0.00000	0.00000
weight_6wk	weight_10wk	0.95219	0.00000	0.00000
weight_14wk	weight_sac	0.96038	0.00000	0.00000
weight_10wk	weight_14wk	0.96370	0.00000	0.00000
f_Coriobacteriaceae	g_Adlercreutzia	0.97790	0.00000	0.00000
f_Ruminococcaceae	g_Oscillospira	0.98008	0.00000	0.00000
7b28c20e72c6c95b3e604f0849245770	g_Akkermansia	0.99334	0.00000	0.00000
f_Streptococcaceae	g_Lactococcus	0.99658	0.00000	0.00000
f_Verrucomicrobiaceae	g_Akkermansia	1.00000	0.00000	0.00000
f_Turicibacteraceae	g_Turicibacter	1.00000	0.00000	0.00000

Supplemental Table 3.1. Exact sequence variant (ESV) taxonomy assignment for core measurable microbiota (CMM). 'NA' indicates taxonomic rank could not assigned.

Rank	ESV ID	Taxonomy (Phyla;Class;Order;Family;Genus;Species)
ESV	06538b77450f85903962437f449fa60a	Actinobacteria;Coriobacteriia;Coriobacteriales;Coriobacteriaceae;Adlercreutzia;NA
ESV	727992952afbbf55406b97e29855c9c2	Actinobacteria;Coriobacteriia;Coriobacteriales;Coriobacteriaceae;Adlercreutzia;NA
ESV	02408cd609a6b7f8134da2f8955136ac	Bacteroidetes;Bacteroidia;Bacteroidales;S24-7;NA;NA
ESV	029ac1f87abcac7dbac1ababc489cfcc	Bacteroidetes;Bacteroidia;Bacteroidales;S24-7;NA;NA
ESV	0ad13b6d1cd98ad7c8f6cc9909d33114	Bacteroidetes;Bacteroidia;Bacteroidales;S24-7;NA;NA
ESV	1930d2ae4018583d606e705beea31bd1	Bacteroidetes;Bacteroidia;Bacteroidales;S24-7;NA;NA
ESV	72e530a4206f2e4804d502f4dfca8387	Bacteroidetes;Bacteroidia;Bacteroidales;S24-7;NA;NA
ESV	82b6b13c800633859b139cf0569b4cda	Bacteroidetes;Bacteroidia;Bacteroidales;S24-7;NA;NA
ESV	89cb9aae28895b309fa5446ed4fa0817	Bacteroidetes;Bacteroidia;Bacteroidales;S24-7;NA;NA
ESV	bbee9f3d2ff7eae779b70c7ac2c971e4	Bacteroidetes;Bacteroidia;Bacteroidales;S24-7;NA;NA
ESV	cb8b8a6cc6bcfa115f8b24a45fa1c12a	Bacteroidetes;Bacteroidia;Bacteroidales;S24-7;NA;NA
ESV	daae43be6cf06991f62a085ba8bff3b6	Bacteroidetes;Bacteroidia;Bacteroidales;S24-7;NA;NA
ESV	d114fb4c335125128be28401522dd41a	Firmicutes;Bacilli;Lactobacillales;Streptococcaceae;Lactococcus;NA
ESV	df0e3d38eec730326754d8c17a8b8efe	Firmicutes;Bacilli;Turicibacterales;Turicibacteraceae;Turicibacter;NA
ESV	3192f0892c08ec282493637f4fa28d60	Firmicutes;Clostridia;Clostridiales;[Mogibacteriaceae];NA;NA
ESV	62dc76a09326c359be5946507e82e9fa	Firmicutes;Clostridia;Clostridiales;[Mogibacteriaceae];NA;NA
ESV	c6f4aa25b9a6b316e50c3f0308c05ab9	Firmicutes;Clostridia;Clostridiales;Christensenellaceae;NA;NA
ESV	3562d3a0374b9f2cd190c1a7aa7dcdb7	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;[Ruminococcus];gnavus
ESV	682ddc043b6d08cf20021dc1c423446d	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;[Ruminococcus];gnavus
ESV	324a68faea87d16c8e64abc6856e10e5	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Coprococcus;NA
ESV	8e74f6f63a5f9a304aeb8284be71fd23	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Coprococcus;NA
ESV	9b89b811be9292552e327f16a217ee6f	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Coprococcus;NA
ESV	b1652dc664d4e8ce2d123d26687d1204	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Coprococcus;NA
ESV	ba19eb8193a715ebfc3d44d03ec4a608	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Coprococcus;NA
ESV	c75ac6008025d80134bd14c9712f9d5c	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Coprococcus;NA
ESV	4598c7db6dff8f019ff4f6c50dc05df4	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Dorea;NA
ESV	50313555c1a4385c25c85b06a5c8ad42	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Dorea;NA
ESV	a728a1dce17d5afb4677e6cc919c2391	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;Dorea;NA
ESV	270a745647e1f4f2daefddcae78b43c7	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	31217980cfdf305668ba9c8482a34e81	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	538add871396bc31fee52b4c6e73542a	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	6536a7bc84c95500cf05185628c9f407	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	6c77545756ddc39c81c925b6fb1c8b17	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	80dd71f0135ba7b0b937f699b90d18d8	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	811f358e4e89eecbd75298ffce878cba	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	8c65d91e013525d1c8606fab40e0a88b	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	901c9973bde010c1a5c49023e63cc252	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	9e3b0ce9bb690b78f97f20c102089c55	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	9f3ae9f6096976f00cd7c13a4717ff1c	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	b0f809cbabf3ac99182581cc095868b2	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	b155d97b79acc3086cc118451afa0124	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	beb1407c19f7fa27ce3bbc0aa7f36157	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	bfa3561ca0d31bca54be08e2d9f7937f	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	c977f14c76481a059ef7ba611bb8b1d0	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	cbdc0f4fcc7b1e56aacc302c2b9b870c	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	d6eda88bd8370a52076b0dafde12249a	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	e04cb8c96d35fcc0c181a15fc4511d0c	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	f9c752d5c510e07c457e029af2d6d31b	Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;NA;NA
ESV	00cd2f6860312475904748780758 9f27	Firmicutes;Clostridia;Clostridiales;NA;NA;NA
ESV	0e064a94474c099cd422c1ef160ab28c	Firmicutes;Clostridia;Clostridiales;NA;NA;NA
ESV	1c281deaf71c6d702d1ddadfa953eac6	Firmicutes;Clostridia;Clostridiales;NA;NA;NA
ESV	5284495840f983847ec0d0b741a9a471	Firmicutes;Clostridia;Clostridiales;NA;NA;NA
ESV	54dcb911e3ab04e9d5b30c3912100b41	Firmicutes;Clostridia;Clostridiales;NA;NA;NA

Supplemental Table 3.1, *Continued*

Rank	ESV ID	Taxonomy (Phyla;Class;Order;Family;Genus;Species)
ESV	924668f542df9de35c2226c4a6cf09bd	Firmicutes;Clostridia;Clostridiales;NA;NA;NA
ESV	9b4036adb487adfe4ca484346b5b9a7b	Firmicutes;Clostridia;Clostridiales;NA;NA;NA
ESV	b65410081c87ed629211d7fb136ec45f	Firmicutes;Clostridia;Clostridiales;NA;NA;NA
ESV	ef3a40c4f26b5019887d73ceaaab84f2	Firmicutes;Clostridia;Clostridiales;NA;NA;NA
ESV	14e78619269b8c45554ab4a8ddf2f22c	Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Oscillospira;NA
ESV	20ac30224f0d6923615fbc03ac84563f	Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Oscillospira;NA
ESV	6a8118b17d1d40cf877db24449ccb616	Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Oscillospira;NA
ESV	86555947ddfc576c1bb9c42deb3998df	Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Oscillospira;NA
ESV	8da18b476b64f6f053647455b6890f80	Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Oscillospira;NA
ESV	912487364b4fda5993a03f47d60111e9	Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Oscillospira;NA
ESV	958d78a02bef69a795806f97adb117ea	Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Oscillospira;NA
ESV	cc82a5a0b284dea47583c8afa5a99755	Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Oscillospira;NA
ESV	ce78414357068359ddae6a47b6592a08	Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Oscillospira;NA
ESV	f3e9d78daeca42d345080748c31ac3dd	Firmicutes;Clostridia;Clostridiales;Ruminococcaceae;Oscillospira;NA
ESV	7b28c20c72c6c95b3c604f0849245770	Verrucomicrobia;Verrucomicrobiae;Verrucomicrobiales;Verrucomicrobiaceae; Akkermansia;muciniphila

CHAPTER 4: Conclusions and Future Directions

CONCLUSIONS AND FUTURE DIRECTIONS

Inter-individual variation in microbiota composition is associated with differential susceptibility to metabolic disease. The research presented in this thesis attempts to further understand the interplay between the host genome and gut microbiome, and how their interactions contribute to metabolic health outcomes. This work also provides the foundation for future research to elucidate the molecular mechanisms underlying genetic regulation of microbiota composition and how these microbes confer metabolic phenotypes.

Early evidence from studies using related individuals hinted at a contribution of common genetic variants in shaping the microbiome (Erwin G. Zoetendal, Antoon D. L. Ak, 2001; Turnbaugh et al., 2009a). However, these studies were limited by sample size and results were confounded by environmental factors, making it difficult to determine the relative genetic contributions. In **Chapter 2**, we interrogated the interactions among host genotype, diet, gut microbiome, and metabolic phenotypes using a panel of genetically diverse inbred mouse strains. The mouse model allowed us to control for environmental differences and circumvent confounding factors found in human studies. We used eight inbred founder strains (Churchill et al., 2012), herein referred to as “founder strains”. These founder strains include five laboratory-derived and three wild-derived strains, which together exhibit comparable genetic variation to what is found within the human population (Churchill et al., 2004). We placed the founder strains on either a control or high-fat high-sucrose (HF/HS) diet for 22 weeks. At the end of the study, we found significant variation in metabolic phenotypes and microbiota composition as a function of genotype and diet.

Using the results obtained from these strains, we next investigated whether metabolic phenotypes could be transferred via gut microbes. We took the microbiota from two strains that

exhibit disparate metabolic phenotypes, HF/HS diet-susceptible (B6) and diet-resistant (CAST) strains and colonized germ-free B6 mice. After 16 weeks on the HF/HS diet, we found significant differences in body weight and glucose homeostasis among these animals, where the CAST microbiota protected the mice from diet-induced metabolic disease. Pancreatic β-cell function also differed by microbiota composition, where the islets from CAST-colonized animals secreted significantly less insulin than islets from B6-colonized mice. Concomitant alterations in bile acid profiles were also overserved among these animals. Interestingly, bile acids are associated with metabolic disease (Kuipers et al., 2014) and are capable of stimulating β-cell insulin secretion (Düfer et al., 2012). Overall, this study identified microbiota compositions associated with metabolic phenotypes and provided novel insight into how the microbiota can influence glucose homeostasis via β-cells.

The transplant experiments with B6 and CAST microbiota present several directions for future studies. First, while we have associations between microbial taxa and host phenotypes, specific genotype-selected microbial taxa need to be identified and tested for their role in metabolic disease. We also showed that the microbiota can transfer resistance to diet-induced metabolic phenotypes in diet-sensitive genotypes (B6). To further evaluate the contributions of genotype-specific microbiota composition on susceptibility to diet-induced metabolic disease, it would be of interest to repeat the transplant experiment using a mouse strain resistant to metabolic syndrome like CAST. This would enable further characterization of bacterial communities and taxa associated with susceptibility or resistance to diet-induced metabolic disease. It would particularly interesting to examine *ex vivo* insulin secretion in germ-free CAST mice colonized with either CAST or B6 microbiota, since the CAST microbiota conferred such a strong insulin secretion phenotype.

In general, this transplant study highlighted the importance of the microbiota on insulin phenotypes and additional work needs to be done to identify specific microbial taxa and metabolites or microbial-dependent signals that affect β-cell function. Our study suggests variation in bile acid profiles may contribute to differences islet insulin secretion. There is evidence that specific bile acids stimulate insulin secretion via bile acid receptors (Düfer et al., 2012; Kumar et al., 2012). However, the effects of different bile acid profiles on insulin secretion remains relatively unknown. In addition to bile acids, other microbial derived metabolites may be responsible for changes in insulin secretion. For example, microbial production of the short chain fatty acid (SCFA) acetate was causally shown to modulate β-cell function through stimulation of the parasympathetic nervous system (Perry et al., 2016). Future studies examining microbial-driven differences in insulin secretion should take a systems biology approach and integrate metagenomic profiling of the intestinal microbiome, plasma and cecal metabolomics, and transcriptomics of the ileum and β-cells.

In **Chapter 3**, we build on the characterization of genotype-specific microbiome from Chapter 2 to identify specific host loci associated with the abundance of gut bacteria, as well as bile acid levels. We characterized the fecal microbiota composition and levels of 27 bile acid species in 400 genetically unique Diversity Outbred (DO) mice. The DO mouse stock is population used for high-resolution genetic mapping derived from an outbreeding scheme using the eight founder strains profiled in Chapter 2. The genome of each DO mouse is a mosaic of the eight founder strains and every position can be attributed to one of the founder haplotypes. Furthermore, this breeding scheme allows for high-resolution genetic mapping to better identify candidate genes (Svenson et al., 2012). This is an improvement from previous microbial QTL studies in mice,

which have limited mapping resolution and unevenly distributed genetic variation due to the intercross breeding schemes (Svenson et al., 2012; Yang et al., 2007).

We found significant variation in the gut microbiota and bile acid profiles among the DO mice. In fact, we identified loci associating with both gut microbes and bile acids, suggesting possible pleiotropy, or genomic intervals that influence multiple traits. For example, we found a locus on chromosome 8 containing the bile acid transporter *Slc10a2* was associated with the abundance of *Turicibacter sp*. and plasma cholic acid levels. We also found associations with the metabolically beneficial bacterium *Akkermansia muciniphila* (Everard et al., 2013) and plasma bile acids on chromosome 1 mapping to the same position as previously identified cholesterol QTLs (Ishimori et al., 2004; Su et al., 2009a). This is particularly striking given the association of *A. muciniphila* to obesity and cholesterol levels (Everard et al., 2013; Fu et al., 2015; Plovier et al., 2017). Thus, additional work should investigate whether cholesterol explains the relationship between this locus and *A muciniphila* and bile acids levels.

The instances of overlapping microbial and bile acid QTL are particularly interesting because they may yield important insight into host-microbe interactions. Overlapping QTL may be influenced by a pleiotropic locus, or a locus that effects multiple traits, or two closely linked loci. Additional causal inference testing can then be used to elucidate the directionality of the relationships and to determine if the underlying genetic variation effects both traits or whether one trait is causal for the other. The results of causal testing can be applied to designing validation experiments to gain mechanistic insight. For example, we used a genetics approach to identify a correlative relationship between *Turicibacter sp*. and plasma cholic acid levels. The correlative relationship between the traits suggested there may be an interaction between this microbe and bile acid. So, we designed follow-up experiments to identify a novel metabolic capability of

Turicibacter sp. and found this microbe is sensitive to high concentrations of bile acids, expanding what is known about microbe-metabolite interactions in the intestine. Even though the effect of the variant is still unclear, this approach yielded novel insight into factors that shape the gut microbiome.

Furthermore, our findings in Chapter 3 successfully replicate patterns from earlier human and mouse genetic mapping studies (Benson et al., 2010; Goodrich et al., 2016; McKnite et al., 2012; Org et al., 2015; Wang et al., 2016). While genetic variation only explains a small fraction (1% - 8.1%) of the variability in the microbiome among individuals, the congruence among these studies provides strong evidence that specific bacterial taxa are modulated by host genetics.

While there are many instances of microbial taxa associating to the mouse genome, it is important to acknowledge these associations do not equate to causation. Functional validation by additional experimental work using *in vivo* and *in vitro* approaches is required to establish causality and elucidate the molecular mechanisms underlying these associations. A major limitation of genetic mapping studies is the inability to validate candidate genes since the majority of associations are found in intergenic regulatory regions and identified loci often contain too many genes to individually test (Chen and Tian, 2016; Maurano et al., 2012). The inclusion of additional "omics" data would allow for a systems genetics approach and likely provide a deeper understanding of the underlying biology (Civelek and Lusis, 2014). The ability to assign candidate genes is vastly improved when transcriptome data is incorporated (Chick et al., 2016). Metabolomics data also provides another layer to dissect underlying mechanisms and causal pathways. Specific functional interactions likely underlie host gene – microbe associations, therefore it would likely be beneficial to incorporate metagenomics data to characterize the functional capacity of the microbiota. Taken together, the incorporation of these additional

“omics” data would greatly enhance our ability to decipher host-microbe interactions in the DO mice.

It is clear that the microbiome is a complex, high dimensional trait governed by an interplay of environmental and genetic factors. Overall, the information resulting from this thesis project enhances the understanding of the role of genetics in modulating gut microbiota and provides a framework for mechanistic studies. Ultimately, this work will help guide therapeutic strategies to manipulate the gut microbiome to treat human disease.

REFERENCES

Benson, A.K., Kelly, S.A., Legge, R., Ma, F., Low, S.J., Kim, J., Zhang, M., Oh, P.L., Nehrenberg, D., Hua, K., et al. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. Proc. Natl. Acad. Sci. U. S. A. *107*, 18933–18938.

Chen, J., and Tian, W. (2016). Explaining the disease phenotype of intergenic SNP through predicted long range regulation. Nucleic Acids Res. *44*, 8641–8654.

Chick, J.M., Munger, S.C., Simecek, P., Huttlin, E.L., Choi, K., Gatti, D.M., Raghupathy, N., Svenson, K.L., Churchill, G.A., and Gygi, S.P. (2016). Defining the consequences of genetic variation on a proteome-wide scale. Nature *534*, 500–505.

Churchill, G.A., Airey, D.C., Allayee, H., Angel, J.M., Attie, A.D., Beatty, J., Beavis, W.D., Belknap, J.K., Bennett, B., Berrettini, W., et al. (2004). The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat. Genet. *36*, 1133–1137.

Churchill, G.A., Gatti, D.M., Munger, S.C., and Svenson, K.L. (2012). The Diversity Outbred mouse population. Mamm. Genome *23*, 713–718.

Civelek, M., and Lusis, A.J. (2014). Systems genetics approaches to understand complex traits. Nat. Rev. Genet. *15*, 34–48.

Düfer, M., Hörth, K., Wagner, R., Schittenhelm, B., Prowald, S., Wagner, T.F.J., Oberwinkler, J., Lukowski, R., Gonzalez, F.J., Krippeit-Drews, P., et al. (2012). Bile acids acutely stimulate insulin secretion of mouse β-cells via farnesoid X receptor activation and K(ATP) channel inhibition. Diabetes *61*, 1479–1489.

Erwin G. Zoetendal, Antoon D. L. Ak, A.D.L.A.W.M.A.V.J.A.G.M. de V.W.M. de V. (2001). The Host Genotype Affects the Bacterial Community in the Human Gastronintestinal Tract. Microb. Ecol. Health Dis. *13*, 129–134.

Everard, A., Belzer, C., Geurts, L., Ouwerkerk, J.P., Druart, C., Bindels, L.B., Guiot, Y., Derrien, M., Muccioli, G.G., Delzenne, N.M., et al. (2013). Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. Proc. Natl. Acad. Sci. U. S. A. *110*, 9066–9071.

Fu, J., Bonder, M.J., Cenit, M.C.M.C., Tigchelaar, E.F., Maatman, A., Dekens, J.A.M.M., Brandsma, E., Marczynska, J., Imhann, F., Weersma, R.K., et al. (2015). The Gut Microbiome Contributes to a Substantial Proportion of the Variation in Blood Lipids. Circ. Res. *117*, 817–824.

Goodrich, J.K., Davenport, E.R., Beaumont, M., Jackson, M.A., Knight, R., Ober, C., Spector, T.D., Bell, J.T., Clark, A.G., and Ley, R.E. (2016). Genetic Determinants of the Gut Microbiome in UK Twins. Cell Host Microbe *19*, 731–743.

Ishimori, N., Li, R., Kelmenson, P.M., Korstanje, R., Walsh, K.A., Churchill, G.A., Forsman-Semb, K., and Paigen, B. (2004). Quantitative trait loci analysis for plasma HDL-cholesterol concentrations and atherosclerosis susceptibility between inbred mouse strains C57BL/6J and 129S1/SvImJ. Arterioscler. Thromb. Vasc. Biol. *24*, 161–166.

Kuipers, F., Bloks, V.W., and Groen, A.K. (2014). Beyond intestinal soap--bile acids in metabolic control. Nat. Rev. Endocrinol. *10*, 488–498.

Kumar, D.P., Rajagopal, S., Mahavadi, S., Mirshahi, F., Grider, J.R., Murthy, K.S., and Sanyal, A.J. (2012). Activation of transmembrane bile acid receptor TGR5 stimulates insulin secretion in pancreatic β cells. Biochem. Biophys. Res. Commun. *427*, 600–605.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

McKnite, A.M., Perez-Munoz, M.E., Lu, L., Williams, E.G., Brewer, S., Andreux, P.A., Bastiaansen, J.W.M., Wang, X., Kachman, S.D., Auwerx, J., et al. (2012). Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. PLoS One *7*, e39191.

Org, E., Parks, B.W., Joo, J.W.J., Emert, B., Schwartzman, W., Kang, E.Y., Mehrabian, M., Pan, C., Knight, R., Gunsalus, R., et al. (2015). Genetic and environmental control of host-gut microbiota interactions. Genome Res. *25*, 1558–1569.

Perry, R.J., Peng, L., Barry, N.A., Cline, G.W., Zhang, D., Cardone, R.L., Petersen, K.F., Kibbey, R.G., Goodman, A.L., and Shulman, G.I. (2016). Acetate mediates a microbiome-brain-β-cell axis to promote metabolic syndrome. Nature *534*, 213–217.

Plovier, H., Everard, A., Druart, C., Depommier, C., Van Hul, M., Geurts, L., Chilloux, J., Ottman, N., Duparc, T., Lichtenstein, L., et al. (2017). A purified membrane protein from Akkermansia muciniphila or the pasteurized bacterium improves metabolism in obese and diabetic mice. Nat. Med. *23*, 107–113.

Su, Z., Cox, A., Shen, Y., Stylianou, I.M., and Paigen, B. (2009). Farp2 and Stk25 are candidate genes for the HDL cholesterol locus on mouse chromosome 1. Arterioscler. Thromb. Vasc. Biol. *29*, 107–113.

Svenson, K.L., Gatti, D.M., Valdar, W., Welsh, C.E., Cheng, R., Chesler, E.J., Palmer, A.A., McMillan, L., and Churchill, G.A. (2012). High-resolution genetic mapping using the Mouse Diversity outbred population. Genetics *190*, 437–447.

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., et al. (2009). A core gut microbiome in obese and lean twins. Nature *457*, 480–484.

Wang, J., Thingholm, L.B., Skiecevičienė, J., Rausch, P., Kummen, M., Hov, J.R., Degenhardt, F., Heinsen, F.-A., Rühlemann, M.C., Szymczak, S., et al. (2016). Genome-wide association

analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. Nat. Genet. *48*, 1396–1406.

Yang, H., Bell, T.A., Churchill, G.A., and Pardo-Manuel de Villena, F. (2007). On the subspecific origin of the laboratory mouse. Nat. Genet. *39*, 1100–1107.

APPENDIX A: Diet-Microbiota Interactions Mediate Global Epigenetic Programming in Multiple Host Tissues

The work presented in this chapter has been published:

Krautkramer KA[*], ***Kreznar JH***, *Romano KA, Vivas EI, Barrett-Wilt GA, Rabaglia ME, Keller MP, Attie AD, Rey FE, and Denu JM*[+]

[*] *indicates lead author,* [+] *indicates corresponding author*

Supplemental methods and results available online.

ABSTRACT

Histone-modifying enzymes regulate transcription and are sensitive to availability of endogenous small molecule metabolites, allowing chromatin to respond to changes in environment. The gut microbiota produces a myriad of metabolites that affect host physiology and susceptibility to disease, however the underlying molecular events remain largely unknown. Here we demonstrate that microbial colonization regulates global histone acetylation and methylation in multiple host tissues in a diet-dependent manner: consumption of a “Western-type” diet prevents many of the microbiota-dependent chromatin changes that occur in a polysaccharide rich diet. Finally, we demonstrate that supplementation of germ-free mice with short-chain fatty acids, major products of gut bacterial fermentation, is sufficient to recapitulate chromatin modification states and transcriptional responses of colonization on host epigenetic programming. These findings have profound implications for understanding the complex functional interactions between diet, gut microbiota, and host health.

INTRODUCTION

The eukaryotic genome is organized into a highly compressed nucleoprotein structure known as chromatin. Histone proteins are major components of chromatin and act as spools, wrapping DNA into fundamental nucleosome units that can fold into higher order structures. Histones undergo a myriad of covalent post-translational modifications (PTMs), and along with histone variant replacement, comprise what is known as the "histone code," wherein the local PTM-state dictates whether chromatin is repressive or activating toward transcription (Jenuwein and Allis, 2001). For example, histone acetylation is generally associated with open chromatin and active transcription, whereas trimethylation of histone H3 K27 (H3K27me3) is associated with transcriptional repression (Cao et al., 2002; Lachner et al., 2001). Histone PTMs exist in a combinatorial manner and can serve as a signal integration platform, sensing changes in environment and allowing for adaptive responses (Johnson and Dent, 2013). Importantly, the enzymes that add and remove PTMs are known to be exquisitely sensitive to the availability of endogenous small molecule metabolites (Fan et al., 2015). For example, acetyl-coenzyme A (acetyl-CoA) is a necessary substrate for histone acetyltransferases (HATs), and increased availability results in increased HAT activity.

The gut microbiota produces a variety of metabolites that are present at detectable levels in host circulation (Wikoff et al., 2009), including small organic acids, bile acids, vitamins, choline metabolites, and lipids (thoroughly reviewed in (Nicholson et al., 2012)). Dietary poly- and oligosaccharides that are resistant to digestion by the mammalian host's limited repertoire of carbohydrate active enzymes are not broken down and absorbed in the small intestine, but rather pass to the distal gut where they serve as a source of carbon and energy for gut bacteria. Through fermentative reactions, the gut microbiota can metabolize

these complex carbohydrates to produce small organic acids, the majority of which are comprised of the short chain fatty acids (SCFAs) acetate, propionate, and butyrate (≥95%) (Besten et al., 2013). These metabolites are present in the proximal gut lumen at roughly 70-140 mM total concentration, depending on dietary substrate (Topping and Clifton, 2001). A number of robust associations between gut microbiota and host metabolic outcomes have been made in recent years, including cardiovascular disease (Karlsson et al., 2012; Wang et al., 2011), metabolic syndrome (Cabreiro et al., 2013; Chassaing et al., 2015; Suez et al., 2014; Vijay-Kumar et al., 2010), obesity (Bäckhed et al., 2004; Ley et al., 2005; Ridaura et al., 2013; Turnbaugh et al., 2006; Zhao, 2013), diabetes mellitus (Amar et al., 2011; Qin et al., 2012), non-alcoholic fatty liver disease (Henao-Mejia et al., 2012), hepatic steatosis (Singh et al., 2015), and even inflammatory bowel disorders and malignancy (Chassaing et al., 2015; Donohoe et al., 2012; 2014). Interestingly, SCFAs have recently been associated with both disease promoting (Belcheva et al., 2014; Perry et al., 2016; Samuel et al., 2008; Singh et al., 2015; Turnbaugh et al., 2006) and therapeutic effects (Canfora et al., 2015; Donohoe et al., 2014; Tan et al., 2014; Tilg and Moschen, 2014), prompting a need for increased understanding of the underlying molecular mechanisms.

It is possible that gut microbial metabolites, which are present at detectable levels in host circulation, play regulatory roles at the level of host chromatin. For example, acetate, propionate, and butyrate can be converted to acetyl-CoA, thereby activating HAT activity. Further, butyrate is a known histone deacetylase inhibitor (HDACi) (Riggs et al., 1977). Therefore, these small organic acids may increase histone acetylation. Histone methylation may also be affected by microbial metabolites. Gut bacteria produce a number of B-vitamins (Hill, 1997) including B2 (riboflavin), B9 (folate), and B12 (cobalamin), which may promote

availability of the methyl-donor S-adenosyl methionine (SAM), increasing histone methylation. The gut microbiota may also compete with the host for choline (Romano et al., 2015), ultimately limiting SAM availability. Therefore, there are a number of ways in which gut microbial metabolites may impact the host epigenome, however experimental evidence linking changes in the abundance of these metabolites to global variations in histone PTMs *in-vivo* is lacking.

Here, we explore whether the gut microbiota affects host epigenetic programming in a variety of tissues and how this relationship is affected by host diet. We provide the first evidence of gut microbiota- mediated changes in global histone acetylation and methylation not only in colon, which is in direct contact with microbes and their metabolites, but in tissues outside the gut. We demonstrate that this regulatory relationship is sensitive to host diet, wherein a “Western-type” diet limits microbial SCFA production, abolishes the effects of microbiota on host chromatin states, and results in functionally relevant alterations in hepatic gene expression. Finally, we identify an underlying mechanism that reveals SCFA supplementation of germ-free mice is sufficient to recapitulate the epigenetic phenotype associated with gut colonization.

RESULTS

Gut microbiota affect host tissue epigenetic states

To investigate whether gut microbes and their metabolites affect host chromatin states, we examined histone PTM states as a function of colonization. We focused our analysis on proximal colon, liver, and white adipose. The proximal colon harbors the largest microbial community in the body and is a region exposed to the highest levels of microbial metabolites. Nearly all metabolite-rich venous blood that drains from the gut enters the liver via the convergence of several mesenteric veins into the hepatic portal vein. Given this anatomical feature and known associations between gut microbiota and hepatic steatosis, liver tissue was a logical choice to investigate the effects of bacterial metabolites. Lastly, we selected white adipose tissue (WAT) as a tissue more distant from the gut that is known to be affected by gut microbial colonization (Bäckhed et al., 2004). The experimental workflow is described in figure 1A. Briefly, mice were either maintained germ- free (GF) throughout the experiment, allowed to acquire a microbiota from birth to adulthood (conventionally raised, ConvR), or colonized with a complete (uncultured) microbiota (conventionalized, ConvD) harvested from ConvR donors. The use of a ConvD mouse model allows for the determination of whether the phenotype observed in ConvR animals is transferrable via the gut microbiota alone. Additionally, since ConvR animals experience different environmental exposure early in life and have developmental differences (K. Smith et al., 2007) that may exhibit phenotypic differences *vs.* their GF controls, the use of ConvD mice for relatively short time periods allows for dissection of effects more directly related to differences in microbial metabolism. At the time of sacrifice, tissues were harvested, and histones were extracted and prepared for mass spectrometry analysis using an in-house data independent acquisition mass spectrometry workflow (Krautkramer et al., 2015).

We surveyed 55 unique and combinatorial acetylated and methylated histone PTM states in proximalcolon, liver, and WAT (Supp. Table 1). Colonization induced robust increases in histone H4 acetylation in all three tissues (Fig. A.1B). This peptide includes the first 4 lysines (K5, K8, K12, and K16) on the histone H4 N-terminal tail. Thus H4: 0ac indicates peptides where no lysine residues are acetylated, whereas H4:1ac-4ac indicates peptides where any 1-4 of the 4 lysines are acetylated. In ConvR animals, there was a significant 2.1-fold increase in both triply and quadruply acetylated histone H4 (H4: 3ac and H4:4ac, respectively). Similarly, ConvR animals showed a 3-fold and 1.30-fold increase in the highly acetylated H4: 4ac peptide in proximal colon and adipose tissue, respectively, relative to GF mice (Fig. A.1B). The effects of colonization on histone H4 acetylation were even more robust in ConvD mice, with a 4.5-, 6.0-, and 12.0-fold increase in H4: 4ac of proximal colon, adipose, and liver, respectively (Fig. A.1B). Triply acetylated histone H4 peptides also increased 2.1- to 4.2-fold in proximal colon, adipose, and liver (Fig. A.1B). It is noteworthy that these two histone H4 states collectively account for just over 1% of the total histone H4 states, suggesting that this open chromatin state is confined to very specific loci along the genome. As acetylated forms of the H4 N-terminal tail increased in abundance in colonized animal tissues, the completely unmodified form of this peptide decreased significantly (H4: 0ac, 1.33 to 3.3-fold across the tissues surveyed), consistent with the conversion of unacetylated states to higher acetylation content in colonized animals.

Microbes also induced acetylation of canonical histone H3 and the variant histone H3.3. Similar to the patterns observed on histone H4, there were significant increases in acetylation in response to gut colonization. The doubly acetylated canonical histone H3 K9ac+K14ac and H3 K18ac+K23ac peptides increased significantly in ConvD mouse livers

and trended toward a similar magnitude of increase in proximal colon of ConvR and ConvD mice (Fig. A.1B). Similar to highly acetylated histone H4 peptides, these two doubly acetylated canonical histone H3 peptides account for roughly 2% or less of total histone PTM states observed in each peptide family, again supporting a more loci-specific role for these modified nucleosomes along the genome (Fig. A.1B). Interestingly, the singly acetylated peptides K9ac+K14un and K9un+K14ac decrease concomitantly with an increase in the doubly acetylated K9ac + K14ac peptide, and a similar pattern occurs on the singly acetylated and coeluting K18ac/K23ac peptides (Fig. A.1B). These results are consistent with a shift away from a singly acetylated state toward a maximally acetylated state.

Histone H3 methylation patterns are also altered as a function of gut colonization status. There is a modest, yet statistically significant increase in histone peptide H3 K27me3+K36un in proximal colon, liver, and adipose tissues from ConvD mice vs. their GF controls (1.4- to 1.5-fold increase, Fig. A.1B). This modification accounts for ~12% of total PTM states, suggesting more broad regulatory effects in comparison to highly acetylated states of histones H3 and H4 which accounted for only a very small fraction of total chromatin. There were significant increases in peptides containing highly methylated forms of K27 and K36 (i.e. me2 and me3) on both the canonical histone H3 and the histone variant H3.3, however these effects were not present consistently across all three tissues, suggesting some tissue- specificity in the response to colonization (Fig. A.1B). Notably, histone monomethylation at H3 K18 decreased significantly across all three ConvD tissues (Fig. A.1B, 2.1 to 8.3-fold across liver, proximal colon, and adipose tissue). A similar pattern was present for the combinatorial K27me2+K36me1 peptide on both histone H3 and the variant H3.3 (Fig. A.1B). Together, these results demonstrate that gut microbiota affect host tissue

acetylated and methylated chromatin states in a site-specific and combinatorial fashion, strongly supporting a role for the gut microbiota as a driver of host tissue chromatin regulation. Finally, while some histone PTM states appear to be similarly regulated across all tissues surveyed, other changes are unique within a tissue.

A key feature of our mass spectrometry data is the ability to detect histone PTMs within the context of neighboring modifiable sites on the same peptide, thereby capturing some of the combinatorial nature of the histone code and allowing for detection of histone PTM states that account for a very small percentage of the total. Thus, our quantitative mass spectrometry results identified changes in histone PTM states that are not necessarily resolvable by orthologous techniques such as western blot analysis (Aebersold et al., 2013). Indeed, we performed several western blots and found no statistically significant differences in histone H3 K9ac, H3 K27me3, or pan-acetyl(K) detectable via western blot in histone extracts from GF, ConvR, and ConvD mouse livers (Supplemental Fig. 1A-B). To more directly compare our mass spectrometry data with that obtained via western blot, we summed all possible permutations of peptides containing K9ac and K27me3 on canonical histone H3 to obtain single western blot-like estimates of K9ac and K27me3 abundance for GF, ConvR, and ConvD mouse liver histone PTM states (equations available in *Supplemental Methods*). We then calculated fold changes relative to GF. Using this method, both K9ac and K27me3, as a total among all combinatorial forms quantified, are predicted to remain relatively unchanged between colonized and GF mice (Supplemental Fig. 1C), which is consistent with quantitative western blot results (Supplemental Fig. 1B). It is also noteworthy that robust changes in histone PTM states, including triply and quadruply acetylated histone H4 peptides, cannot be probed using antibodies, since there are no antibodies currently available that can

specifically detect these combinatorial states. Therefore, our histone proteomics workflow allowed for identification of altered chromatin states that would not otherwise have been detectable.

Gut microbiota-mediated changes in chromatin state are sensitive to host diet

Since host diet is known to affect both gut microbial community composition and metabolism (Daniel et al., 2013; David et al., 2013; Turnbaugh et al., 2009b), we next evaluated the effects of host diet on microbiota-mediated regulation of host chromatin states (Supp. Table 1). ConvR and GF mice were fed a "Western-type" high fat, high sucrose diet (HF/HS), which is provides low levels of fermentable substrate for the gut microbiota, for 16 weeks prior to sacrifice at 19 weeks of age. At the time of sacrifice, tissues were harvested and a number of physiological parameters and histone PTM states were measured as described in the workflow (Fig. A.2A). As anticipated, ConvR mice fed a HF/HS diet weighed significantly more than diet-matched GF controls (Fig. A.2B). HF/HS-fed ConvR mice also displayed significantly higher hepatic total cholesterol and triglycerides *vs.* diet-matched GF controls and chow-fed mice (Fig. A.2C-D). Thus, HF/HS feeding significantly impacted host metabolic state in a microbiota-dependent manner. To determine whether HF/HS feeding altered SCFA production, we measured acetate, propionate and butyrate concentrations in cecal contents (i.e. at the principal site of fermentation) of ConvR, ConvD, and GF mice on both chow and HF/HS diets. SCFAs are mostly derived from microbial fermentation of complex polysaccharides, thus we anticipated differences in SCFA production in response to altered diet composition. Indeed, gut microbial colonization resulted in an increase of these metabolites in the ceca of mice, and this increase was more pronounced in mice fed a chow

diet compared to HF/HS diet (Fig. A.2E). In HF/HS-fed ConvR mice, cecal acetate, propionate, and butyrate were present at significantly lower levels relative to chow-fed ConvR mice (1.9-6.0 fold, Fig. A.4A, Supplemental Fig. 5A), in accordance with the lack of microbiota-accessible carbohydrates in this diet. Interestingly, chow-fed ConvD mice had greater cecal SCFAs than ConvR mice (Fig. A.2E). This pattern in cecal SCFA levels is consistent with that of histone PTM changes in ConvD and ConvR mice on chow, wherein PTM states trend in the same direction, but the magnitude of change is larger in ConvD mice *vs.* ConvR (Fig. A.1B). This suggests that SCFA availability influences histone PTM states. In peripheral venous blood, levels of these SCFAs were unchanged (Supplemental Fig. 5B), which is consistent with these organic acids having already undergone significant metabolism in the liver prior to reaching peripheral venous blood.

As predicted by cecal SCFA data, the gut microbiota-host epigenome relationship was significantly altered in response to HF/HS feeding. While there was a microbiota-dependent increase in histone H4 acetylation in ConvR and ConvD tissues of chow-fed mice (Fig. A.1B), HF/HS-feeding abolished the effects of gut colonization in liver and WAT (Fig. A.2F and A.2G). Interestingly, the microbiota-dependent effects on histone H4 acetylation were attenuated, but still significantly increased relative to GF controls in HF/HS-fed mouse proximal colon (Fig. A.2H). This pattern of diet-dependence was also present in other histone PTM states. The response to gut microbiota on histone H3 K18 and K23 also trended as a function of diet: K18me1 and K23me1 peptides both decreased significantly in livers of both ConvR and ConvD chow fed mice, but remained unchanged in HF/HS-fed mice (Fig. A.2I). The coeluting peptides K18ac and K23ac (i.e. K18ac/K23ac) were unchanged in response to gut microbiota in livers of chow-fed mice, yet decreased significantly in HF/HS-fed mouse

livers (Fig. A.2I). However, in proximal colon this diet-dependency was again absent and histone PTM states trended in similar directions regardless of dietary conditions (Fig. A.2J). There was also a diet-dependent, microbiota-independent increase in basal histone H4 acetylation in liver, proximal colon, and adipose tissue. In comparison to histone H4 acetylation in tissues from chow-fed ConvR and GF animals, there were significant increases in nearly all forms of 1ac-4ac histone H4 peptides in HF/HS-fed tissues ranging from 1.2 to 7.2-fold (Supplemental Fig. 6).

Although the direction of change in both acetylated and methylated histone PTM states generally remained similar in ConvR and ConvD tissues, there were differences in the magnitude of PTM changes (Fig. A.1B, A.2F-J) that trended with differences in cecal SCFA levels (Fig. A.2E). To investigate whether these differences reflected alterations to microbial community composition, we performed 16S rRNA sequencing. Principal Coordinates Analysis (PCoA) of weighted UniFrac distances revealed that the microbial community composition of ConvR and ConvD mice on a chow diet was more similar to each other than it was to the microbial community from mice fed HF/HS diet (Supplemental Fig. 2A). ConvR mice fed a HF/HS diet had significantly fewer Bacteroidetes and a greater abundance of Firmicutes (Supplemental Figures 2B-C) than chow fed ConvR and ConvD mice. This is consistent with previous observations that diet and obesity alter the ratio of Bacteroidetes to Firmicutes in the gut (Ley et al., 2005; Turnbaugh et al., 2006). Furthermore, the relative abundance of these two major phyla in chow-fed ConvD mice were intermediate between chow- and HF/HS-fed ConvR mice (Supplemental Figures 2B- D). Together, these data suggest that both gut microbial community composition and metabolite production, which are

inherently linked, are important factors that mediate the relationship between gut microbiota and regulation of host chromatin states.

Functional impact: Co-regulation of hepatic genes associates with altered chromatin states

To assess whether microbiota-dependent changes in host chromatin state affected tissue gene expression, we performed RNA-seq analyses on livers of colonized (both ConvR and ConvD) and GF mice on the two diets described above. Each group of colonized mice was compared to its diet-matched GF control: i.e., chow-fed ConvR *vs.* GF, chow-fed ConvD *vs.* GF, and HF/HS-fed ConvR *vs.* GF. A total of 623 genes were differentially expressed (DE) among these three groups, as determined by an FDR cut-off of 0.05 (Supp. Table 2). K-means clustering of hepatic DE genes revealed 6 optimal clusters, each enriched for unique biological pathways (Fig. A.3A, Supp. Table 3). When comparing ConvR mice on either chow or HF/HS diets to their respective GF controls, cluster 2 contained genes that are co-regulated as a function of both diet and microbiota (Fig. A.3A-B). This group of genes was enriched for processes involved in insulin, SREBP, and PPAR signaling, and adaptive immunity. Additionally, this cluster contains a number of genes that may regulate histone PTM states via modulation of small molecule metabolite availability. Clusters 4 and 6 contain genes whose expression patterns differ in ConvR animals as a function of diet (Fig. A.3A, C-D). Cluster 4 genes are enriched for pathways involved in cholesterol, retinol, and amino acid metabolism as well as host immunity, whereas cluster 6 contains a number of genes involved in lipid and amino acid metabolism as well as a group of genes involved in regulation of folate, which ultimately affects the availability of the one-carbon donor SAM to histone methyltransferases. It is noteworthy that a significant proportion of DE genes (4% of DE genes) are known hepatic targets of SREBP and PPAR.

To specifically assess the effects of diet alone, we made comparisons in GF and ConvR mice as a function of diet. Thus, rather than comparing each group of mice to their respective GF control, mice were compared within colonization groups (i.e. GF or ConvR) across dietary conditions (i.e. HF/HS *vs.* chow). This comparison yielded 868 differentially expressed genes, 413 of which increased and 455 of which decreased in expression in HF/HS-fed mice relative to chow-fed controls (Supplemental Fig. 4A- B). Although a fraction of DE up (Supplemental Fig. 4A) and DE down (Supplemental Fig. 4B) genes are regulated in both ConvR and GF mice, the fact that there are 1.8-fold more total DE genes in response to diet in ConvR mice vs. GF mice suggests that gut microbiota drive a significant portion of the response to HF/HS feeding in liver. Additionally, KEGG pathway analysis of unique and overlapping DE genes in GF and ConvR mice revealed several oppositely regulated pathways as a function of diet (Supplemental Figures 4C-D). For example, while starch and sucrose metabolism is enriched in DE up genes of ConvR mice, this same pathway is significantly enriched in DE down genes of GF mice. The same pattern is present for pathways involved in arachidonic acid metabolism, cytokine-cytokine receptor interaction, and endocytosis. Other key differences include DE up gene enrichment for processes involved in the TCA cycle and propionate metabolism in GF mice only, and butyrate metabolism in ConvR mice only (Supplemental Fig. 4C). Importantly, genes involved in PPAR signaling, insulin signaling, and diabetes mellitus were enriched in DE up genes shared by both GF and ConvR mice (Supplemental Fig. 4C). Pathways involved in de-novo cholesterol synthesis are significantly enriched in DE down genes shared by both GF and ConvR mice, suggesting that HF/HS feeding decreases host de-novo cholesterol synthesis irrespective of gut colonization status.

Thus, there are a number of DE genes that associate with altered chromatin states in liver whose expression is mediated by this diet-microbiota interaction. More detailed information about specific gene cluster membership, expression, and pathway enrichment among clusters is described in *Supplemental Tables 2-4, Supplemental Figures 3-4,* and the *Supplemental Information*.

SCFA-supplementation partially phenocopies the effects of colonization on host epigenetic programming

We hypothesized that the SCFAs acetate, propionate, and butyrate were key mediators of systemic microbiota-induced changes in host chromatin states. To further investigate this idea, we supplemented germ-free mice with acetate, propionate, and butyrate (GF+SCFA) and harvested liver and proximal colon for histone PTM and gene expression analyses (Fig. A.4A). These GF+SCFA mice were then compared to GF and ConvD mice, as negative and positive controls, respectively. Hierarchical clustering revealed that histone PTM states of GF+SCFA and ConvD mice were strikingly similar across both acetylated and methylated peptides (Fig. A.4B). Further, GF+SCFA and ConvD histone PTM signatures were highly correlated, with a Pearson's correlation coefficient of 0.74 - 0.75, for proximal colon and liver samples ($p = 1.2 \times 10^{-10}$ and 5.7×10^{-11}, respectively; Fig. A.4C-D). These data reveal that the global chromatin states induced by SCFAs mimic, in part, gut colonization.

While levels of acetate, propionate, and butyrate were significantly increased in the cecal contents of ConvD mice, there were no significant differences between GF+SCFA mice and GF controls (Supplemental Fig. 5C). Similar to previous observations (Supplemental Fig. 5B), peripheral blood levels of SCFAs were unchanged as a function of either gut colonization or SCFA-supplementation (Supplemental Fig. 5D). Thus, while conventionalization results in

increased local concentration of cecal SCFAs, oral supplementation of SCFAs did not, as these organic acids are expected to be absorbed to some extent across earlier segments of the alimentary tract. Additionally, each bolus of ingested SCFA in GF+SCFA mice is likely to be absorbed very quickly across the gut epithelium, which may result in differential luminal accumulation *vs.* constant local production by the gut microbiota. Despite these potential differences, the physiological impacts are likely to be similar, given that venous blood from the small intestine also drains to the liver via the hepatic portal vein where it mixes with blood traveling from the colon.

To determine whether these highly similar global chromatin states elicited similar biological effects in ConvD and GF+SCFA mice, we used RNAseq analyses to examine hepatic gene expression in ConvD, GF+SCFA, and GF mice. Consistent with histone PTM observations, GF+SCFA mice had highly similar transcriptional profiles to ConvD mice (Fig. A.4E & A.4G). K-means clustering of 537 DE genes revealed 6 clusters of co-regulated genes that were enriched for a number of metabolic and immunological processes (Fig. A.4E-F, Supp. Tables 5-6). In particular, clusters *b* and *c* were enriched for GO-terms involved in immunity (cluster *b*) and regulation, storage, and metabolism of lipids and cholesterol (cluster *c*, Supp. Table 5). Finally, there was striking overlap of DE genes between GF+SCFA and ConvD mice, wherein >50% of DE genes in GF+SCFA livers overlapped with DE genes in ConvD mice (Fig. A.4G). Together these data suggest that SCFAs are partially causative metabolites in the complex regulatory relationship between diet, gut microbiota, and host tissue epigenetic programming.

DISCUSSION

Eukaryotic histone modifying enzymes have evolved to sense and integrate environmental signals, ultimately programming gene expression patterns and mediating phenotype. Given the sensitivity of these enzymes to endogenous small molecule metabolites, we hypothesized gut microbial metabolites absorbed and metabolized by the host may exert similar control. Here, we present the first evidence that global histone acetylation and methylation are mediated by gut microbiota in multiple host tissues, not limited solely to the gut itself.

Gut microbiota-diet interactions influence host chromatin states

Gut colonization drives robust increases in acetylation of histones H3 (K9, K14, K18, and K23) and H4 (K5, K8, K12, and K16) in a diet- and tissue-dependent manner (Fig. A.1B). While there were increases in histone acetylation in liver and adipose tissue from colonized chow-fed animals, the effects of gut colonization were lost in animals fed a HF/HS diet (Fig. A.2F-H). Interestingly, the diet-dependency of microbiota-driven changes in histone H4 acetylation was attenuated in proximal colon. Although HF/HS-feeding diminished the magnitude of change in H4 acetylation, the change was not abolished (Fig A.2H). This difference in tissue-specific response may be due to changes in availability of SCFAs. Production of SCFAs is reduced 1.9 to 6-fold in HF/HS-fed animals *vs.* their colonized counterparts (Fig. A.2E, Supplemental Fig. 5A). These dietary effects were expected, given that HF/HS-feeding is known to reduce gut microbial biomass and significantly alter both community composition and metabolite production (Daniel et al., 2014). Therefore, while SCFAs may still be present at high enough concentration to affect histone modifying enzymes in the gut of HF/HS-fed animals, the amount available in peripheral circulation may not be

sufficient to elicit effects in distal tissues in HF/HS-fed mice. It is worth noting that although there was no effect of microbiota on histone H4 acetylation in tissues of HF/HS-fed mice, HF/HS-feeding alone induced an increase in baseline histone H4 acetylation (Supplemental Fig. 6). The nature of this increase is unclear, but could result from increased production of acetyl-CoA through beta-oxidation of dietary lipids or due to elevated lipid-based signaling pathways in the HF/HS diet. While it may be possible that this basal increase in histone acetylation masked microbiota-driven changes, the fact that acetylation is increased still further in proximal colon, where SCFAs are generated, under HF/HS-feeding suggests that the loss of effect seen in distal tissues is due to limiting amounts of bacterial SCFA rather than an elevated baseline.

SCFAs play a dual role as both substrates for metabolism and as signaling molecules (Besten et al., 2013). Mice and humans derive ~10% of their energy from oxidation of bacterial SCFAs. Colonocytes obtain 60-70% of their energy from oxidation of SCFAs. The remaining SCFAs drain from the gut via the superior and inferior mesenteric veins, which converge and empty into the liver via the hepatic portal vein. As much as 70% of acetate and roughly 30% of propionate is taken up by the liver, where they both serve as sources of energy. Acetate can also serve as a substrate for cholesterol, long-chain fatty acid, glutamine, and glutamate synthesis in the liver. The remainder is metabolized by other tissues, including white adipose. As ligands of the G-protein coupled receptors FFAR2 (GPR43) and FFAR3 (GPR41), SCFAs play a role in lipid and glucose metabolism. Thus, these metabolites play complex roles in regulation of host metabolic phenotype. Whether SCFAs contribute to beneficial or pathogenic effects in the host remains unclear. While SCFAs have been associated with anti-inflammatory effects, improvement of insulin sensitivity, glucose

homeostasis, and weight control; and protection from colorectal cancer (Besten et al., 2013; Canfora et al., 2015; Donohoe et al., 2012; 2014; P. M. Smith et al., 2013), they have also conversely been associated with increased capacity for energy harvest, inflammation, hepatic steatosis, and the promotion of colorectal cancer (Belcheva et al., 2014; Singh et al., 2015; Turnbaugh et al., 2006). Thus, further investigation is needed to reveal the complete set of molecular mechanisms underlying SCFA-associated phenotypes.

Here, we show that SCFA supplementation of germ-free mice is sufficient to recapitulate many of the epigenetic effects of gut colonization (Fig. A.4A-G). Interestingly, SCFA supplementation mimics effects of colonization on both histone acetylation and methylation, to the extent that collective histone PTM states in liver and proximal colon of GF+SCFA and ConvD mice have highly significant Pearson's correlation values of 0.74-0.75 (Fig. A.4B-D). While the link between histone acetylation and SCFAs is more clear, the link between SCFAs and histone methylation is less clear. Treatment with SCFAs in a cell culture system has been shown to suppress expression of the histone methyltransferases EZH2 and SUV39H1and decrease of the repressive histone methylation modifications H3 K9me3 and H3 K27me3 (Yu et al., 2014), however the underlying mechanisms are not currently understood. Regulation of histone methylation may be due to either antagonism by acetylation at the same site (Pasini et al., 2010) or regulation by combinatorial effects of acetylated states at nearby sites. While the effects of SCFA- supplementation broadly recapitulated both chromatin states and gene expression patterns of ConvD mice, it is worth noting that the magnitude of the effects were generally less in SCFA+GF tissues relative to ConvD (Fig. A.4B, A.4E). This suggests that while SCFAs are at least partially causative, there are likely other bacterial metabolites that regulate the host epigenome.

The increases observed in histone acetylation generally accounted for very small mean percentages of total global chromatin states (≤2.1% of the total across all conditions for highly acetylated peptides containing H3 K9, K14, K18, and K23 and H4 K5, K8, K12, and K16). In contrast, H3 K27me3, which increased by 1.4- to 1.5-fold in ConvD tissues relative to GF controls, accounted for a mean of nearly 12% of the global peptide family total across all conditions. Therefore, it is possible that gut microbiota drive active chromatin states at very specific loci in the genome (given the small % of global peptide family totals), whereas the microbiota-induced repressive chromatin states are more broadly distributed across the genome (given their larger % of total global peptide family states).

While the direction of change remained highly similar across PTM states in ConvR and ConvD tissues, the magnitude of change was often greater in ConvD mice *vs.* ConvR (Fig. A.1B). Notably, this trend in PTM states mirrored that of cecal SCFA contents (Fig. A.2E), which supports a role for SCFAs as key metabolites in the microbiota-host epigenome response. This is somewhat unexpected but is likely due, at least in part, to differences in microbial community composition between ConvR and ConvD mice. This variance may be due to loss of key taxa during colonization of GF mice and/or differences in the ability of specific taxa to compete in a GF environment. It is also possible that differences in the host tissue chromatin response are due to the fact that ConvD animals were colonized only one month prior to sacrifice, whereas ConvR animals were allowed to acquire a microbiota from birth onward. In this scenario, the increased magnitude of change in histone PTM states of ConvD mice vs. their ConvR counterparts may be due to the fact that ConvD hosts are still metabolically adapting to a previously inaccessible source of nutrition via the gut microbiota.

Hepatic gene co-regulation is linked to altered chromatin states

Altered histone PTM states associated with differential expression of genes involved in metabolic homeostasis and immune regulation in both colonized and SCFA-supplemented mice (Fig. A.3A-D, Fig. A.4F-H). A surprisingly large number of hepatic DE genes were involved in either glucose or lipid and cholesterol homeostasis, which is underscored by significant pathway enrichment for insulin and PPAR signaling (Fig. A.3, Supplemental Information). The fact that genes involved in insulin signaling were downregulated in both chow and HF/HS fed mice *vs.* their GF controls suggests that insulin sensitivity may be decreased in colonized mice, irrespective of diet. Interestingly, hepatic *Scd1* expression was significantly increased in colonized mice relative to their GF controls, and the magnitude of change was greater in chow- *vs.* HF/HS fed mice (Fig. 3B). Expression of *Scd1* was recently shown to be driven by gut microbiota in a SCFA-dependent manner and to contribute to the development of metabolic syndrome and increased *de novo* hepatic lipogenesis in toll-like receptor 5 knockout (T5KO) mice (Singh et al., 2015). Given this, it is unclear why the microbiota-driven increase in *Scd1* expression is greater in mice fed a standard chow diet than in those fed a diet more permissive to the development of metabolic syndrome (HF/HS), but the increased response in chow-fed mice may simply be due to increased SCFA availability relative to their HF/HS-fed counterparts. Further, T5KO mice are prone to the development of metabolic syndrome, thus, their metabolic response to microbiota may differ from that of a wild type mouse (Singh et al., 2015).

The gut microbiota also altered expression of a number of genes involved in availability of small molecule metabolites that are known to regulate histone PTM addition or removal. One prime example is *ATP citrate lyase* (*Acly*). Under both chow and HF/HS

feeding, *Acly* expression decreased in ConvR *vs.* GF mice. This enzyme has been demonstrated to be essential for glucose-driven histone acetylation in mammalian cells, however acetate-driven histone acetylation was not affected by knockdown of *Acly* expression (Wellen et al., 2009). Interestingly, *Acly* expression decreases most in HF/HS-fed ConvR mouse livers. This begs the question of whether *Acly* expression is decreased in the setting of decreased tissue glucose-dependency due to the presence of other sources of carbon and energy, such as bacterial SCFAs or highly energetic lipids from HF/HS-feeding. Availability of NAD^+, a necessary co-substrate for Class III HDACs, may also be modulated by gut microbiota. *NAMPT,* which catalyzes the rate-limiting step in NAD^+ biosynthesis, was differentially expressed in colonized vs. GF livers. Finally, there were four genes in cluster 6 linked to regulation of folate, which can affect the availability of the methyl donor SAM for histone methyltransferases: *Sardh, Dmgdh*, *Amt,* and *Gldc* (Fig. A.3D). Loss of *Gldc*, the first enzyme of the glycine cleavage system that breaks down glycine to produce one carbon formate, has been associated with neural tube defects, growth retardation, and decreased levels of one carbon-carrying folates in tissue that can be rescued by supplementation with formate (Pai et al., 2015). Thus, decreased expression of these enzymes may affect tissue SAM availability, via modulation of folate levels.

Finally, a number of genes associated with host immunity and inflammation were differentially expressed in colonized as well as GF+SCFA mice relative to their GF controls (Fig. 3A-D, 4F-G). Given that GF mice are completely naive to microbiota, it is not unexpected that colonized mice have altered expression of immunomodulatory genes relative to GF mice. Further, it is known that gut microbiota are important for the development and function of host immunity (Belkaid and Hand, 2014). SCFAs have been previously shown to

play an important role in both adaptive and innate immunity, in FFAR2- and FFAR3-dependent manners (Belkaid and Hand, 2014; Correa et al., 2016; P. M. Smith et al., 2013). Whether the multitude of DE metabolic and immunomodulatory genes observed here are directly regulated by histone PTMs remains to be determined. However, it is noteworthy that there are a number of regulated genes that have no known association with FFAR2 or FFAR3 signaling. The close association between histone PTM states and gene expression patterns, particularly in GF+SCFA mice, further supports a role for histone PTMs in the response to microbiota. Finally, gene activation via FFAR2 and FFAR3 signaling need not be mutually exclusive with concurrent alterations in histone PTM states. Further work will be required to elucidate which genes are directly regulated by microbiota- mediated histone PTM states, however the work presented here offers an extensive resource that will be invaluable for future exploration of this nature.

CONCLUSIONS

The gut microbiota is a diverse, metabolically rich community that has co-evolved with its mammalian hosts (Ley et al., 2008). The impact of this metabolic and immune regulatory organ on the host has become increasingly evident in recent years. Here, we demonstrate that gut microbiota exert a regulatory role on both methylated and acetylated host chromatin states in multiple organ systems that is at least partially driven by microbial SCFAs. These SCFAs can be either directly converted (acetate) or oxidized (propionate and butyrate) to acetyl-CoA, the substrate for HAT enzymes. Further, butyrate is a known HDAC inhibitor. Both scenarios result in increased histone acetylation. Consequently, it is possible that eukaryotic histone-modifying enzymes have evolved to “sense” not only endogenous small molecule metabolites, but also those produced by commensal microbiota. In this manner, the host epigenetic machinery guides phenotype in response to altered metabolic states, such as increased availability of SCFAs, driving specific gene responses. While robust associations between the gut microbiota and the host in health and disease have been demonstrated, in many cases the underlying molecular mechanisms remain to be elucidated. Here we show that the gut microbiota and their metabolites exert systemic regulatory effects at the level of host tissue epigenetic programming. This approach may provide valuable insight into a variety of gut microbiota-mediated host metabolic and immunologic phenotypes, enhancing our ability to not only understand how the gut dysbiosis affects host disease, but also to harness this metabolic organ to promote host health.

EXPERIMENTAL PROCEDURES

Mouse husbandry **–** Animal care and study protocols were approved by the University of Wisconsin- Madison Animal Care and Use Committee. Mice were housed in the Microbial Sciences Building vivarium. Conventionally-raised (ConvR) and germ-free (GF) C57BL/6J mice were bred at the University of Wisconsin-Madison to generate mice used in this study. GF mice were housed in separate plastic flexible vinyl gnotobiotic isolators.

Mice were group housed by colonization status and diet (3-5 mice/cage) under standard conditions (12 h light:dark, temperature- and humidity-controlled conditions), and received ad libitum access to water and food. After 3 weeks of age, mice were maintained on either a control breeder chow (5021, Lab Diet, 23.7%-kcal fat, 53.2% carbohydrate, 23.1% protein) or a high-fat high-sucrose (HF/HS) diet (TD.08811, Envigo Teklad, 44.6%-kcal fat, 40.6% carbohydrate, 14.8% protein). Diets were sterilized by irradiation and autoclaving. Sterility of germ-free animals was assessed by incubating freshly collected fecal samples under aerobic and anaerobic conditions using standard microbiology methods. Final dissection and data collection were performed at 19-weeks of age.

Conventionalized (ConvD) mice were generated by colonizing GF C57BL/6J mice with fresh cecal contents were collected from 15-week old conventionally-raised C57BL/6J mice maintained on a control breeder chow diet (n = 2 mice per donor cecal microbiota sample per experiment; 5021, Lab Diet). Immediately after sacrifice, fresh cecal contents from donor mice were re-suspended in Mega Medium (1:100 w/v) in an anaerobic chamber (Romano et al., 2015). Suspensions were transferred into anaerobic sealed tubes and used to colonize mice in a sterile biological safety cabinet. Germ-free 15-week-old C57BL/6J male mice were

inoculated via a single oral gavage with ~0.2 ml of cecal inocula (Turnbaugh et al., 2009b) and kept in sealed filter-top gnotobiotic cages for 4 weeks.

SCFA Supplementation **–** Germ-free (GF) C57BL/6J male mice were maintained in sterile HEPA filter cages and fed water and autoclaved chow ad libitum. At 12 weeks of age a subset of GF mice were supplemented with a mixture of short chain fatty acids (SCFA) (acetate 67.5mM, butyrate 40mM, propionate 25.9mM (P. M. Smith et al., 2013)) via drinking water supplied ad libitum or conventionalized (ConvD) with a C57BL/6J intestinal microbial community by gavage. Inoculum for colonization was prepared by suspending freshly collected fecal pellets in Mega Medium. Water was sterilized by autoclave prior to SCFA addition and filter sterilized after SCFA addition through a 0.2 µm filter before being supplied to the mice. SCFA supplemented water was freshly prepared and changed every 5 days and again 24h before sacrifice. Sterility of germ-free animals was assessed by incubating freshly collected fecal samples under aerobic and anaerobic conditions using standard microbiology methods. Prior to sacrifice and tissue collection at 14 weeks of age all mice (GF, GF+SCFA, ConvD) were anesthetized with 1-5% inhalant isoflurane supplied in oxygen.

The methods used to measure histone PTMs, 16S rRNA sequencing, gene expression, hepatic total cholesterol and triglycerides, and cecal and peripheral blood SCFAs are described in the *Supplemental Methods*.

CONTRIBUTIONS

KAK conceived the project, performed experiments, interpreted results, prepared figures and write the manuscript. Julia Kemis, Kymberleigh Romano and Greg Barrett-Wilt performed experiments, interpreted results, and contributed to writing the manuscript. Specifically, Julia Kemis collected fecal samples and performed 16S rRNA sequencing and analysis to profile the gut microbiota composition. Eugenio Vivas and Mary Rabaglia performed experiments. Alan Attie and Mark Keller interpreted results and contributed to writing the manuscript. John Denu and Federico Rey conceived the project, interpreted results, and wrote the manuscript.

ACKNOWLEDGEMENTS

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.

K.A.K. is supported by NIH F30 DK108494, J.A.D by NIH GM059789-15/P250VA. Additional support: Clinical and Translational Science Award program through the NIH National

Center for Advancing Translational Sciences grants UL1TR000427 and KL2TR000428 (F.E.R.), NIH DK108259 (F.E.R.), DK101573 (A.D.A.).

REFERENCES

Aebersold, R., Burlingame, A.L., Bradshaw, R.A., 2013. Western Blots versus Selected Reaction Monitoring Assays: Time to Turn the Tables? Molecular & Cellular Proteomics 12, 2381–2382. doi:10.1074/mcp.E113.031658

Amar, J., Serino, M., Lange, C., Chabo, C., Iacovoni, J., Mondot, S., Lepage, P., Klopp, C., Mariette, J., Bouchez, O., Perez, L., Courtney, M., Marre, M., Klopp, P., Lantieri, O., Doré, J., Charles, M.A., Balkau, B., Burcelin, R., D.E.S.I.R. Study Group, 2011. Involvement of tissue bacteria in the onset of diabetes in humans: evidence for a concept. Diabetologia 54, 3055–3061. doi:10.1007/s00125-011-2329-8

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

Bäckhed, F., Ding, H., Wang, T., Hooper, L.V., Koh, G.Y., Nagy, A., Semenkovich, C.F., Gordon, J.I., 2004. The gut microbiota as an environmental factor that regulates fat storage. Proc Natl Acad Sci USA 101, 15718–15723. doi:10.1073/pnas.0407076101

Belcheva, A., Irrazabal, T., Robertson, S.J., Streutker, C., Maughan, H., Rubino, S., Moriyama, E.H., Copeland, J.K., Kumar, S., Green, B., Geddes, K., Pezo, R.C., Navarre, W.W., Milosevic, M., Wilson, B.C., Girardin, S.E., Wolever, T.M.S., Edelmann, W., Guttman, D.S., Philpott, D.J., Martin, A., 2014. Gut microbial metabolism drives transformation of MSH2-deficient colon epithelial cells. Cell 158, 288– 299. doi:10.1016/j.cell.2014.04.051

Belkaid, Y., Hand, T.W., 2014. Role of the microbiota in immunity and inflammation. Cell. doi:10.1016/j.cell.2014.03.011

Besten, den, G., van Eunen, K., Groen, A.K., Venema, K., Reijngoud, D.-J., Bakker, B.M., 2013. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. J. Lipid Res. 54, 2325–2340. doi:10.1194/jlr.R036012

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

Cabreiro, F., Au, C., Leung, K.-Y., Vergara-Irigaray, N., Cochemé, H.M., Noori, T., Weinkove, D., Schuster, E., Greene, N.D.E., Gems, D., 2013. Metformin retards aging in C. elegans by altering microbial folate and methionine metabolism. Cell 153, 228–239. doi:10.1016/j.cell.2013.02.035

Canfora, E.E., Jocken, J.W., Blaak, E.E., 2015. Short-chain fatty acids in control of body weight and insulin sensitivity. Nature Publishing Group 11, 577–591. doi:10.1038/nrendo.2015.128

Cao, R., Wang, L.J., Wang, H.B., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., Zhang, Y., 2002. Role of histone H3 lysine 27 methylation in polycomb-group silencing. Science 298, 1039–1043. doi:10.1126/science.1076997

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L., Knight, R., 2010a. PyNAST: a flexible tool for align ng 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

Chassaing, B., Koren, O., Goodrich, J.K., Poole, A.C., Srinivasan, S., Ley, R.E., Gewirtz, A.T., 2015. Dietary emulsifiers impact the mouse gut microbiota promoting colitis and metabolic syndrome. Nature 519, 92–96. doi:10.1038/nature14232

Correa, R.O., Fachi, J.L., Vieira, A., Sato, F.T., Vinolo, M.A.R., 2016. Regulation of immune cell function by short-chain fatty acids. Clinical and Translational Immunology 2–8. doi:10.1038/cti.2016.17

Daniel, H., Gholami, A.M., Berry, D., Desmarchelier, C., Hahne, H., Loh, G., Mondot, S., Lepage, P., Rothballer, M., Walker, A., hm, C.B.O., Wenning, M., Wagner, M., Blaut, M., Schmitt-Kopplin, P., Kuster, B., Haller, D., Clavel, T., 2013. High-fat diet alters gut microbiota physiology in mice. ISME J 8, 295–308. doi:10.1038/ismej.2013.155

Daniel, H., Moghaddas Gholami, A., Berry, D., Desmarchelier, C., Hahne, H., Loh, G., Mondot, S., Lepage, P., Rothballer, M., Walker, A., Böhm, C., Wenning, M., Wagner, M., Blaut, M., Schmitt-Kopplin, P., Kuster, B., Haller, D., Clavel, T., 2014. High-fat diet alters gut microbiota physiology in mice. ISME J 8, 295–308. doi:10.1038/ismej.2013.155

David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A.V., Devlin, A.S., Varma, Y., Fischbach, M.A., Biddinger, S.B., Dutton, R.J., Turnbaugh, P.J., 2013. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505, 559–563. doi:10.1038/nature12820

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

Donohoe, D., Collins, L., Wali, A., Bigler, R., Sun, W., Bultman, S., 2012. The Warburg Effect Dictates the Mechanism of Butyrate-Mediated Histone Acetylation and Cell Proliferation. Molecular Cell 48, 612– 626. doi:10.1016/j.molcel.2012.08.033

Donohoe, D.R., Holley, D., Collins, L.B., Montgomery, S.A., Whitmore, A.C., Hillhouse, A., Curry, K.P., Renner, S.W., Greenwalt, A., Ryan, E.P., Godfrey, V., Heise, M.T., Threadgill, D.S., Han, A., Swenberg, J.A., Threadgill, D.W., Bultman, S.J., 2014. A Gnotobiotic Mouse Model Demonstrates that Dietary Fiber Protects Against Colorectal Tumorigenesis in a Microbiota- and Butyrate-Dependent Manner. Cancer Discovery CD–14–0501. doi:10.1158/2159-8290.CD-14-0501

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

Fan, J., Krautkramer, K.A., Feldman, J.L., Denu, J.M., 2015. Metabolic regulation of histone post-translational modifications. ACS Chem. Biol. 10, 95–108. doi:10.1021/cb500846u

Henao-Mejia, J., Elinav, E., Jin, C., Hao, L., Mehal, W.Z., Strowig, T., Thaiss, C.A., Kau, A.L., Eisenbarth, S.C., Jurczak, M.J., Camporez, J.P., Shulman, G.I., Gordon, J.I., Hoffman, H.M., Flavell, R.A., 2012. Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. Nature 482, 179–185. doi:10.1038/nature10809

Hill, M.J., 1997. Intestinal flora and endogenous vitamin synthesis. Eur. J. Cancer Prev. 6 Suppl 1, S43–5.

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

Jenuwein, T., Allis, C.D., 2001. Translating the histone code. Science 293, 1074–1080. doi:10.1126/science.1063127

Johnson, D.G., Dent, S.Y.R., 2013. Chromatin: receiver and quarterback for cellular signals. Cell 152, 685–689. doi:10.1016/j.cell.2013.01.017

Karlsson, F.H., Fåk, F., Nookaew, I., Tremaroli, V., Fagerberg, B., Petranovic, D., Bäckhed, F., Nielsen, J., 2012. Symptomatic atherosclerosis is associated with an altered gut metagenome. Nat Commun 3, 1245. doi:10.1038/ncomms2266

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

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., Jenuwein, T., 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature 410, 116–120. doi:10.1038/35065132

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

Ley, R.E., Bäckhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D., Gordon, J.I., 2005. Obesity alters gut microbial ecology. Proc Natl Acad Sci USA 102, 11070–11075. doi:10.1073/pnas.0504978102

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R., Gordon, J.I., 2008. Evolution of mammals and their gut microbes. Science 320, 1647–1651. doi:10.1126/science.1155725

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

Nicholson, J.K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., Pettersson, S., 2012. Host-gut microbiota metabolic interactions. Science 336, 1262–1267. doi:10.1126/science.1223813

Pai, Y.J., Leung, K.-Y., Savery, D., Hutchin, T., Prunty, H., Heales, S., Brosnan, M.E., Brosnan, J.T., Copp, A.J., Greene, N.D.E., 2015. Glycine decarboxylase deficiency causes neural tube defects and features of non-ketotic hyperglycinemia in mice. Nat Commun 6, 6388. doi:10.1038/ncomms7388

Pasini, D., Malatesta, M., Jung, H.R., Walfridsson, J., Willer, A., Olsson, L., Skotte, J., Wutz, A., Porse, B., Jensen, O.N., Helin, K., 2010. Characterization of an antagonistic switch between histone H3 lysine 27 methylation and acetylation in the transcriptional regulation of Polycomb group target genes. Nucleic Acids Research 38, 4958–4969. doi:10.1093/nar/gkq244

Perry, R.J., Peng, L., Barry, N.A., Cline, G.W., Zhang, D., Cardone, R.L., Petersen, K.F., Kibbey, R.G., Goodman, A.L., Shulman, G.I., 2016. Acetate mediates a microbiome-brain-β-cell axis to promote metabolic syndrome. Nature 534, 213–217. doi:10.1038/nature18309

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., Peng, Y., Zhang, D., Jie, Z., Wu, W., Qin, Y., Xue, W., Li, J., Han, L., Lu, D., Wu, P., Dai, Y., Sun, X., Li, Z., Tang, A., Zhong, S., Li, X., Chen, W., Xu, R., Wang, M., Feng, Q., Gong, M., Yu, J., Zhang, Y., Zhang, M., Hansen, T., Sanchez, G., Raes, J., Falony, G., Okuda, S., Almeida, M., LeChatelier, E., Renault, P., Pons, N., Batto, J.-M., Zhang, Z., Chen, H., Yang, R., Zheng, W., Li, S., Yang, H., Wang, J., Ehrlich, S.D., Nielsen, R., Pedersen, O., Kristiansen, K., Wang, J., 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490, 55–60. doi:10.1038/nature11450

Ridaura, V.K., Faith, J.J., Rey, F.E., Cheng, J., Duncan, A.E., Kau, A.L., Griffin, N.W., Lombard, V., Henrissat, B., Bain, J.R., Muehlbauer, M.J., Ilkayeva, O., Semenkovich, C.F., Funai, K., Hayashi, D.K., Lyle, B.J., Martini, M.C., Ursell, L.K., Clemente, J.C., Van Treuren, W., Walters, W.A., Knight, R., Newgard, C.B., Heath, A.C., Gordon, J.I., 2013. Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science 341, 1241214. doi:10.1126/science.1241214

Riggs, M.G., Whittaker, R.G., Neumann, J.R., Ingram, V.M., 1977. n-Butyrate causes histone modification in HeLa and Friend erythroleukaemia cells. Nature 268, 462–464. doi:10.1038/268462a0

Romano, K.A., Vivas, E.I., Amador-Noguez, D., Rey, F.E., 2015. Intestinal microbiota composition modulates choline bioavailability from diet and accumulation of the proatherogenic metabolite trimethylamine-N-oxide. MBio 6, e02481. doi:10.1128/mBio.02481-14

Samuel, B.S., Shaito, A., Motoike, T., Rey, F.E., Bäckhed, F., Manchester, J.K., Hammer, R.E., Williams, S.C., Crowley, J., Yanagisawa, M., Gordon, J.I., 2008. Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. PNAS 105, 16767– 16772. doi:10.1073/pnas.0808567105

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

Singh, V., Chassaing, B., Zhang, L., San Yeoh, B., Xiao, X., Kumar, M., Baker, M.T., Cai, J., Walker, R., Borkowski, K., Harvatine, K.J., Singh, N., Shearer, G.C., Ntambi, J.M., Joe,

B., Patterson, A.D., Gewirtz, A.T., Vijay-Kumar, M., 2015. Microbiota-Dependent Hepatic Lipogenesis Mediated by Stearoyl CoA Desaturase 1 (SCD1) Promotes Metabolic Syndrome in TLR5-Deficient Mice. Cell Metabolism 22, 983– 996. doi:10.1016/j.cmet.2015.09.028

Smith, K., McCoy, K.D., Macpherson, A.J., 2007. Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. Semin. Immunol. 19, 59–69. doi:10.1016/j.smim.2006.10.002

Smith, P.M., Howitt, M.R., Panikov, N., Michaud, M., Gallini, C.A., Bohlooly-Y, M., Glickman, J.N., Garrett, W.S., 2013. The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. Science 341, 569–573. doi:10.1126/science.1241165

Suez, J., Korem, T., Zeevi, D., Zilberman-Schapira, G., Thaiss, C.A., Maza, O., Israeli, D., Zmora, N., Gilad, S., Weinberger, A., Kuperman, Y., Harmelin, A., Kolodkin-Gal, I., Shapiro, H., Halpern, Z., Segal, E., Elinav, E., 2014. Artificial sweeteners induce glucose intolerance by altering the gut microbiota. Nature 514, 181–186. doi:10.1038/nature13793

Tan, J., McKenzie, C., Potamitis, M., Thorburn, A.N., Mackay, C.R., Macia, L., 2014. The role of short- chain fatty acids in health and disease. Adv. Immunol. 121, 91–119. doi:10.1016/B978-0-12-800100- 4.00003-9

Tilg, H., Moschen, A.R., 2014. Microbiota and diabetes: an evolving relationship. Gut 63, 1513–1521. doi:10.1136/gutjnl-2014-306928

Topping, D.L., Clifton, P.M., 2001. Short-chain fatty acids and human colonic function: roles of resistant starch and nonstarch polysaccharides. Physiological Reviews 81, 1031–1064.

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., 2009a. A core gut microbiome in obese and lean twins. Nature 457, 480–484. doi:10.1038/nature07540

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., Gordon, J.I., 2006. An obesity- associated gut microbiome with increased capacity for energy harvest. Nature 444, 1027–131. doi:10.1038/nature05414

Turnbaugh, P.J., Ridaura, V.K., Faith, J.J., Rey, F.E., Knight, R., Gordon, J.I., 2009b. The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. Sci Transl Med 1, –6ra14. doi:10.1126/scitranslmed.3000322

Vijay-Kumar, M., Aitken, J.D., Carvalho, F.A., Cullender, T.C., Mwangi, S., Srinivasan, S., Sitaraman, S.V., Knight, R., Ley, R.E., Gewirtz, A.T., 2010. Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. Science 328, 228–231. doi:10.1126/science.1179721

Wang, Z., Klipfell, E., Bennett, B.J., Koeth, R., Levison, B.S., DuGar, B., Feldstein, A.E., Britt, E.B., Fu, X., Chung, Y.-M., Wu, Y., Schauer, P., Smith, J.D., Allayee, H., Tang, W.H.W., DiDonato, J.A., Lusis, A.J., Hazen, S.L., 2011. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. Nature 472, 57–63. doi:10.1038/nature09922

Wellen, K.E., Hatzivassiliou, G., Sachdeva, U.M., Bui, T.V., Cross, J.R., Thompson, C.B., 2009. ATP- Citrate Lyase Links Cellular Metabolism to Histone Acetylation. Science 324, 1076–1080. doi:10.1126/science.1164097

Wikoff, W.R., Anfora, A.T., Liu, J., Schultz, P.G., Lesley, S.A., Peters, E.C., Siuzdak, G., 2009. Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. Proc Natl Acad Sci USA 106, 3698–3703. doi:10.1073/pnas.0812874106

Yu, X., Shahir, A.-M., Sha, J., Feng, Z., Eapen, B., Nithianantham, S., Das, B., Karn, J., Weinberg, A., Bissada, N.F., Ye, F., 2014. Short-chain fatty acids from periodontal pathogens suppress histone deacetylases, EZH2, and SUV39H1 to promote Kaposi's sarcoma-associated herpesvirus replication. Journal of Virology 88, 4466–4479. doi:10.1128/JVI.03326-13

Zhao, L., 2013. The gut microbiota and obesity: from correlation to causality. Nat Rev Micro 11, 639– 647. doi:10.1038/nrmicro3089

FIGURES

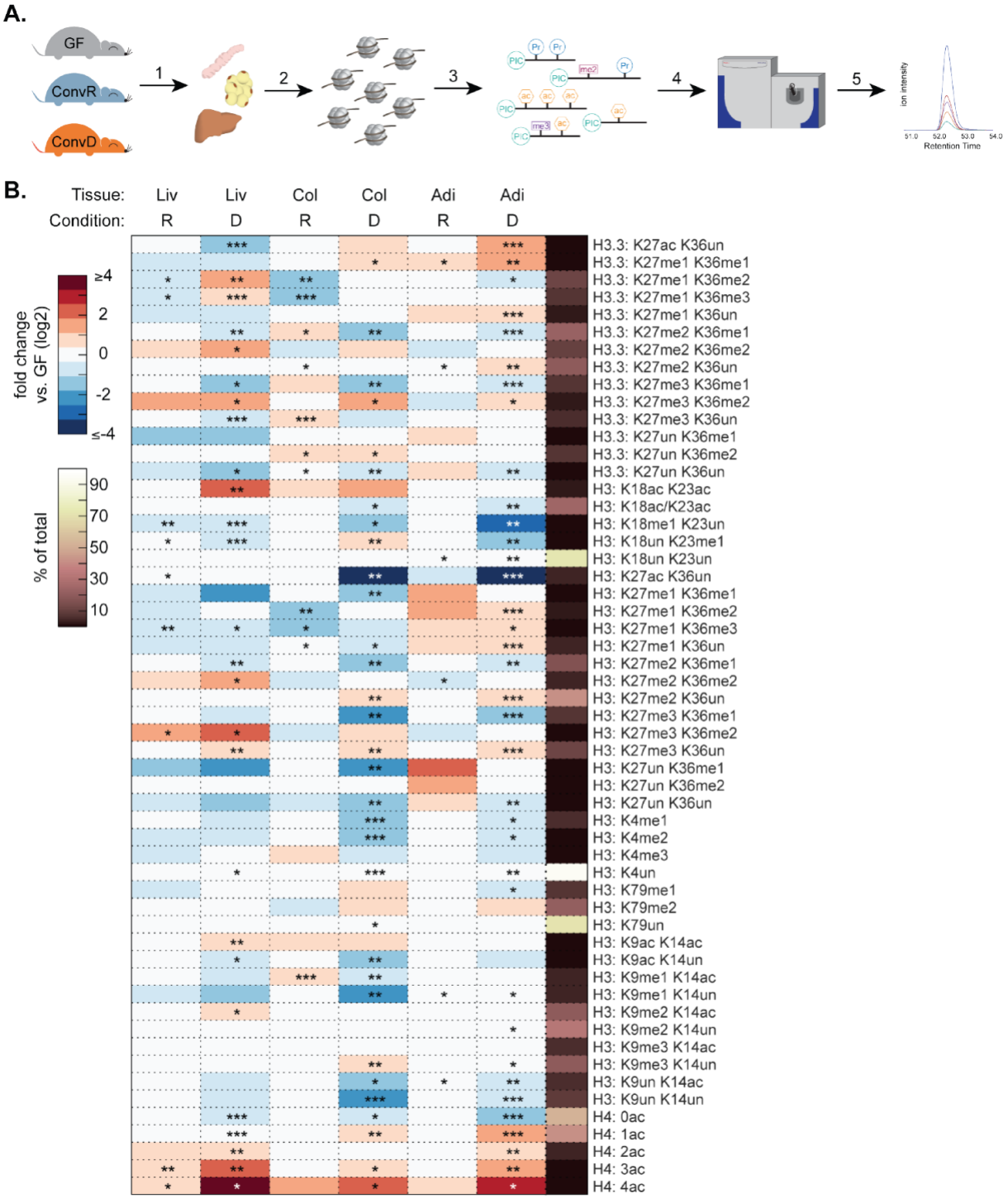

A.
GF
ConvR
ConvD
1
2
3
4
5
ion intensity
Retention Time
B.
Tissue: Liv Liv Col Col Adi Adi
Condition: R D R D R D
fold change vs. GF (log2)
≥4
2
0
-2
≤-4
% of total
90
70
50
30
10
H3.3: K27ac K36un
H3.3: K27me1 K36me1
H3.3: K27me1 K36me2
H3.3: K27me1 K36me3
H3.3: K27me1 K36un
H3.3: K27me2 K36me1
H3.3: K27me2 K36me2
H3.3: K27me2 K36un
H3.3: K27me3 K36me1
H3.3: K27me3 K36me2
H3.3: K27me3 K36un
H3.3: K27un K36me1
H3.3: K27un K36me2
H3.3: K27un K36un
H3: K18ac K23ac
H3: K18ac/K23ac
H3: K18me1 K23un
H3: K18un K23me1
H3: K18un K23un
H3: K27ac K36un
H3: K27me1 K36me1
H3: K27me1 K36me2
H3: K27me1 K36me3
H3: K27me1 K36un
H3: K27me2 K36me1
H3: K27me2 K36me2
H3: K27me2 K36un
H3: K27me3 K36me1
H3: K27me3 K36me2
H3: K27me3 K36un
H3: K27un K36me1
H3: K27un K36me2
H3: K27un K36un
H3: K4me1
H3: K4me2
H3: K4me3
H3: K4un
H3: K79me1
H3: K79me2
H3: K79un
H3: K9ac K14ac
H3: K9ac K14un
H3: K9me1 K14ac
H3: K9me1 K14un
H3: K9me2 K14ac
H3: K9me2 K14un
H3: K9me3 K14ac
H3: K9me3 K14un
H3: K9un K14ac
H3: K9un K14un
H4: 0ac
H4: 1ac
H4: 2ac
H4: 3ac
H4: 4ac

Figure A.1. Gut microbiota affect host tissue epigenetic states. (A) Experimental design: 1. proximal colon, liver, and white adipose tissue was harvested from germ-free (GF), conventionally raised (ConvR), and conventionalized (ConvD) mice. 2-3. Histones were extracted, chemically derivatized and trypsinized to generate peptides amenable to mass spectrometry analysis. 4-5. Histone peptides were injected onto a Thermo Q-Exactive mass spectrometer and data was acquired on >60 unique histone PTM states. (B) Relative abundance of histone PTMs on histone H3, histone H3.3, and histone H4. Values are reported as a fold change vs. GF controls (log2). The mean % of peptide family total across all samples is displayed in the right-most column. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, n=4 mice per condition.

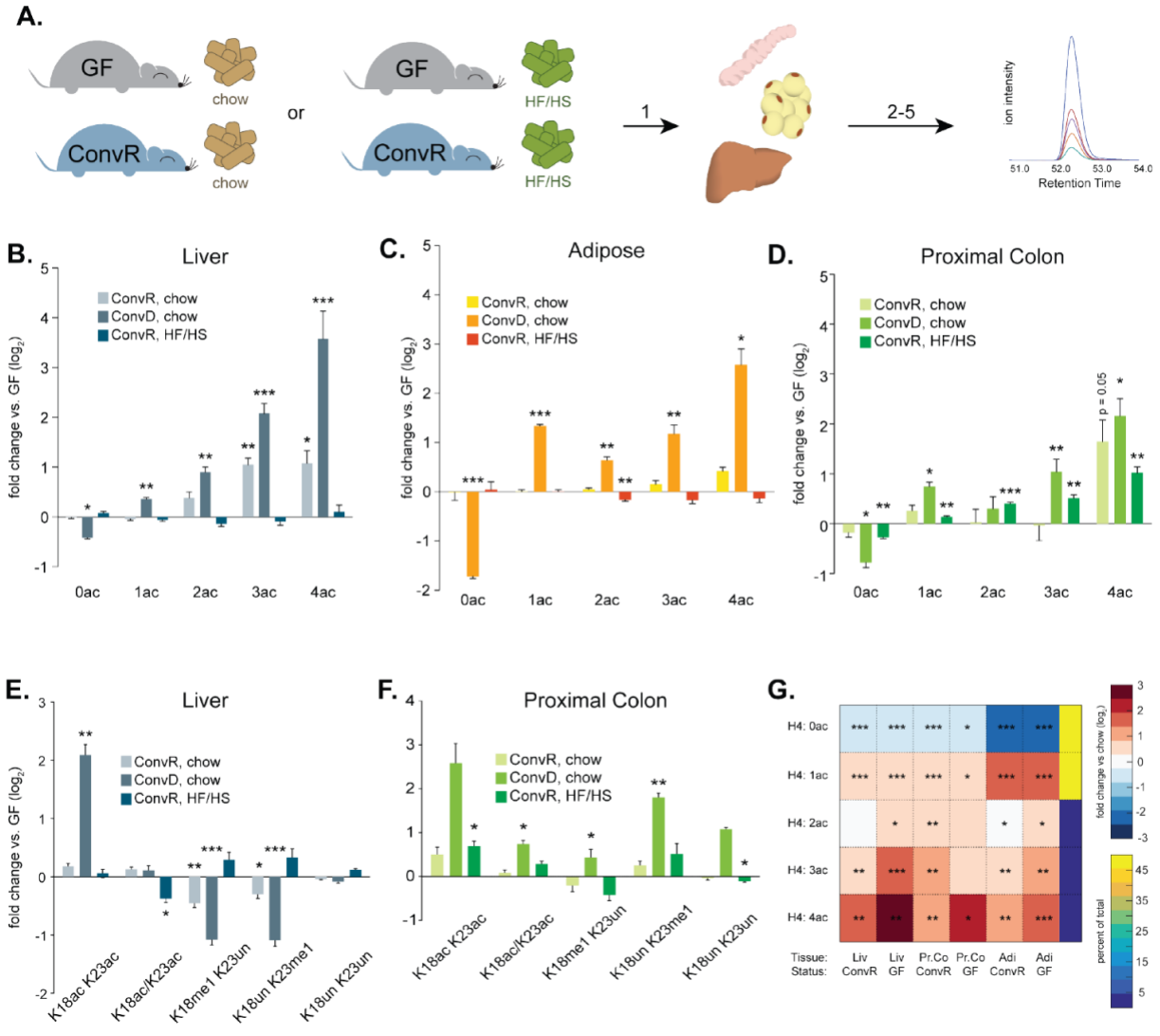


Figure A.2. Gut microbiota-mediated epigenetic changes are sensitive to diet. (A) Experimental design: 1.GF and ConvR mice were raised on either chow or a HF/HS ("Westernized") diet. 2-5. Tissues were harvested and histone extracts were prepared for mass spectrometry analysis as described in Figure 3.1. (B-D) Histone H4 (K5, K8, K12, and K16) acetylation in colonized liver (B), proximal colon (C), and adipose tissue (D) relative to GF controls (fold change, log2). (E-F) Histone H3 K18 and K23 methylation and acetylation in colonized liver (E) and proximal colon (F) relative to GF control (fold change, log2). (G) Histone

H4 (K5, K8, K12, and K16) acetylation in tissues of HF/HS-fed vs. chow-fed mice. Mean % of peptide family totals are displayed in the right-most column. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, error bars represent standard error from the mean, n=4 mice per condition.

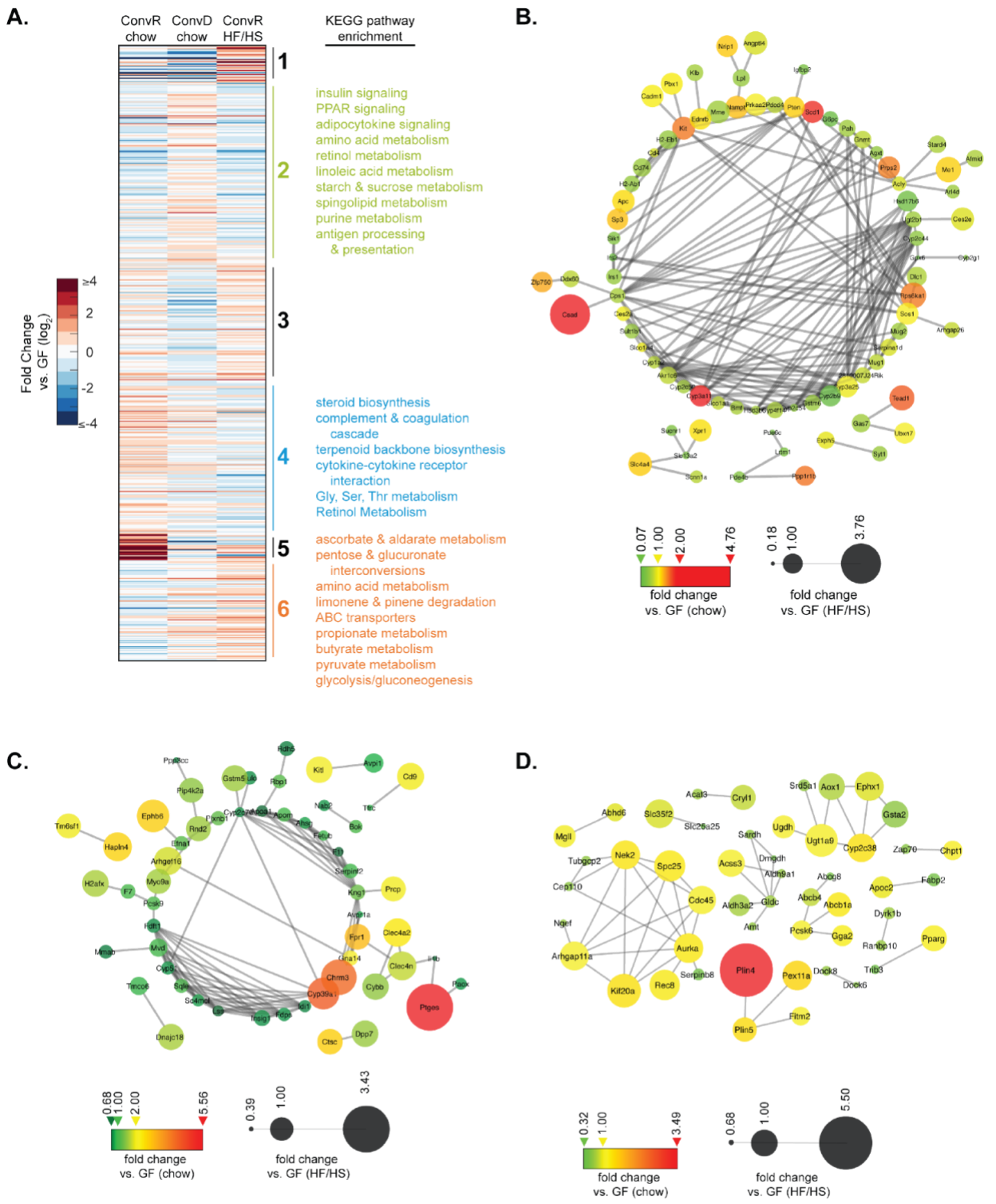
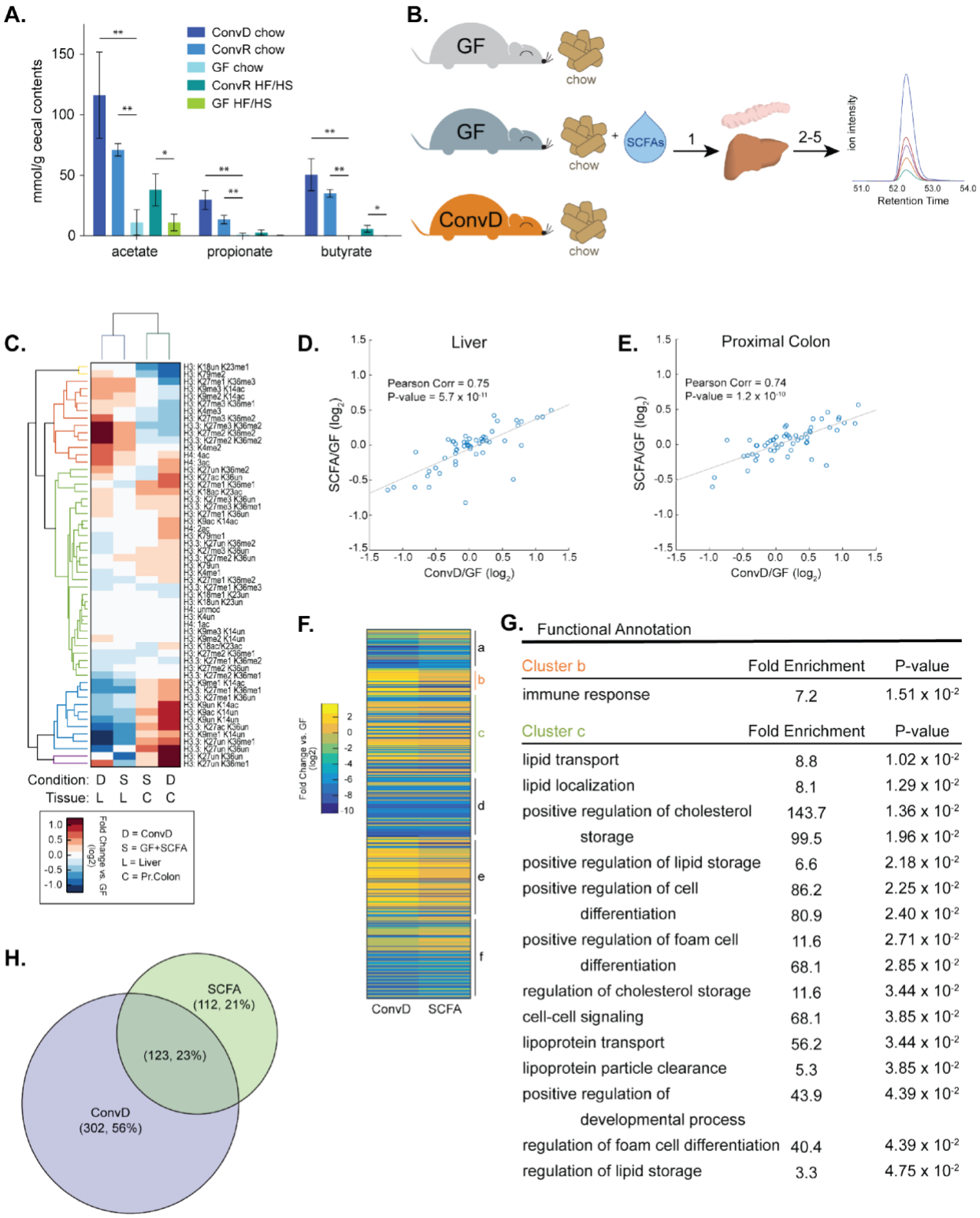


Figure A.3. Hepatic genes are co-regulated in colonized mice as a function of diet or colonization status. (A) K-means clustering of differentially expressed hepatic genes (left, fold

change vs. GF, log2) and KEGG pathway enrichment terms (right). (B-D) Interaction network for cluster 2 (B), cluster 4 (C), and cluster 6 (D). Only genes with at least one reported interaction are graphed. Edges indicate interaction. Node size indicates relative expression in HF/HS-fed mouse livers. Node color indicates relative expression in chow-fed mouse livers.



G. Functional Annotation

Cluster b	Fold Enrichment	P-value
immune response	7.2	1.51×10^{-2}

Cluster c	Fold Enrichment	P-value
lipid transport	8.8	1.02×10^{-2}
lipid localization	8.1	1.29×10^{-2}
positive regulation of cholesterol	143.7	1.36×10^{-2}
storage	99.5	1.96×10^{-2}
positive regulation of lipid storage	6.6	2.18×10^{-2}
positive regulation of cell	86.2	2.25×10^{-2}
differentiation	80.9	2.40×10^{-2}
positive regulation of foam cell	11.6	2.71×10^{-2}
differentiation	68.1	2.85×10^{-2}
regulation of cholesterol storage	11.6	3.44×10^{-2}
cell-cell signaling	68.1	3.85×10^{-2}
lipoprotein transport	56.2	3.44×10^{-2}
lipoprotein particle clearance	5.3	3.85×10^{-2}
positive regulation of	43.9	4.39×10^{-2}
developmental process		
regulation of foam cell differentiation	40.4	4.39×10^{-2}
regulation of lipid storage	3.3	4.75×10^{-2}

Figure A.4. SCFA-supplementation mimics colonization-induced epigenetic programming. (A) SCFA measurement in cecal contents from GF, ConvR, and ConvD animals on chow and HF/ HS

diet. *$p < 0.05$, **$p < 0.01$, error bars represent standard deviation, n = 4 mice per condition. (B) Experimental design: 1. Germ-free mice were supplemented with SCFAs (GF+SCFA) or colonized (ConvD) and tissues were harvested. 2-5. Histone extracts were prepared as described in figure 1. (C) Hierarchical clustering of histone PTMs in colonized and GF+SCFA mouse tissues (fold change vs. GF, log2). (D-E) Pearson's correlation of ConvD and GF+SCFA mouse tissue histone PTM states in liver (D) and proximal colon (E). (F) K-means clustering of differentially expressed hepatic genes in ConvD and GF+SCFA mice. FDR cutoff for differential expression = 0.05, n = 3 mice per condition. (G) GO-term enrichment in clusters b and c. (H) Overlap of differentially expressed genes between ConvD and GF+SCFA mice.

APPENDIX B: Social Relationships, Social Isolation, and the Human Gut Microbiota

The work presented in this appendix is currently under review:

Dill-McFarland KA[*], *Tang ZZ*[*], ***Kemis JH***, *Kerby RL, Chen G, Palloni A, Sorenson T, Rey FE*[+], *Herd P*[+]

[*] *indicates lead author,* [+] *indicates corresponding author*

Supplemental data files available online

ABSTRACT

Social relationships shape human health and mortality via behavioral, psychosocial, and physiological mechanisms, including inflammatory and immune responses. Though not tested in human studies, recent primate studies indicate that the gut microbiome may also be a biological mechanism linking relationships to health. Integrating microbiota data into the 60-year-old Wisconsin Longitudinal Study, we found that socialness with family and friends is associated with differences in the human fecal microbiota. Analysis of spouse (N = 94) and sibling pairs (N = 83) further revealed that spouses have more similar microbiota and more bacterial taxa in common than siblings, with no observed differences between sibling and unrelated pairs. These differences held even after accounting for dietary factors. The differences between unrelated individuals and married couples was driven entirely by couples who reported close relationships; there were no differences in similarity between couples reporting somewhat close relationships and unrelated individuals. Moreover, the microbiota of married individuals, compared to those living alone, has greater diversity and richness, with the greatest diversity among couples reporting close relationships, which is notable given decades of research documenting the health benefits of marriage. These results suggest that human interactions, especially sustained, close marital relationships, influence the gut microbiota.

INTRODUCTION

Social relationships exert a sustained influence on human health and mortality with social isolation having strong negative consequences and high levels of social integration far exceeding the protective effects on mortality of individual level behaviors such as smoking cessation or maintaining a normal weight [1,2]. Research in the social sciences has shown that individuals who cohabitate in marriage and marital like relationships have better health than do unpartnered adults[3]. For both social relationships generally, and marriage specifically, health benefits are largely achieved in the context of high-quality relationships. The robust links between these relationships and health are related to stress, behaviors, and psychosocial resources, among other factors [2]. In part, social support may impact one's health by reinforcing healthy habits, reducing the impacts of stress, and preventing the use of unhealthy "self-medications" like smoking and drinking [2]. Additional research points to stress-related biological processes that may also contribute to the positive impacts of social relationships through changes in inflammatory processes, metabolic syndrome, and neurological functioning [4,5].

Recent work in the field of microbiology points to another possible biological mechanism linking human relationships and health: the microbiome. The microbial communities that inhabit mammals have profound effects on biology and health [6]. Gastrointestinal (GI) microbial communities impact host health by modulating the epigenome [7], brain function [8], and metabolism of drugs and nutrients [9] as well as impacting immune system function [10] and development [11]. While the microbiota reaches an adult-like configuration by three to five years of age [12], considerable variation exists between adults [13], and differences are mediated by a number of factors. Most notable among these are diet [14] and host genetics [15], which also correlate with health. An individual's microbiota structure (*i.e.* relative abundance) and composition (*i.e.* who's there) can

change rapidly in response to inputs like diet [16] and antibiotics [17]. Nonetheless, there is evidence that an individual's microbiota remains relatively stable over many years [18–20], perhaps in part because a person's behaviors also tend to be consistent over many years.

While a number of factors like diet are known to impact both the microbiota and health [21], less is known regarding social relationships. Most existing research has focused on animal models, which has produced compelling evidence that social interactions, via a range of different types of physical contact, influences the gut microbiota through microbial sharing between individuals [22–26]. Additionally, states of isolation, such as maternal neglect, influence the gut microbial composition in animal models [27] at least in part through stress [28,29]. Thus, the gut microbiota may play a role in some of the long-term health effects of social relationships.

But despite this tantalizing evidence, studies in human populations remain relatively small in number [30]. There are a few studies exploring how mother-infant interactions influence the development of the infant's gut microbiome and even how broader social interactions influence the milk microbiome [31,32]. In terms of adults, there is evidence regarding the influence of cohabitation, may influence the gut microbiome. A few recent studies have found that individuals living together had more similar gut [33] and skin [33,34] microbiota. Interestingly, however, another study found that married cohabitating couples had no more similarity in the composition of their gut microbiota than did unrelated individuals [35].

Thus, while it does appear that living together may influence the gut microbiome, human studies have not investigated how adult relationships, rather than just simply living in the same space, may influence the gut microbiome. The quality of the relationship may matter. Closer relationships likely lead to even closer shared environments, via mechanisms such as time spent physically together. Indeed, one recent study of wild baboons found that close partners within

social groups had more similar gut microbiotas [36]. Studies have also have not more generally compared how living alone versus living with an intimate partner influences the gut microbiome; individuals living alone are on average, de facto, more socially isolated than those living with someone, and animal studies have generally shown that social isolation leads to decreased microbial diversity [22,37–39]. Though causality is not certain, decreased microbial diversity is associated with obesity, cardiac disease, and type 2 diabetes, and a range of other inflammatory disorders [40–47]. More broadly, there is extensive evidence that cohabitating couples in later life have substantially improved physical and psychological well-being compared to single adults [48–50]. Thus, similar mechanisms might explain some of the variance in findings in humans.

An important hindrance to research examining social relationships and the GI microbiota is the availability of human samples with sufficiently well-characterized life course measures of broader social environments and conditions. Thus, most microbiological research in this field is based on animal models [22–25]. However, there are now a wide array of well-characterized longitudinal studies in the social sciences that have generated decades of research documenting relationships between broader social environments and mortality [5,51–54]. These data can provide a platform for studies of the human microbiota to advance knowledge for both social scientists and microbiologists, including whether social conditions influence the gut microbiota and whether the gut microbiota is a mediating biological mechanism explaining how social conditions influence health.

Here, we leverage a multidisciplinary collaboration to investigate the links between human interaction, the microbiota, and human health. We utilized data in the nearly 60-year Wisconsin Longitudinal Study (WLS) [54], which constitutes a random sample of 1 in 3 1957 Wisconsin high school graduates (N = 10,317), as well as selected spouses and siblings surveyed periodically

during their adult life. We correlate the fecal microbiota of 408 older individuals (58 – 91 yo) from WLS with extensive health and behavioral data, as well as compare spouse and sibling pairs within the cohort. Overall, this project demonstrates the promise of joint participation between social scientists and microbiologists in efforts to more fully understand the gut microbiota and its impacts on human health.

RESULTS AND DISCUSSION

We employed 16S rRNA gene sequencing to characterize the fecal microbiota of 408 individuals, including Wisconsin Longitudinal Study (WLS) graduates (N = 179, 76 ± 0.5 years old), siblings of graduates (134, 74 ± 6.4), spouses of graduates (63, 76 ± 3.7), and spouses of siblings (32, 73 ± 6.1). We then correlated these communities to longitudinal survey data collected from 1957 to 2015 as part of WLS [54]. For more details on this data collection, see [55]. A total of 24.5 million high-quality sequences were obtained for 408 fecal samples (60,000 ± 19,000 SD sequences per sample) after quality filtering in mothur. All samples achieved sufficient coverage as determined by Good's coverage > 99% (Dataset S1).

In the WLS graduate cohort, we identified several factors correlated with gastrointestinal (GI) microbiota including sex, antibiotics, dietary protein, high blood sugar, and heart disease (Fig.B.1, Fig. S1, Table S1). These factors were reported in the previous literature [57,58] with diet playing a particularly strong role [14,16,56]. Thus, we assessed diet across a number of measures including habitual intake of protein, vegetables, and fruits (Text S1) during the year prior to the fecal sample collection (for details, see METHODS, Statistical analysis for graduates). While overall dietary dissimilarity (Bray-Curtis and Jaccard) across these three categories correlated with gut microbiota dissimilarity, only the total frequency of dietary protein consumption was robustly associated with microbial composition using either univariate or multivariate analyses (Table S1). Thus, we note that all analyses have adjusted for potential confounders including age, sex, antibiotics, dietary protein, and chronic conditions (diabetes and heart disease) unless stated otherwise. In some analyses –that are noted below—we do also include vegetable and fruit dietary data.

Social interactions and the human fecal microbiota

Human interactions were also associated with differences in gut microbiota and diversity. Specifically, we found that individuals that were cohabitating with a spouse or partner had more similar microbiota composition with their cohabitating spouse/partner as well as higher diversity and richness than unmarried, non-cohabitating individuals (unweighted UniFrac $P = 0.029$ Shannon $P = 0.005$, Chao $P = 0.011$, Fig. B.2). Since all cohabitating pairs were male-female and sex was a strong determinant of the microbiota in this study ($P < 0.001$, Table S2), increased diversity may be partially due to sustained exchange of microorganisms between the sexes, though we were not able to test this given that there were no same sex couples in these data. Increases in diversity seen here are consistent with a previous cohabitation study in pigs [59] and may have implications for human health, as previous work indicates that increased gut microbial diversity is associated with lower risks of irritable bowel syndrome (IBS), Crohn's disease, ulcerative colitis, and other GI afflictions [60]. Social interactions with relatives and friends were stronger predictors of gut microbial diversity in non-cohabiting individuals than cohabiting spouses/partners (unweighted UniFrac $P = 0.0030$, Shannon $P = 0.042$, Chao $P = 0.063$, Fig. S2) (Table S2). Here, social interactions were defined as the sum of "How many times during the past four weeks have you gotten together with relatives/friends?" The associations may have been weaker for cohabitating spouses due to their higher microbial diversity; ecological theory supports that diverse communities are more resilient and resistant to invasion by new species [61]. Thus, one explanation for these differential associations is that the more diverse microbiotas of individuals already cohabitating with a spouse may not have been as strongly influenced by increasing social interactions while the less diverse microbiotas of those living alone were more strongly influenced by invasion of new species through social exposures. It is also possible that cohabitating couples share the same friends and socialize together with these friends. However,

factors contributing to the resilience of the human gut microbiota require further exploration to confirm this hypothesis.

Spouses have more similar microbes than siblings and unrelated individuals

Previous studies have established that the GI microbiota reaches an adult-like configuration by 3 to 5 years of age[18,62,63] and that during adulthood, communities are stable on the time scale of years [19,20]. Thus, microbial communities established in early life may persist and, aside from extreme perturbation, remain stable across one's adult lifetime. However, our analyses comparing sibling, couple, and unrelated pairs challenge the assumption that microbial communities established in early life will be largely unperturbed in later life (for details, see METHODS, Statistical Analysis for spouse and siblings). In fact, we find no evidence for a remaining influence of early life on the composition of the gut microbiota among older adults. In this older cohort, spouses were more similar than unrelated subjects (unweighted UniFrac P = 3.2E-5) or sibling pairs (unweighted UniFrac P = 0.033, Fig. B.3). Further, the length of the cohabitating marital relationship was positively correlated with similarity (unweighted UniFrac, P = 0.031) In contrast, siblings were no more similar than unrelated pairs by any beta-diversity metric ($P > 0.3$, Fig. B.3A, Fig. S3A, D, G) (Table S2). We also found no evidence that the physical proximity of siblings—as measured by physical distance between siblings—influenced gut microbial similarity. Thus, adult factors like marriage with cohabitation (spouses) appear to have a greater influence on the adult gut microbiota than early-life environment or genetics (siblings).

This is further supported by our findings that childhood farm status was not associated with microbial richness (Chao P = 0.342) while working on a farm as an adult correlated with higher richness (Chao P = 0.005). Farm-driven differences in the microbiota are of particular interest, because adolescents that grew up on a farm have more diverse microbial communities [64] and

reduced risk of asthma and other atopic diseases both during childhood [65] and as adults [66]. Given the results here, it appears that the microbially-driven protective effects of early farm exposures are not due to the persistence of protective microorganisms acquired in early-life. Protection may, instead, be conferred by immune development and training by early-life microbes as suggested previously [67].

Our results are also in contrast with previous work showing that genetically related individuals harbor more similar microbial communities than unrelated individuals, regardless of current cohabitation [35,68–70]. However, these previous studies investigated children [68], young adults[35,69], or a wide age range [70], and therefore, cumulative changes across a lifetime may not have reached a level sufficient to overcome early-life factors impacting the microbiota. Additionally, sibling pairs in other studies were twins [35,68–70], and many focused on monozygotic twins (same sex and age) [35,69,70] as opposed to this study where siblings were often of opposite sexes (43%) and ranged from less than a year to 18 years apart in age. Also, the unrelated group in this study may have exhibited higher homogeneity than unrelated groups in other studies, because most grew up in and/or currently live in the state of Wisconsin. Thus, compared to previous studies, siblings were likely less similar and unrelated pairs more similar across our cohort. Furthermore, genetic effects on the microbiota are often small [70] and detection may require a larger human cohort than used here. Taken together, these factors may have contributed to the lack of significant differences observed between sibling and unrelated groups even though average sibling beta-diversity was intermediate between spouses and unrelated individuals.

Increased microbial similarity, diversity, and richness in closer relationships

For both spouse and sibling relationships, microbiota similarity was associated with self-reported relationship closeness (unweighted UniFrac P = 0.0079). Closeness was measured by

participant responses to “How close are you and your current spouse/sibling?” on a scale of not at all (1) to very (4). Due to the small sample sizes in the categories “Not very” (N=13) and “Not at all” (N=4), we combined these two groups into “Not” close. Across spouses and siblings, individuals in very close relationships harbored gut microbial communities more similar to their close social partners than those in not very close relationships (Fig. B.3B), though this relationship was not significant within the spousal and sibling pair groups separately (Fig. B.3C). Moreover, differences between spouses and unrelated individuals, in terms of closeness (Fig. B.2), as well as the enhanced diversity and richness in cohabitating couples versus individuals living alone (Fig. B.2) were driven by spouses reporting very close relationships. This was in contrast to couples reporting only somewhat close relationships as these pairs did not have higher gut microbiota similarity than unrelated pairs (Table S3) nor did they display microbial diversity or richness different from non-cohabitating individuals (Table S3). Importantly, the apparent impacts of relationship closeness do not appear to be mediated by similarities in diet since overall dietary dissimilarity (Bray-Curtis and Jaccard) did not significantly differ according to relationship closeness (ANOVA $P > 0.5$; Table S4). We note that these included sensitivity tests that modeled diet based on the protein consumption, but also overall diet that captured vegetable and fruit consumption.

While diet is often correlated with the GI microbiota [56], closeness points to the less well-understood contributions of human interactions and shared behaviors. Close proximity and frequent physical contact were correlated with microbiota similarity among primates with direct microbial sharing between individuals contributing to similarity [22,23]. In this study, relationship closeness may represent a summative measure of time spend together, physical affection, and other human interactions with the potential to result in microbial sharing. Indeed, there is evidence that

the salivary microbiome influences the gut microbiome and the salivary microbiome may be influenced by kissing [71,72]. In these data, this is supported by the fact that spouses had more operational taxonomic units (OTUs, a proxy for microbial species) in common (30.4 ± 7.32%) than siblings (26.4 ± 7.47%, t-test P = 4.39E-04) (Dataset S2). Also, when comparing the spouse and sibling pair within a family represented in this dataset, a person tended to have more OTUs in common with his or her very close spouse (25.4 ± 7.9%) than his or her very close sibling (22.2 ± 6.4%, N = 12 families, P = 0.074, Fig. 3D). This is also true when comparing very close spouses (22.9 ± 5.8%) and somewhat close siblings within a family (20.6 ± 5.5%, N = 17 families, P = 0.027, Fig. B.3E).

Shared taxa with close human relationships.

In general, highly abundant genera and OTUs were shared between many spouse and sibling pairs while less abundant shared taxa were specific to one pair type and shared by a small number of pairs within that type (Dataset S3). OTUs that were commonly found among spouses or siblings (> 50% of pairs) but rare in the unrelated dataset (< 70% individuals, < 49% unrelated pairs) may represent bacterial species easily shared by close human interaction. These OTUs were predominately from the phylum Firmicutes (16 of 22 OTUs) with representatives of families Lachnospiraceae and Ruminococcaceae (Dataset S4). Interestingly, most of these potentially shared OTUs were from strictly anaerobic taxa, indicating that persisting in an oxygen-rich environment in-between hosts may not be a limiting factor in very close human relationships. Transmission, in these cases, could be mediated by direct contact similar to mechanisms of vertical transmission from mother to child [73].

Taxa commonly associated with reduced disease incidence or severity like *Akkermansia muciniphila* [74], *Bifidobacterium* sp. [75,76], *Collinsella aerofaciens* [76], and *Ruminococcus bromii* [77]

as well as potentially harmful taxa like *Clostridium spiroforme* [78,79] were often present in both persons in a spouse or sibling pair. Several of these potentially shared OTUs were associated with disease incidence in the larger dataset. In particular, *Ruminococcus bromii*, *Lachnospira* sp. and unclassified Ruminococcaceae and Lachnospiraceae OTUs were less abundant in those with high blood sugar (Fig. 4, Dataset S4). These results are in contrast to previous reports of more abundant Ruminococcaceae/*Ruminococcus* [80,81] and Lachnospiraceae [80] associated with diabetes in humans and may point to important differences in the impacts of the microbiota on metabolic health in older populations. Overall, though, this indicates that GI microbial species with the potential to impact host health may be shared by close human interactions. However, it cannot be discounted that these apparent health associations may be mediated by diet as those with high blood sugar often consume specific diets to manage disease.

Overall, our findings indicate that in order to understand environmental influences on the gut microbiota, we must now consider the many microbiotas with which this individual interacts. Socialness with family and friends is associated with differences in the fecal microbiota. These differences held even after accounting for dietary factors, though given this is the first study of its kind, it will be critical for future work to validate this finding. Thus, it is possible that relationships with others may influence the gut microbiota and consequent health outcomes, either through direct microbial transfer or reinforcement of healthy microbiota behaviors. We further found not only that married couples had more similar gut microbiota but also that the microbiota of married individuals, compared to those living alone, has greater diversity and richness. Key to both of these findings, however, was that they were driven by individuals reporting that they were very close to their spouse as opposed to somewhat close. Close marriage relationships had a stronger influence than the shared genetic factors and early life environments among siblings. This finding is

interesting, in part, because it parallels an extensive body of evidence demonstrating robust links between high quality marriages and morbidity and mortality. Future work could attempt to disentangle the mechanisms linking close relationships to microbial composition. For example, while we did not find evidence that shared diet was primarily responsible for these findings, we could not test precise frequencies of physical contact and intimacy as an alternative explanatory mechanism. Importantly, the types of physical contact and intimacy change over the life course, with sexual intimacy becoming far less frequent in later life, but other kinds of intimate physical contact remaining important. Regardless of the mediating mechanism, from a social and population health science perspective, decades of evidence that social relationships, especially close ones like marriage, influence morbidity and mortality make the central finding of significant interest. For example, even if future work finds a greater role for shared diets, it is still the social relationships that drive that shared diet.

Overall, these results provide support for the gut microbiome as a possible mediating pathway between social relationships, especially marriage, and health and mortality. These findings, in the context of the robust body of evidence linking social relationships to human morbidity and mortality, provide fodder for further work examining the role of the gut microbiome as a possible biological mediator in these relationships [32]. Further microbiota work across time in a more diverse population should be undertaken with the many longitudinal social science studies currently underway in an effort to increase our understanding of the complex interactions between human behavior, the microbiota and health.

EXPERIMENTAL PROCEDURES

Wisconsin Longitudinal Study (WLS). WLS is based on a one-third sample of all 1957 Wisconsin high school graduates (N = 10,317) as well as selected siblings and spouses [54]. Graduates originally enrolled with an in-person questionnaire upon graduating high school in 1957, which was followed by data collection in 1964, 1975, 1992, 2004, and 2011. Siblings were surveyed in 1977, 1994, 2005, and 2011; spouses were surveyed in 2004 or 2006. The content of WLS surveys changed to reflect the participants' life course with an education focus in the initial data collection, familial and career outcomes in young adulthood / midlife, and health, cognitive functioning, psychological well-being, non-work activities, caregiving, bereavement, social support, and end-of-life preparations in later rounds. WLS data collection was approved by the Institutional Review Board (IRB) at the University of Wisconsin-Madison (2014-1066, 2015-0955). Informed consent, the content and procedures of which were included in the IRB approval, was obtained from participants. All methods were performed in accordance with relevant guidelines and regulations.

Study design. A total of 500 individuals were randomly drawn from the full WLS dataset constrained based on the following: 1) participated in the 2011 interviews; 2) lived in one of 10 counties in Wisconsin that included both northern rural counties and southern more urban counties; and 3) were part of a sibling pair. Individuals were removed from the study if they did not give consent, their sample did not arrive for processing chilled, but not frozen, within 48 hrs of collection, or their sample did not yield at least 10,000 sequences for analysis. This resulted in 408 individuals being included in this study. An additional survey was administered at the time of fecal sampling, which detailed dietary data from the prior three days, prescription/antibiotic use, current

living situation, and additional health information. This as well as selected data from the larger WLS study focused on health, spouse/sibling relationships, and social interactions were used in this study (Text S1). Data, documentation, and other materials are accessible at http://www.ssc.wisc.edu/wlsresearch/. Access to the full dataset can be obtained through wls@ssc.wisc.edu.

Sample collection. Stool samples were collected by participants in November 2014, January 2015, or April 2015 following provided instructions (Text S2). Participants stored samples at ~4 °C in their refrigerator or in a NanoCool box (Albuquerque, NM) with cooling cartridge and customized foam insert, supplemented with a single ice pack. Interviewers picked-up samples from participants within 24 hours of collection and shipped samples in fresh NanoCool boxes for arrival at UW-Madison within 48 hours of collection. Upon arrival, an aliquot of feces was collected for DNA extraction and immediately stored at -80°C until further processing. The use of WLS and fecal microbiota data were approved by the Institutional Review Board at the University of Wisconsin-Madison (2017-0600).

DNA extraction. Genomic DNA was extracted from fecal aliquots using a bead-beating protocol[45]. Briefly, feces (~100 mg) were re-suspended in a solution containing 500 µl of extraction buffer 344 [200 mM Tris (pH 8.0), 200 mM NaCL, 20 mM EDTA], 210 µl of 20% SDS, 500 µl phenol:chloroform:isoamyl alcohol (pH 7.9, 25:24:1) and 500 µl of 0.1-mm diameter zirconia/silica beads. Samples were mechanically disrupted using a bead beater (BioSpec Products, Barlesville, OK; maximum setting for 3 min at room temperature), followed by centrifugation, recovery of the aqueous phase, and precipitation with isopropanol. QIAquick 96-well PCR

Purification Kit (Qiagen, Germantown, MD) was used to remove contaminants. Isolated DNA was eluted in 5 mM Tris/HCl (pH 8.5) and was stored at -80 °C until further use. We also note that we used negative controls.

Sequencing. PCR was performed using universal primers flanking the variable 4 (V4) region of the bacterial 16S rRNA gene [82]. We used negative controls for each PCR reaction. PCR reactions where the negative control yielded a product were not sequenced until the problem was solved. Samples were processed all together, not in batches, in a random order (i.e., not clustered by family). Additionally, unlike other specimens (e.g., saliva, skin), DNA contamination from reagents is in general not a problem for fecal samples given the high DNA content of the sample (10^{12} microbes/g of feces). In one reaction per sample, 10 - 50 ng DNA, 10 µM each primer, 12.5 µl 2X HotStart ReadyMix (KAPA Biosystems, Wilmington, MA, USA), and water to 25 µl were used. Cycling conditions were initial denaturation of 95 °C for 3 min followed by 25 cycles of 95°C for 30 s, 55 °C for 30 s, and 72 °C for 30 s, with a final extension of 72 °C for 5 min. PCR products were purified with the QIAquick 96-well PCR Purification Kit (Qiagen, Germantown, MD, USA). Samples were quantified by Qubit Fluorometer (Invitrogen, Carlsbad, CA, USA) and equimolar pooled. The pool plus 5% PhiX control DNA was sequenced through the U. of Wisconsin-Madison Biotechnology Center with the MiSeq 2x250 v2 kit (Illumina, San Diego, CA, USA) using custom sequencing primers [82]. All DNA sequences are available upon institutional review board (IRB) or other ethics board approval through wls@ssc.wisc.edu.

Sequence clean-up. All sequences were demultiplexed on the Illumina MiSeq. Sequence clean-up and processing was performed with mothur v.1.36.1 [83] following a protocol similar to [82]. Briefly,

paired-end sequences were combined into contigs with default parameters (match bonus 373 = 1, mismatch penalty = -1, gap penalty = -2, gap extend penalty = -1, insert quality ≥ 20, mismatch quality difference ≥ 6). Poor-quality sequences, including those with ambiguous basepairs, homopolymers greater than 8, or outside 200 – 500 bp in length, were discarded. Sequences were then aligned to the SILVA 16S rRNA gene reference alignment database [84] and trimmed to the V4 region. To reduce sequencing error, sequences with 2 or fewer differences were pre-clustered. Chimera detection and removal were performed using UCHIME [85]. Final sequences were then classified to the GreenGenes database [86]. Singletons were removed to facilitate downstream analyses. All sequences were grouped into 98% operational taxonomic units (OTUs) by uncorrected pairwise distances and average neighbor clustering in mothur. Clustering performed on uncorrected pairwise distances revealed no differences in clusters at 97 vs 98% similarity. Therefore, the stricter cutoff was reported Coverage was assessed by Good's coverage, and then samples were normalized to whole number counts by percent relative abundance to approximately 10,000 sequences per sample (9,914 - 10,061 after rounding.

Statistical analysis for graduates. Graduates were assessed separately from siblings and spouses to avoid potential interactions, and the graduate subset was not significantly different from other groups (PERMANOVA *P* Bray-Curtis *P* = 0.56, Jaccard *P* = 0.57, weighted UniFrac *P* = 0.33, unweighted UniFrac *P* = 0.24). Alpha-diversity was assessed with Shannon's diversity and Chao's richness calculated in mothur. Differences in alpha-metrics were assessed in R v3.3.2 [87] by linear regression with the Benjamini-Hochberg correction for multiple comparisons across each metric. Microbial beta-diversity was assessed for Bray-Curtis, Jaccard, weighted, and unweighted UniFrac metrics with results shown for unweighted UniFrac unless otherwise noted. Dietary beta- diversity

was assessed for Bray-Curtis and Jaccard metrics as well as corresponding nMDS axes calculated from habitual intake of specific sources of protein (N = 4), vegetables (N = 76), and fruits (N = 24) expressed as times consumed per week (protein), proportions of total types (all), and presence/absence of individual types (all). Differences in beta-diversity were tested with permutational analysis of variance (PERMANOVA, adonis) in the vegan package [88] with the Benjamini-Hochberg correction for multiple comparisons across each metric and a maximum of 5000 permutations. All variables were modeled using independent, univariate tests and dietary variables were additionally modeled using multivariate tests of all components (protein, vegetables, fruits). Co-variance of microbial and dietary beta metrics was measured using Mantel's test. The factors that associated with the microbiome in univariate models (*i.e.* age, sex, antibiotics, dietary protein, high blood sugar, and heart disease [57]) were adjusted for in regression models as potential confounders. Beta-diversity was visualized by non-metric multidimensional scaling (nMDS) plots with arrows from significant variables (PERMANOVA) fitted to the ordination using maximum correlation (envfit, vegan). All tests were assessed at significance $P < 0.05$ and trends $0.05 < P < 0.1$.

Statistical analysis for spouses and siblings. For the spouse and sibling similarity analysis, the unit of the observation is the pair (*i.e.* spouse, sibling, or unrelated pair defined below) and the variables used in the analysis are distance in individual measurements between the two members of the pair. Specifically, beta-diversity metrics were used to quantify the distance in microbial and overall diet whereas absolute difference were calculated to quantify the distance in all the other variables (*e.g.* age, sex, dietary protein). We sampled unrelated pairs from the data in order to compare the spouse or sibling pair with unrelated pairs. In particular, the unrelated individuals

cannot be siblings, spouses, or in-laws, and each unrelated pair will match the corresponding spouse or sibling pair in sex and antibiotics usage. Beta-diversity distances were compared among spouse, sibling, and unrelated pairs using linear regression while adjusting for the distance in age, sex, dietary protein, health conditions (if available). P-values were averaged across 1000 rounds of unrelated pair sampling. For closeness analysis, we removed age and sex from the model because the two variables are highly correlated with pair type (*i.e.* sibling/spouse pair can be accurately classified using the difference of the age or sex between the two members of the pair). For comparing OTU sharing among spouse and sibling within a family, we used mixed-effect models to account for family clustering. All tests were assessed at significance $P < 0.05$ and trends $0.05 < P < 0.01$.

CONTRIBUTIONS

Federico Rey and Pamela Herd designed the research and conceived the project. Julia Kemis, Robert Kerby and Thomas Sorenson completed sample processing and sequencing. Alberto Palloni contributed to the conceptual and analytic approach to conducting the analyses. Kimberly Dill-McFarland, Zheng-Zheng Tang and Guanhua Chen analyzed the data. Kimberly Dill-McFarland, Zheng-Zheng Tang, Guanhua Chen, Federico Rey and Pamela Herd contributed to the final manuscript.

ACKNOWLEDGEMENTS

The authors wish to thank members of the Rey laboratory for their insights and support. This work was supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture (2016-67017-24416, FER), the National Institutes for Aging, National Institutes for Health (AG041868, PH), the Center for the Demography of Health and Aging (AG017266, PH), and the Clinical and Translational Science Award (CTSA), National Institutes for Health National Center for Advancing Translational Sciences (NCATS) (UL1TR000427, ZZT). Additional support was provided by the Vice Chancellor for Research and Graduate Education (VCGRE) at the U. of Wisconsin-Madison.

REFERENCES

1. House, J. S., Landis, K. R. & Umberson, D. Social relationships and health. *Science (80).* **241,** 540–545 (1988).

2. Umberson, D., Crosnoe, R. & Reczek, C. Social relationships and health behavior across life course. *Annu. Rev. Sociol.* **36,** 139–157 (2010).

3. Robles, T. F. & Kiecolt-Glaser, J. K. The physiology of marriage: pathways to health. *Physiol. Behav.* **79,** 409–416 (2003).

4. McEwen, B. S. Brain on stress: How the social environment gets under the skin. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 17180–17185 (2012).

5. Yang, Y. C. *et al.* Social relationships and physiological determinants of longevity across the human life span. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 578–583 (2016).

6. Grice, E. A. & Segre, J. A. The human microbiome: Our second genome. *Annu. Rev. Genomics Hum. Genet.* **13,** 151–170 (2012).

7. Krautkramer, K. A. *et al.* Diet-microbiota interactions mediate global epigenetic programming in multiple host tissues. *Mol. Cell* **64,** 982–992 (2016).

8. Rieder, R., Wisniewski, P. J., Alderman, B. L. & Campbell, S. C. Microbes and mental health: A review. *Brain. Behav. Immun.* (2017). doi:10.1016/j.bbi.2017.01.016

9. Flint, H. J., Bayer, E. A., Rincon, M. T., Lamed, R. & White, B. A. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat Rev Micro* **6,** 121–131 (2008).

10. Lin, L. & Zhang, J. Role of intestinal microbiota and metabolites on gut homeostasis and human diseases. *BMC Immunol.* **18,** 2 (2017).

11. Gensollen, T., Iyer, S. S., Kasper, D. L. & Blumberg, R. S. How colonization by microbiota in early life shapes the immune system. *Science* **352,** 539–544 (2016).

12. Rodríguez, J. M. *et al.* The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb. Ecol. Health Dis.* **26,** 10.3402/mehd.v26.26050 (2015).

13. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473,** 174–180 457 (2011).

14. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. 459 *Science* **334,** 105–108 (2011).

15. Bonder, M. J. *et al.* The effect of host genetics on the gut microbiome. *Nat Genet* **48,** 1407–1412 (2016).

16. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505,** 559–563 (2014).

17. Panda, S. *et al.* Short-term effect of antibiotics on human gut microbiota. *PLoS One* **9,** e95476 (2014).

18. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486,** 222–227 (2012).

19. Faith, J. J. *et al.* The long-term stability of the human gut microbiota. *Science* **341,** 237439 (2013).

20. Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326,** 1694–1697 (2009).

21. Flint, H. J., Scott, K. P., Louis, P. & Duncan, S. H. The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol* **9,** 577–589 (2012).

22. Moeller, A. H. *et al.* Social behavior shapes the chimpanzee pan-microbiome. *Sci. Adv.* **2,** (2016).

23. Amato, K. R. *et al.* Patterns in gut microbiota similarity associated with degree of sociality among sex classes of a neotropical primate. *Microb. Ecol.* 1–9 (2017). 478 doi:10.1007/s00248-017-0938-6

24. Amaral, W. Z. *et al.* Social influences on Prevotella and the gut microbiome of young monkeys. *Psychosom. Med.* **Epub ahead,** (2017).

25. Tung, J. *et al.* Social networks predict gut microbiome composition in wild baboons. *Elife* **4,** e05224 (2015).

26. Amaral, W. Z., Lubach, G. R. & Coe, C. L. 114. Social influences on the gut microbiome of developing monkeys. *Brain. Behav. Immun.* **40,** (2014).

27. Bailey, M. T. & Coe, C. L. Maternal separation disrupts the integrity of the intestinal microflora in infant rhesus monkeys. *Dev. Psychobiol.* **35,** 146–155 (1999).

28. O'Mahony, S. M. *et al.* Early life stress alters behavior, immunity, and microbiota in rats: implications for irritable bowel syndrome and psychiatric illnesses. *Biol. Psychiatry* **65,** 263–267 (2009).

29. Bailey, M. T. *et al.* Exposure to a social stressor alters the structure of the intestinal microbiota: implications for stressor-induced immunomodulation. *Brain. Behav. Immun.* **25,** 397–407 (2011).

30. Archie, E. A. & Tung, J. Social behavior and the microbiome. *Curr. Opin. Behav. Sci.* **6,** 28–34 (2015).

31. Meehan, C. L. *et al.* Social networks, cooperative breeding, and the human milk microbiome. *Am. J. Hum. Biol.* e23131 (2018).

32. Münger, E., Montiel-Castro, A. J., Langhans, W. & Pacheco-López, G. Reciprocal interactions between gut microbiota and host social behavior. *Front. Integr. Neurosci.* **12,** 21 (2018).

33. Song, S. J. *et al.* Cohabiting family members share microbiota with one another and with their dogs. *Elife* **2,** e00458 (2013).

34. Ross, A. A., Doxey, A. C. & Neufeld, J. D. The skin microbiome of cohabiting couples. *mSystems* **2,** (2017).

35. Erwin G. Zoetendal Wilma M. Akkermans-van Vliet, J. Arjan G. M. de Visser, Willem M. 505 de Vos, A. D. L. A. The host genotype affects the bacterial community in the human gastronintestinal tract. *Microb. Ecol. Health Dis.* **13,** 129–134 (2001).

36. Grieneisen, L. E., Livermore, J., Alberts, S., Tung, J. & Archie, E. A. Group living and male dispersal predict the core gut microbiome in wild baboons. *Integr. Comp. Biol.* **57,** 770–785 (2017).

37. Levin, I. I. *et al.* Stress response, gut microbial diversity and sexual signals correlate with social interactions. *Biol. Lett.* **12,** 20160352 (2016).

38. Lombardo, M. P. Access to mutualistic endosymbiotic microbes: an underappreciated benefit of group living. *Behav. Ecol. Sociobiol.* **62,** 479–497 (2008).

39. Li, H. *et al.* Pika population density is associated with the composition and diversity of gut microbiota. *Front. Microbiol.* **7,** 758 (2016).

40. Manichanh, C. *et al.* Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55,** 205–211 (2006).

41. Scher, J. U. *et al.* Decreased Bacterial Diversity Characterizes an Altered Gut Microbiota in Psoriatic Arthritis and Resembles Dysbiosis of Inflammatory Bowel Disease. *Arthritis Rheumatol. (Hoboken, N.J.)* **67,** 128–139 (2015).

42. de Goffau, M. C. *et al.* Fecal Microbiota Composition Differs Between Children With β-Cell Autoimmunity and Those Without. *Diabetes* **62,** 1238–1244 (2013).

43. Wang, M. *et al.* Reduced diversity in the early fecal microbiota of infants with atopic eczema. *J. Allergy Clin. Immunol.* **121,** 129–134 (2008).

44. Schippa, S. *et al.* A distinctive 'microbial signature' in celiac pediatric patients. *BMC Microbiol.* **10,** 175 (2010).

45. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457,** 480 52 (2008).

46. Lambeth, S. M. *et al.* Composition, Diversity and Abundance of Gut Microbiome in Prediabetes and Type 2 Diabetes. *J. diabetes Obes.* **2,** 1–7 (2015).

47. Menni, C. *et al.* Gut microbial diversity is associated with lower arterial stiffness in women. *Eur. Heart J.* **39,** 2390–2397 (2018).

48. Kim, J. & Waite, L. J. Relationship quality and shared activity in marital and cohabiting dyads in the National Social Life, Health, and Aging Project, Wave 2. *Journals Gerontol. Ser. B Psychol. Sci. Soc. Sci.* **69,** S64–S74 (2014).

49. Lewin, A. C. Health and Relationship Quality Later in Life: A Comparison of Living Apart Together (LAT), First Marriages, Remarriages, and Cohabitation. *J. Fam. Issues* **38,** 1754–1774 (2016).

50. Wong, J. S. & Waite, L. J. Marriage, social networks, and health at older ages. *J. Popul. Ageing* **8,** 7–25 (2015).

51. Levine, M. E., Cole, S. W., Weir, D. R. & Crimmins, E. M. Childhood and later life stressors and increased Inflammatory gene expression at older ages. *Soc. Sci. Med.* **130,** 16–22 (2015).

52. Seeman, M., Merkin, S., Karlamangla, A., Koretz, B. & Seeman, T. Social status and biological dysregulation: The "status syndrome" and allostatic load. *Soc. Sci. Med.* **118,**143–151 (2014).

53. Das, A. Psychosocial distress and inflammation: Which way does causality flow? *Soc. Sci. Med.* **170,** 1–8 (2016).

54. Herd, P., Carr, D. & Roan, C. Cohort profile: Wisconsin longitudinal study (WLS). *Int. J. Epidemiol.* **43,** 34–41 (2014).

55. Herd, P. *et al.* The influence of social conditions across the life course on the human gut microbiota: a pilot project with the Wisconsin Longitudinal Study. *J Gerontol B Psychol Sci Soc Sci* (2017). doi:10.1093/geronb/gbx029

56. Scott, K. P., Gratz, S. W., Sheridan, P. O., Flint, H. J. & Duncan, S. H. The influence of diet on the gut microbiota. *Pharmacol. Res.* **69,** 52–60 (2013).

57. Borgo, F. *et al.* Body Mass Index and sex affect diverse microbial niches within the gut. *Front. Microbiol.* **9,** 213 (2018).

58. Knight, R. *et al.* The microbiome and human biology. *Annu. Rev. Genomics Hum. Genet.* **18,** 65–86 (2017).

59. Pajarillo, E. A. B., Chae, J. P., Kim, H. B., Kim, I. H. & Kang, D.-K. Barcoded pyrosequencing-based metagenomic analysis of the faecal microbiome of three purebred pig lines after cohabitation. *Appl. Microbiol. Biotechnol.* **99,** 5647–5656 (2015).

60. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489,** 220–230 (2012).

61. Levine, J. M. & D'Antonio, C. M. Elton revisited: a review of evidence linking diversity and invasibility. *Oikos* **87,** 15–26 (1999).

62. Koenig, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 4578–4585 (2011).

63. Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. Development of the human infant intestinal microbiota. *PLoS Biol* **5,** e177 (2007).

64. Dicksved, J. *et al.* Molecular fingerprinting of the fecal microbiota of children raised according to different lifestyles. *Appl. Environ. Microbiol.* **73,** 2284–2289 (2007).

65. Ege, M. J. *et al.* Exposure to environmental microorganisms and childhood asthma. *N. Engl. J. Med.* **364,** 701–709 (2011).

66. House, J. S. *et al.* Early-life farm exposures and adult asthma and atopy in the Agricultural Lung Health Study. *J. Allergy Clin. Immunol.* (2017). doi:10.1016/j.jaci.2016.09.036

67. Wlasiuk, G. & Vercelli, D. The farm effect, or: when, what and how a farming environment protects from asthma and allergic disease. *Curr. Opin. Allergy Clin. Immunol.* **12,** 461–466 (2012).

68. Stewart, J. A., Chadwick, V. S. & Murray, A. Investigations into the influence of host genetics on the predominant eubacteria in the faecal microflora of children. *J. Med. Microbiol.* **54,** 1239–1242 (2005).

69. Tims, S. *et al.* Microbiota conservation and BMI signatures in adult monozygotic twins. *ISME J.* **7,** 707–717 (2013).

70. Goodrich, J. K. *et al.* Human genetics shape the gut microbiome. *Cell* **159,** 789–799 (2014).

71. Cao, X. Intestinal inflammation induced by oral bacteria. *Science (80-.).* **358,** 308 LP-309 (2017).

72. Kort, R. *et al.* Shaping the oral microbiota through intimate kissing. *Microbiome* **2,** 41 (2014).

73. Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124,** 837–848 (2017).

74. Everard, A. *et al.* Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. *Proc. Natl. Acad. Sci.* **110,** 9066–9071 (2013).

75. Arboleya, S., Watkins, C., Stanton, C. & Ross, R. P. Gut Bifidobacteria populations in human health and aging. *Front. Microbiol.* **7,** 1204 (2016).

76. Malinen, E. *et al.* Association of symptoms with gastrointestinal microbiota in irritable bowel syndrome. *World J. Gastroenterol.* **16,** 4532–4540 (2010).

77. Kang, S. *et al.* Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflamm. Bowel Dis.* **16,** 2034–2042 (2010).

78. Babudieri, S. *et al.* Diarrhoea associated with toxigenic Clostridium spiroforme. *J. Infect.* **12,** 278–279 (1986).

79. Borriello, S. P. Clostridial disease of the gut. *Clin. Infect. Dis.* **20,** S242–S250 (1995).

80. Zhang, X. *et al.* Human gut microbiota changes reveal the progression of glucose intolerance. *PLoS One* **8,** e71108 (2013).

81. Ciubotaru, I., Green, S. J., Kukreja, S. & Barengolts, E. Significant differences in fecal microbiota are associated with various stages of glucose tolerance in African American male veterans. *Transl. Res.* **166,** 401–411 (2015).

82. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* (2013). doi:10.1128/aem.01043-13

83. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-Independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75,** 7537–7541 (2009).

84. Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35,** 7188–7196 (2007).

85. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27,** 2194–2200 (2011).

86. DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72,** 5069–5072 (2006).

87. R Core Team. R: A language and environment for statistical computing. (2017).

88. Oksanen, J. *et al.* vegan: community ecology package. (2015). 625

FIGURES

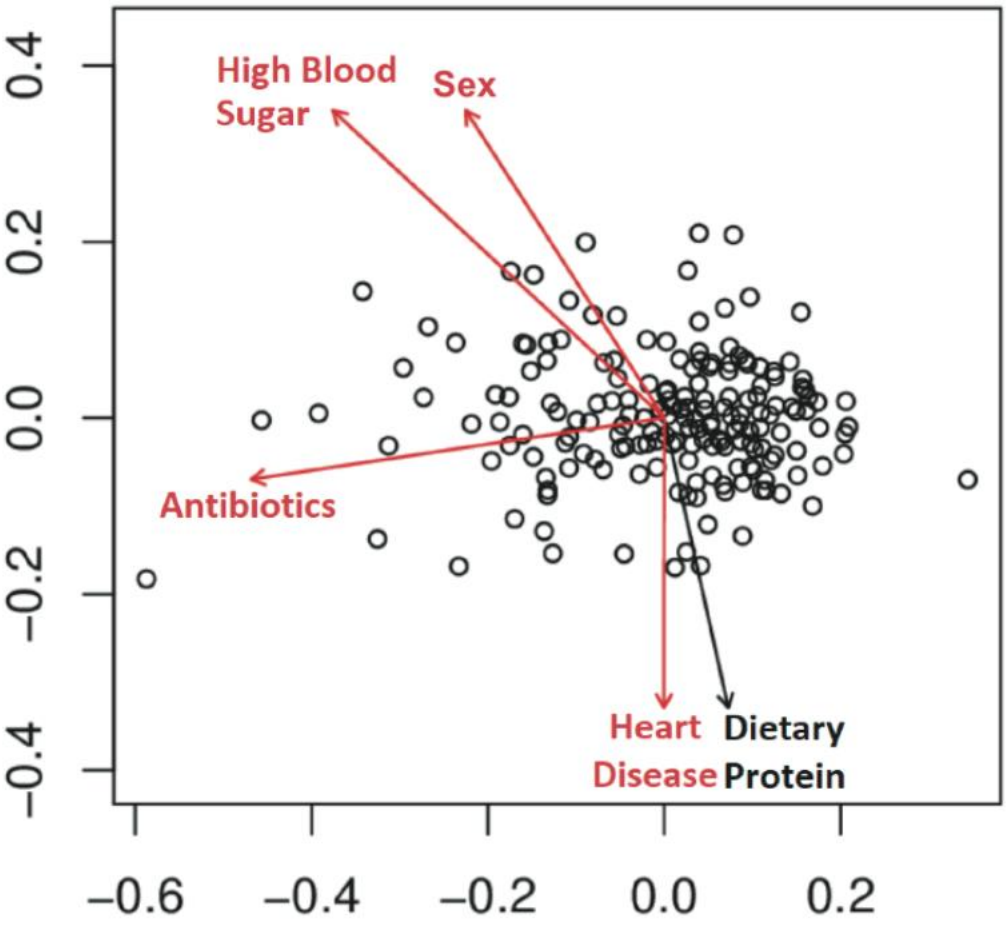


Figure B.1. Factors associated with the overall fecal microbiota. Non-metric multidimensional scaling (nMDS) of unweighted UniFrac for all graduates (N = 179). Variables found to be significant (PERMANOVA $P < 0.05$, red) and trends ($0.05 < P < 0.1$, black) are shown as fitted arrows. Arrows point toward increasing values (dietary protein), toward affirmative responses (high blood sugar, antibiotics, heart disease), or from male to female (sex).

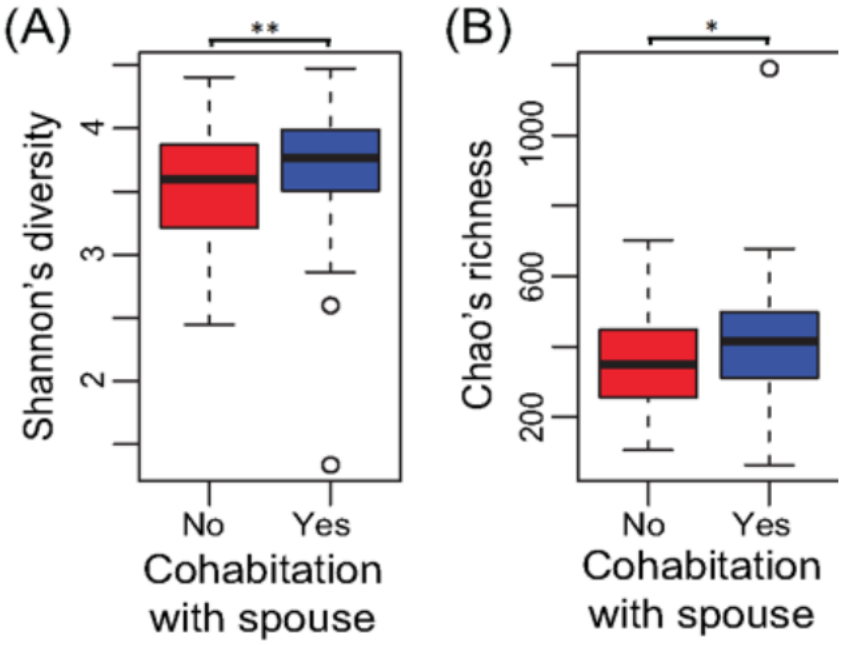


Figure B.2. Cohabitation is associated with increased alpha-diversity. Boxplots of (A) Shannon's diversity and (B) Chao's richness of graduates that are (blue) or are not (red) cohabitating with a spouse or partner. All spouses/partners were cohabitating while all non-cohabitating individuals were unmarried. **P < 0.01, *P < 0.05.

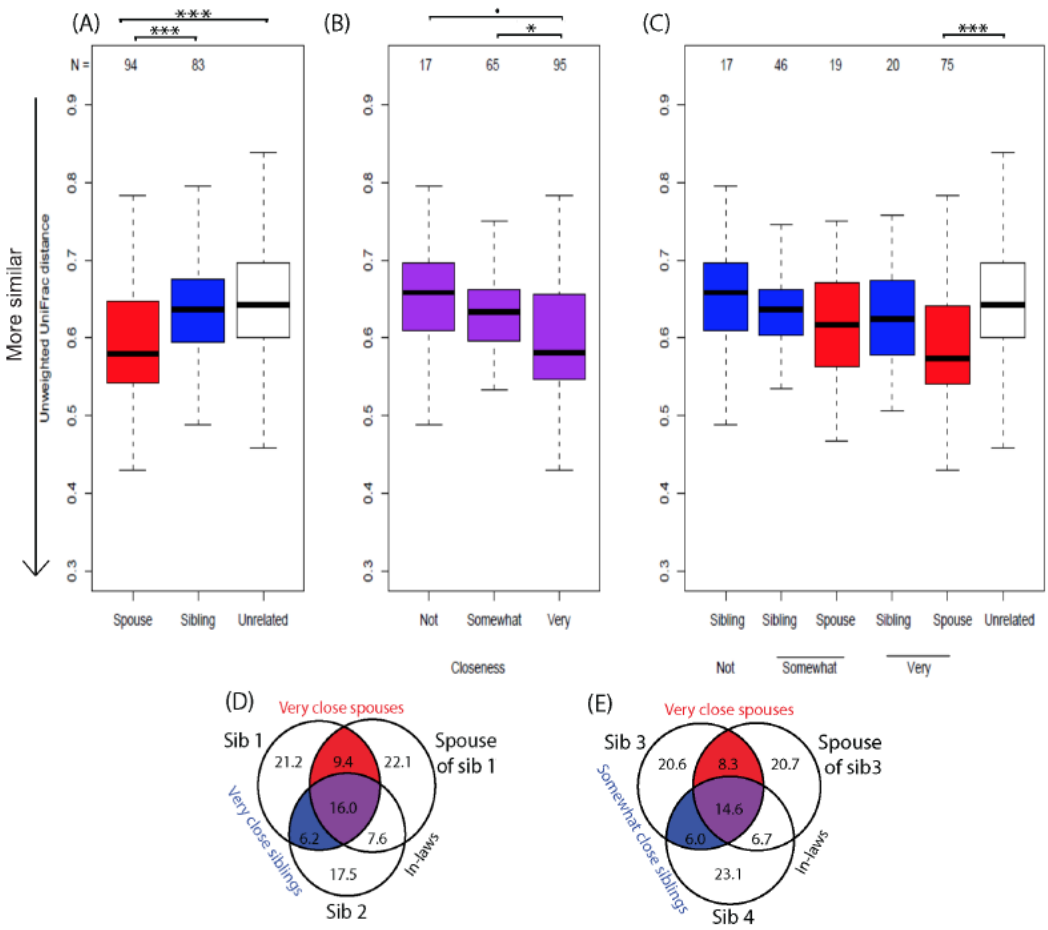


Figure B.3. Microbial sharing in spouse and sibling relationships. Unweighted UniFrac distances of (A) spouse, sibling, and unrelated pairs, (B) spouses and siblings grouped by relationship closeness, and (C) spouses and siblings separated by relationship closeness. Groups (A) were compared in linear regression model adjusting for potential confounders (e.g. age, sex, diet, health conditions). P-values were averaged across 1000 rounds of unrelated pair sampling. Closeness groups (B,C) were compared in linear regression models adjusting for potential confounders. (D,E) Average percentages of shared OTUs within family groups including a related spouse and sibling pair. Families included those with both very close spouses and siblings (D, N = 12) and those with very close spouses and somewhat close siblings (E, N = 17). Percentages are of the total number of OTUs across all three individuals, and circle sizes are proportional to total percentages represented. ***P < 0.001, **P < 0.01, *P < 0.05, •P< 0.1.

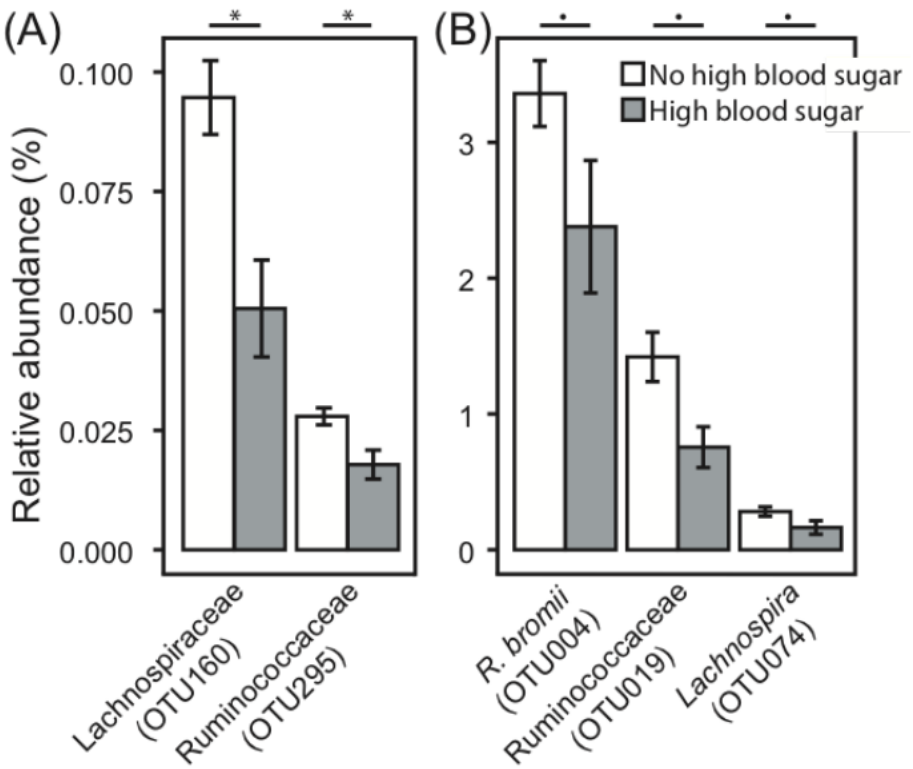


Figure B.4. Percent relative abundance of OTUs that are commonly shared between spouses and that differed between those with (grey) and without (white) high blood sugar. (A) Low abundance and (B) more highly abundant OTUs. Means with standard error bars are shown. Kruskal-Wallis FDR *P < 0.05, •P < 0.1

APPENDIDX C: Chapter 3 Supplemental Results

SUPPLEMENTAL RESULTS

Here, we report additional findings from the analysis of the Diversity Outbred (DO) mice presented in Chapter 3. We found additional QTL and associations among the measured bile acid, microbial and clinical traits that were interesting, but were not included in the version set for publication. Notably, these supplemental results corroborate findings from other studies and provide support for further mechanistic studies.

Correlations between bile acids, clinical traits and gut microbiota abundance

Correlation analysis identified several significant associations between the microbial and clinical weight traits after FDR correction (FDR < 0.05) (Table 4.4, *see Chapter 3*). The significant correlations between weight and microbial traits were attributed to 15 distinct microbial taxa from the Actinobacteria, Bacteroidetes and Firmicutes phyla. ~58% of these associations could be attributed to exact sequence variants (ESVs) assigned to the Lachnospiraceae family and another ~19% to ESVs classified to the S24-7 family. Additionally, the *Ruminococcus* genus and fat pad weight were negatively correlated, and ESVs classified to the *Adlercreutzia* genus and body weight at 14 weeks were positively correlated.

Additionally, we identified significant associations between bile acids and body weight. Body weight over time was inversely correlated with plasma levels of deoxycholic acid (DCA), taurodeoxycholic acid (TDCA) and taurocholic acid (TCA) (Table 4.4). Conversely, cecal levels of muricholic acid (MCA) and ursodeoxycholic acid (UDCA) were positively correlated with body, liver and heart weight. These associations were surprising since elevated levels of DCA have been associated with weight gain and insulin resistance in humans (Brufau et al., 2010; Cariou et al., 2011; Gu et al., 2017).

Four of the ESVs from the Lachnospiraceae family were significantly associated with body weight and fat pad weight. Interestingly, one of these Lachnospiraceae ESVs was positively correlated with the weight traits, while the other three were negatively correlated, indicating the metabolic effects of Lachnospiraceae are possibly genus or even strain dependent. Disparate associations of different members of the Lachnospiraceae on weight traits has been shown in previous mouse genetic studies. For example, one study fed ~110 inbred strains of mice from the Hybrid Mouse Diversity Panel a high-fat high-sucrose diet and found that two taxa from the Lachnospiraceae family were positively associated with obesity and metabolic traits (Org et al., 2015). On the other hand, two separate studies using the eight DO founder strains fed different diets found that Lachnospiraceae family was negatively correlated with body weight (Kreznar et al., 2017; O'Connor et al., 2014).

Despite these significant correlations, QTLs for only two of the Lachnospiraceae ESVs overlapped with weight QTL. These two Lachnospiraceae traits represented the two patterns observed where one was positively correlated and the other negatively correlated with weight traits. QTLs for liver weight and one of the Lachnospiraceae taxa that negatively correlated with liver weight overlapped on chr 9 at ~65-66 Mbp. Furthermore, QTL for the Lachnospiraceae taxa positively correlated with weight traits associated to the same position of the genome as QTLs for body weight on chr 4 at ~150 Mbp. For both of these examples, the microbial and weight traits were driven by different founder haplotypes indicating the co-mapping traits are likely not causally related.

Correlation analysis of the microbial taxa comprising the CMM was also used to provide insight into microbiota community structure in the DO mice (Table 4.4). A positive correlation was found between two members of the small intestine microbiome, *Turicibacter* and

Peptostreptococcaeae (r = 0.33, p = 0.026). A similar correlative relationship was detected in the founder strains (r = 0.65, p = 0.003). This finding is consistent with Goodrich et al., who found these taxa strongly correlated with one another in humans (r = 0.66) (Goodrich et al., 2016). Additionally, both taxa are consistently identified as heritable in humans and mice, and associate to regions of the mouse genome (Benson et al., 2010; Goodrich et al., 2016; O'Connor et al., 2014). The correlation between *Turicibacter* and Peptostreptococcaeae particularly notable since we observe both taxa co-mapping with plasma bile acids and significantly correlating. However, we do not observe QTL for these taxa mapping to the same position. Additionally, both taxa are capable of bile acid metabolism. As shown in Chapter 3, *Turicibacter* efficiently deconjugates bile acids. *Peptostreptococcus productus*, a member of the Peptostreptococcaceae family, has 3α-, 3β-, and 7β-hydroxysteroid dehydrogenases and is capable of oxidation and epimerization of bile acids (Edenharder et al., 1989). These microbes have complimentary bile acid metabolism capabilities as *Turicibacter* provides necessary deconjugation activity for further transformations by members of the Peptostreptococcaeae family. Thus, their co-occurrence may provide a fitness advantage for small intestine colonization. Bile acids must be deconjugated prior to epimerization, so Peptostreptococcaeae may associate with *Turicibacter* in order to utilize this metabolic capability. The consistency in these findings as well as the different bile acid metabolism capabilities warrant further investigation and may provide insight into community dynamics and bile acid metabolism in the small intestine.

Adlercreutzia associates to immune genes on chromosome 10

We identified QTL for two different taxa classified to the *Adlercreutzia* genus that mapped to chr 10 at ~118-119 Mbp (Figure C.1A). *Adlercreutzia* are gram-positive organisms classified to the Coriobacteriaceae family in the Actinobacteria phylum. These two QTL were of particular

interest because Benson et al., previously found a Coriobacteriaceae peak at the identical position with significant dominance effects of the B6 allele (Benson et al., 2010). They also identified *Lactococcus* QTL that overlapped at this locus. While Lactococcus did not map to this position in our study, we did replicate the significant positive correlation between the abundances of Coriobacteriaceae and Lactococcus ($r = 0.353$, $p < 0.0001$), providing additional evidence to support host genetic influence on shaping the abundance of these taxa. Additionally, there are several strong candidate genes under these QTL relating to host immune response and host regulation of gram-positive organisms (Figure C.1B). Strong candidate genes at these loci are the two primary murine lysozyme genes, *Lyz1* and *Lyz2* (Markart et al., 2004). Additionally, the same interval also contains the genes encoding *Irak3*, *IFN-y* and *IL-22*, which play a role in mucosal immunity (Kjerrulf et al., 1997; Nakayama et al., 2004; Zheng et al., 2008).

Upon further investigation, we found that the two QTLs were driven by distinct allele patterns. For one QTL, denoted as *Adlercreutzia sp.* 1, the B6 and 129 founder haplotypes were associated with higher levels of *Adlercreutzia*, while the AJ haplotype was associated with lower levels of *Adlercreutzia* (Figure C.1C). The converse pattern was observed for the second *Adlercreutzia* QTL (*Adlercreutzia sp.* 2), where the AJ haplotype had a positive association and the B6 and 129 haplotypes had a negative association with levels of that strain of *Adlercreutzia* (Figure C.1A). It appears that the B6 haplotype for this locus is associated with increased abundance of one Adlercreutzia species and associated with a decreased abundance of another (Figure C.1E).

These founder allele effect patterns suggest that there are at least two distinct variants that drive *Adlercreutzia* abundance and these QTL may reveal insight into how host genetic variants select for specific species or strains of closely related bacteria. Furthermore, we found a significant

negative correlation between these two strains of *Adlercreutzia* (r = -0.261, p = 3.35e-06), providing additional evidence that host genotype at this locus determines which *Adlercreutzia* strain is capable of colonizing the intestine.

The identification of this QTL in separate studies, along with the interesting genes under the QTLs warrant further investigation.

Christensenellaceae and body weight traits associate to chromosome 1

Christensenellaceae is a gram-negative bacterium that previously identified as a highly heritable in geographically distinct groups of humans (Goodrich et al., 2014b; Lim et al., 2017; Turpin et al., 2016). We identified a QTL for Christensenellaceae spanning 5.5 Mbp on chr 1 at ~59 Mbp (Figure C.2A). A candidate gene within the QTL interval is *Casp8*, which is a key regulator of the host innate immune response and plays a central role in inflammasome-mediated cell death (Figure C.2B). In the intestine, *Casp8* is activated by microbial recognition receptors, such as TLR4 in response to LPS (Monie and Bryant, 2015). It also a known transcriptional regulator of the *Il1b* gene (Gurung et al., 2014), which has previously been shown to be influenced by the composition of the microbiota (Seo et al., 2015).

Genome analysis identified a missense variant (rs32803726) within a coding region of *Casp8* (Q13R) driven by the CAST, PWK, NOD and WSB haplotypes. This missense variant may affect protein structure and have functional consequences on the transcribed protein, leading to changes in intestinal immune environment. Consistent with this notion, Christensenellaceae abundance varied by genotype (p = 0.0048; one-way Kruskal-Wallis) and was significantly greater in DO mice with this variant than in mice with two copies of the B6 allele (0-1 p = 0.0217, 0-2 p = 0.0099; 1-2 p = 0.1210, Wilcoxon test with Benjamini-Hochberg correction) (Figure C.2C). This same pattern was also observed at the peak SNP, where the abundance of Christensenellaceae was

greater in DO mice with this variant than in mice without, but no significant differences were found between genotypes (0-1 p = 0.636, 0-2 p = 0.199; 1-2 p = 0.069) (Figure C.2D). To further validate these findings, we compared the abundance of Christensenellaceae in the founder strains to the estimated coefficients in the DO mice. We found these estimated coefficients closely resembled the abundance of Christensenellaceae in the founder strains, where CAST and PWK harbored the highest abundance of this bacteria (Figure C.2E-F).

Christensenellaceae has been negatively associated with BMI and visceral adiposity in humans (Beaumont et al., 2016; Goodrich et al., 2014b). It was also shown that administration of *Christensenella minuta* attenuated weight gain and total adiposity in germ-free mice colonized with feces from an obese human donor (Goodrich et al., 2014b). Interestingly, QTL for body weight at 14 weeks and sacrifice associate to the same position on chr 1 (Figure C.2A). However, Chirstensenellacae abundance did not correlate with body weight at 14 weeks (r = -0.09, p = 0.206), body weight at sacrifice (r = -0.06, p = 0.406), or fat pad mass per gram body weight (r = 0.12, p = 0.129) (Figure C.2G-I). Furthermore, the body weight QTLs were driven by the 129 and WSB haplotypes, providing additional evidence that the Christensenellaceae and weight QTLs are unlikely to be related. In our study, the relationship between Christensenellaceae and weight may be masked by the HF/HS diet, since the previous work done in mice used animals fed a chow diet. To our knowledge, this is the first instance linking differences in Christensenellaceae to host genetics in mice, thus providing additional insight into the heritability of Christensenellaceae.

Bile acid QTL are found in multiple "hot spots"

QTL analysis revealed several metabolite hotspots where multiple bile acid traits co-map within a <20 Mbp window. These hotspots may indicate traits that interact or are have highly correlated levels. They may also be a consequence of the occurrence of pleiotropic or regulatory

genes controlling multiple metabolite traits. Notable bile acid hotspots were detected on chrs 3, 13, and 18, where at least 5 bile acid traits associated within a <15 Mbp genomic interval. The hotspot on chr 3 includes multiple plasma bile acids driven by the NOD haplotype and is previously discussed. Both plasma and cecal bile acid QTLs are found at the hotspots on chrs 13 and 18.

The hotspot on chr 13 at ~108-119 Mbp was associated with 4 cecal and 1 plasma bile acid QTL (Figure C.3A). All overlapping QTL at this hotspot were for secondary bile acids. These QTL have varying genetic architecture, which suggests more than one causal locus. QTL for cecal levels of DCA, allocholic acid (ACA) and 12-ketolithocholic acid (12-KLCA) were all positively associated with the PWK and NZO alleles and negatively associated with the CAST allele (Figure C.3B-D). Furthermore, these three cecal bile acids were also highly correlated with one another (Table C.1). Mediation analysis found a correlative relationship between DCA and both 12-KLCA and ACA. However, no causal relationship was detected between 12-KLCA and ACA. Interestingly, the mediation analysis captured the relationship of these bile acids. For example, both ACA and DCA are respectively derived from 5α- and 7α-epimerization of CA by gut microbes. 12-KLCA is produced from metabolism of DCA. Thus, we see correlative relationships between bile acids that are either derived from the same metabolite. ACA and 12-KLCA do not have a relationship because they are not directly linked to one another.

Since all the co-mapping metabolites were secondary bile acids, we looked for overlapping microbial QTL that may be causal for levels of these bile acids. Several microbial traits associated to this locus including the Mogibacteriaceae family and taxa classified to the *Adlercreutzia* and *Oscillospira* genera. The Mogibacteriaceae QTL had a positive association with the PWK allele, but mediation analysis did not show a causal relationship. The other microbial QTL did not share genetic architecture with any of the other bile acid QTL.

The largest hotspot was on chr 18 at 32 - 46 Mbp and included 9 bile acid traits, most of which were secondary bile acids except for plasma TCA and cecal tauromuricholic acid (TMCA) (Figure C.4F). This hotspot contains a variety of bile acid traits including conjugated and unconjugated, plasma and cecal, and primary and secondary. In general, there is no shared founder effects pattern shared among these traits. The underlying genetic architecture of these QTL is varied and complicated, which some founder haplotypes are positively associated with some traits and negatively associated with others. For instance, plasma TCA levels are positively and plasma UCA levels are negatively associated with the CAST haplotype. The variability in the underlying genetic architecture of the plasma bile acids is greater than that seen the cecal bile acids. All four overlapping cecal metabolite QTL have a negative association with the NOD and/or CAST haplotypes and are significantly correlated with each other. The variability seen at this hotspot suggests multiple closely linked loci, as opposed to a single pleiotropic locus.

Corroboration of previous human and mouse genetics studies

Although recent studies show environment contributes more to the variability among gut microbiota composition than genetics (Falony et al., 2016; Rothschild et al., 2018; Zhernakova et al., 2016), there are consistencies among different host organisms and geographically discrete populations indicate specific taxa and related traits are under the influence of the host genome. Our results in the DO population corroborate several of these key findings for both microbial and clinical traits. These shared findings can be followed up for mechanistic experiments.

We observed the strongest associations to the host genome with members of the Firmicutes phyla, including unknown members of the Clostridiales order, the Lachnospiraceae, Christensenellacae and S24-7 families, the *Turicibacter* and *Coprococcus* genera, as well as the species *Akkermansia muciniphila* and *Ruminococcus gnavus*. These taxa have consistently been

identified in multiple studies as either highly heritable or associating to positions on the host genome (Benson et al., 2010; Davenport et al., 2015; Leamy et al., 2014; McKnite et al., 2012; Org et al., 2015; Wang et al., 2016). Furthermore, our study replicated correlations between taxa including the Peptostreptococcaeae and Turicibacteraceae families (Goodrich et al., 2016), which may give insight into microbial dynamics that govern bile acid profiles.

Although the majority of these shared taxa seen in our study did not map to the same loci as in previous studies, we did find several microbial taxa and clinical traits that mapped to the same position of the mouse genome as in previous studies. We did find several clinical QTL in the DO population that co-mapped with clinical QTL previously identified in the HMDP population. For example, we found a QTL for body weight at 14 weeks on chr 2 at 135.2 that overlaps with a percent body fat increase QTL between 138.9 - 139.4 Mbp (Parks et al., 2013). We also found a QTL for fat pad weight on chr 7 at ~40 Mbp that falls within the same confidence interval as a HMDP QTL for triglyceride (TG) gonadal fat (Org et al., 2015). Additionally, QTL for taxa classified to the Coriobacteraceae family mapped to chr 10 between ~116 – 120 Mbp in our study and in an advanced intercross line used by Benson et al. (Benson et al., 2010) (Figure C.1A). However, the majority of these shared taxa seen in our study and previous analysis did not map to the same position.

Given the causal contribution of gut microbiota and obesity/metabolic disease, we were surprised to find few instances of overlapping microbial and clinical QTL. This was especially surprising given the overlap between microbial and obesity-related traits seen in other studies (Leamy et al., 2014; Org et al., 2015; Parks et al., 2013). The lack of congruence may be a result of the complexity of each trait or due to the different genetic background of the study populations as well as other factors including diet, age, and experimental design. Our analysis found co-

mapping of body weight with Christensenellaceae and Lachnospiraceae families. However, these co-mapping traits did not share the same founder haplotype effects and did not show a causal relationship as determined by mediation analysis. Therefore, we hypothesize that the microbial and clinical traits are the result of closely linked, but different loci.

Given the high degree of variability in the gut microbiome across subjects and host organisms, these instances of congruence between studies argues that there are specific taxa responsive to host genotype that may warrant follow-up investigation. Our work with the DO population provides an approach to validate these associations.

REFERENCES

Beaumont, M., Goodrich, J.K., Jackson, M.A., Yet, I., Davenport, E.R., Vieira-Silva, S., Debelius, J., Pallister, T., Mangino, M., Raes, J., et al. (2016). Heritable components of the human fecal microbiome are associated with visceral fat. Genome Biol. *17*, 189.

Benson, A.K., Kelly, S.A., Legge, R., Ma, F., Low, S.J., Kim, J., Zhang, M., Oh, P.L., Nehrenberg, D., Hua, K., et al. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. Proc. Natl. Acad. Sci. U. S. A. *107*, 18933–18938.

Brufau, G., Stellaard, F., Prado, K., Bloks, V.W., Jonkers, E., Boverhof, R., Kuipers, F., and Murphy, E.J. (2010). Improved glycemic control with colesevelam treatment in patients with type 2 diabetes is not directly associated with changes in bile acid metabolism. Hepatology *52*, 1455–1464.

Cariou, B., Chetiveaux, M., Zaïr, Y., Pouteau, E., Disse, E., Guyomarc'h-Delasalle, B., Laville, M., and Krempf, M. (2011). Fasting plasma chenodeoxycholic acid and cholic acid concentrations are inversely correlated with insulin sensitivity in adults. Nutr. Metab. (Lond). *8*, 48.

Davenport, E.R., Cusanovich, D.A., Michelini, K., Barreiro, L.B., Ober, C., and Gilad, Y. (2015). Genome-Wide Association Studies of the Human Gut Microbiota. PLoS One *10*, e0140301.

Edenharder, R., Pfützner, A., and Hammann, R. (1989). Characterization of NAD-dependent 3 alpha- and 3 beta-hydroxysteroid dehydrogenase and of NADP-dependent 7 beta-hydroxysteroid dehydrogenase from Peptostreptococcus productus. Biochim. Biophys. Acta *1004*, 230–238.

Falony, G., Joossens, M., Vieira-Silva, S., Wang, J., Darzi, Y., Faust, K., Kurilshikov, A., Bonder, M.J., Valles-Colomer, M., Vandeputte, D., et al. (2016). Population-level analysis of gut microbiome variation. Science *352*, 560–564.

Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J.T., et al. (2014). Human genetics shape the gut microbiome. Cell *159*, 789–799.

Goodrich, J.K., Davenport, E.R., Beaumont, M., Jackson, M.A., Knight, R., Ober, C., Spector, T.D., Bell, J.T., Clark, A.G., and Ley, R.E. (2016). Genetic Determinants of the Gut Microbiome in UK Twins. Cell Host Microbe *19*, 731–743.

Gu, Y., Wang, X., Li, J., Zhang, Y., Zhong, H., Liu, R., Zhang, D., Feng, Q., Xie, X., Hong, J., et al. (2017). Analyses of gut microbiota and plasma bile acids enable stratification of patients for antidiabetic treatment. Nat. Commun. *8*, 1785.

Gurung, P., Anand, P.K., Malireddi, R.K.S., Vande Walle, L., Van Opdenbosch, N., Dillon, C.P., Weinlich, R., Green, D.R., Lamkanfi, M., and Kanneganti, T.-D. (2014). FADD and caspase-8

mediate priming and activation of the canonical and noncanonical Nlrp3 inflammasomes. J. Immunol. *192*, 1835–1846.

Kjerrulf, M., Grdic, D., Ekman, L., Schön, K., Vajdy, M., and Lycke, N.Y. (1997). Interferon-gamma receptor-deficient mice exhibit impaired gut mucosal immune responses but intact oral tolerance. Immunology *92*, 60–68.

Kreznar, J.H., Keller, M.P., Traeger, L.L., Rabaglia, M.E., Schueler, K.L., Stapleton, D.S., Zhao, W., Vivas, E.I., Yandell, B.S., Broman, A.T., et al. (2017). Host Genotype and Gut Microbiome Modulate Insulin Secretion and Diet-Induced Metabolic Phenotypes. Cell Rep. *18*, 1739–1750.

Leamy, L.J., Kelly, S.A., Nietfeldt, J., Legge, R.M., Ma, F., Hua, K., Sinha, R., Peterson, D.A., Walter, J., Benson, A.K., et al. (2014). Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. Genome Biol. *15*, 552.

Lim, M.Y., You, H.J., Yoon, H.S., Kwon, B., Lee, J.Y., Lee, S., Song, Y.-M., Lee, K., Sung, J., and Ko, G. (2017). The effect of heritability and host genetics on the gut microbiota and metabolic syndrome. Gut *66*, 1031–1038.

Markart, P., Faust, N., Graf, T., Na, C.-L., Weaver, T.E., and Akinbi, H.T. (2004). Comparison of the microbicidal and muramidase activities of mouse lysozyme M and P. Biochem. J. *380*, 385–392.

McKnite, A.M., Perez-Munoz, M.E., Lu, L., Williams, E.G., Brewer, S., Andreux, P.A., Bastiaansen, J.W.M., Wang, X., Kachman, S.D., Auwerx, J., et al. (2012). Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. PLoS One *7*, e39191.

Monie, T.P., and Bryant, C.E. (2015). Caspase-8 functions as a key mediator of inflammation and pro-IL-1β processing via both canonical and non-canonical pathways. Immunol. Rev. *265*, 181–193.

Nakayama, K., Okugawa, S., Yanagimoto, S., Kitazawa, T., Tsukada, K., Kawada, M., Kimura, S., Hirai, K., Takagaki, Y., and Ota, Y. (2004). Involvement of IRAK-M in peptidoglycan-induced tolerance in macrophages. J. Biol. Chem. *279*, 6629–6634.

O'Connor, A., Quizon, P.M., Albright, J.E., Lin, F.T., and Bennett, B.J. (2014). Responsiveness of cardiometabolic-related microbiota to diet is influenced by host genetics. Mamm. Genome *25*, 583–599.

Org, E., Parks, B.W., Joo, J.W.J., Emert, B., Schwartzman, W., Kang, E.Y., Mehrabian, M., Pan, C., Knight, R., Gunsalus, R., et al. (2015). Genetic and environmental control of host-gut microbiota interactions. Genome Res. *25*, 1558–1569.

Parks, B.W., Nam, E., Org, E., Kostem, E., Norheim, F., Hui, S.T., Pan, C., Civelek, M., Rau, C.D., Bennett, B.J., et al. (2013). Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. Cell Metab. *17*, 141–152.

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P.I., Godneva, A., Kalka, I.N., Bar, N., et al. (2018). Environment dominates over host genetics in shaping human gut microbiota. Nature *555*, 210–215.

Seo, S.-U., Kamada, N., Muñoz-Planillo, R., Kim, Y.-G., Kim, D., Koizumi, Y., Hasegawa, M., Himpsl, S.D., Browne, H.P., Lawley, T.D., et al. (2015). Distinct Commensals Induce Interleukin-1β via NLRP3 Inflammasome in Inflammatory Monocytes to Promote Intestinal Inflammation in Response to Injury. Immunity *42*, 744–755.

Turpin, W., Espin-Garcia, O., Xu, W., Silverberg, M.S., Kevans, D., Smith, M.I., Guttman, D.S., Griffiths, A., Panaccione, R., Otley, A., et al. (2016). Association of host genome with intestinal microbial composition in a large healthy cohort. Nat. Genet.

Wang, J., Thingholm, L.B., Skiecevičienė, J., Rausch, P., Kummen, M., Hov, J.R., Degenhardt, F., Heinsen, F.-A., Rühlemann, M.C., Szymczak, S., et al. (2016). Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. Nat. Genet. *48*, 1396–1406.

Zheng, Y., Valdez, P.A., Danilenko, D.M., Hu, Y., Sa, S.M., Gong, Q., Abbas, A.R., Modrusan, Z., Ghilardi, N., de Sauvage, F.J., et al. (2008). Interleukin-22 mediates early host defense against attaching and effacing bacterial pathogens. Nat. Med. *14*, 282–289.

Zhernakova, A., Kurilshikov, A., Bonder, M.J., Tigchelaar, E.F., Schirmer, M., Vatanen, T., Mujagic, Z., Vila, A.V., Falony, G., Vieira-Silva, S., et al. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science *352*, 565–569.

FIGURES

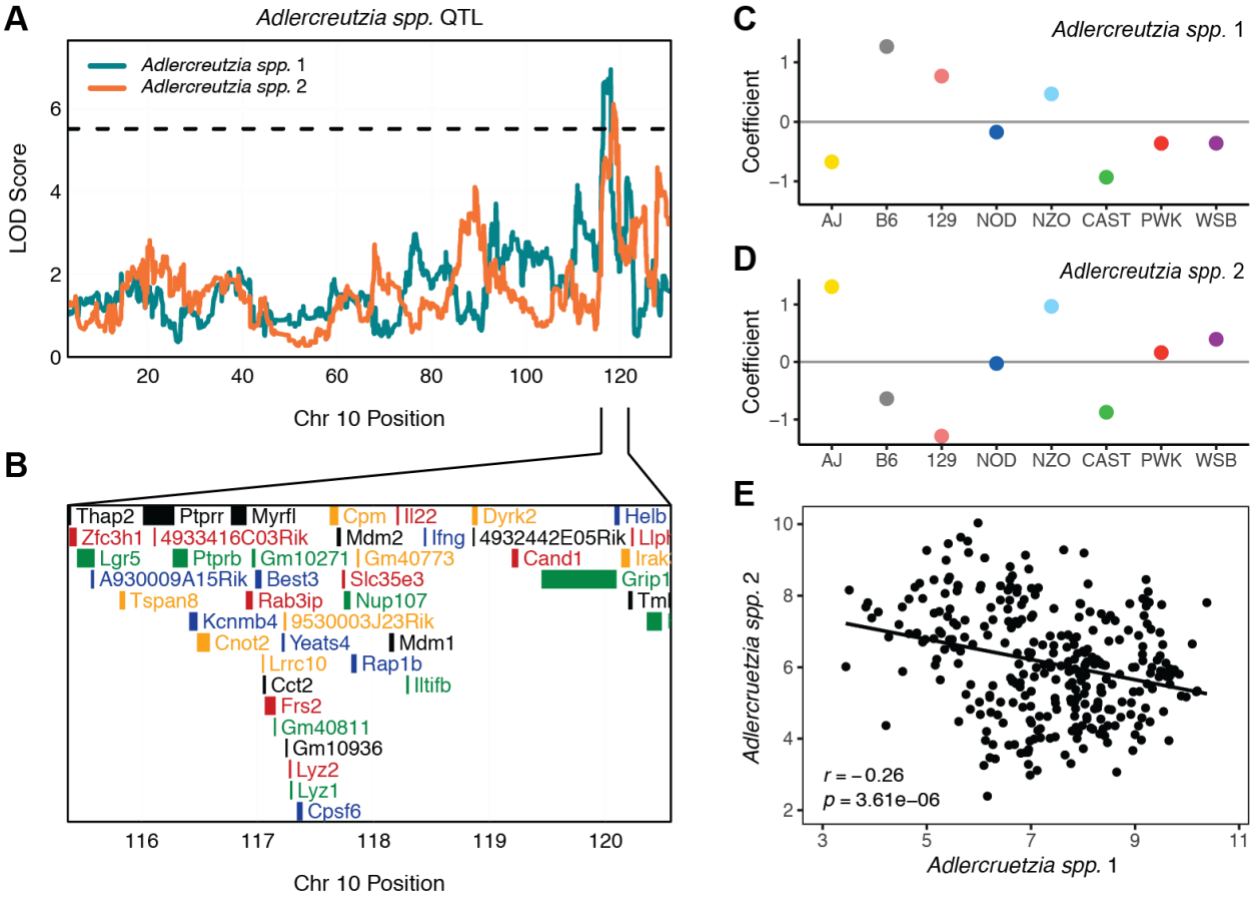


Figure C.1. Multiple Adlercreutzia sp. QTL map to immune genes on chromosome 10. (A) Association of two *Adlercreutzia* exact sequence variants (ESVs) mapping along chr 10. Dashed line denotes LOD threshold of 5.5. (B) Genes under the QTLs. Diversity Outbred (DO) founder coefficients at the QTL peak showing the effects of each founder allele on the abundance of (C) *Adlercreutzia sp.* 1 and (D) *Adlercreutzia sp*. 2. (E) Correlation of the normalized abundance of *Adlercreutzia sp.* 1 and *Adlercreutzia sp.* 2 in the DO mice (n = 309). Spearman correlation; p-value adjusted for multiple tests using Benjamini-Hochberg correction.

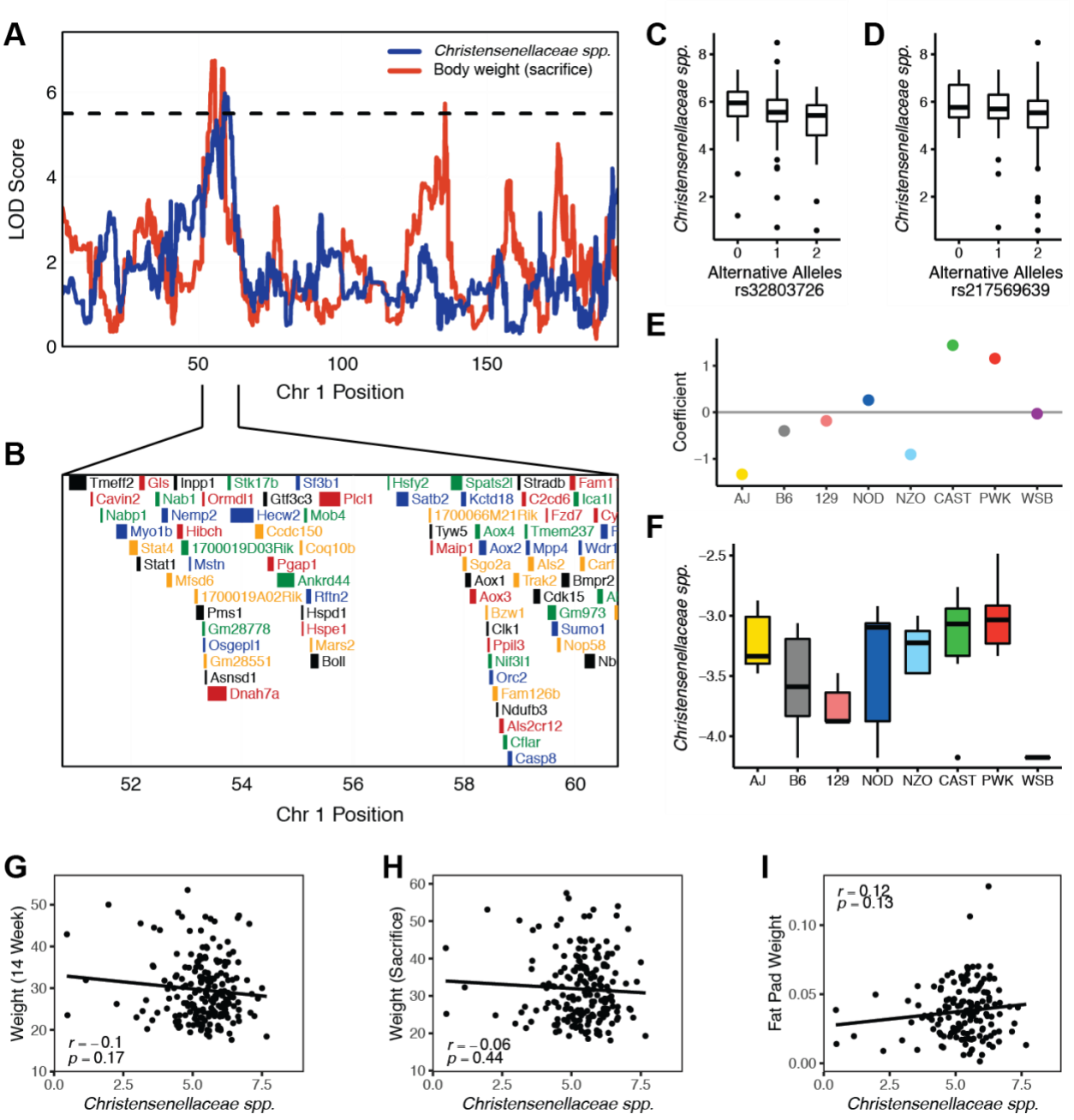


Figure C.2. *Christensenellaceae sp.* and body weight QTL map to chromosome 1. (A) Scan of chr 1 of *Christensenellaceae sp.* and body weight at sacrifice. Dashed line denotes LOD threshold of 5.5. (B) Genes under QTL. (C) Normalized *Christensenellaceae sp.* abundance within each genotype at the missense SNP (rs32803726) and (D) peak SNP (rs217569639). (E) Estimated founder allele effects and (F) observed abundance of Christensenellaceae family in the founder

strains. Pearson correlations in the DO mice between the abundance of *Christensenellaceae sp.* and (G) body weight at 14 weeks (n = 199), (H) body weight at sacrifice (n = 196), and (I) fat pad weight per gram body weight at sacrifice (n = 151). Data are presented as mean ± SEM; Kruskal Wallis one-way test followed by Wilcoxon pair-wise multiple comparisons with Benjamini-Hochberg correction; * $p < 0.05$, ** $p < 0.01$.

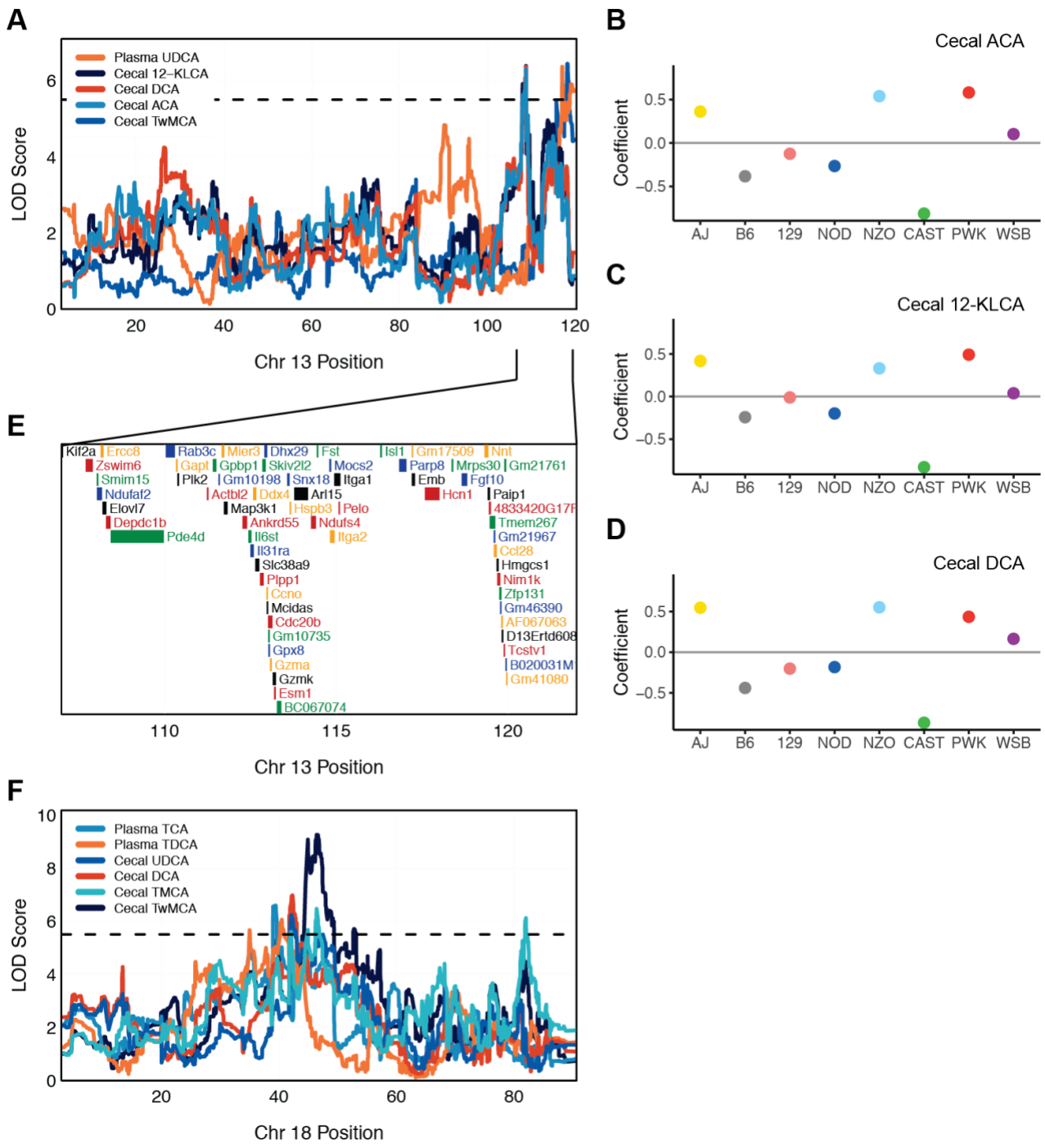


Figure C.3. Bile acids (BAs) associate to hotspots on chromosomes 13 and 18 of mouse genome. (A) Scan of chr 13 BA QTL hotspot where QTL for plasma ursodeoxycholic acid (UDCA), and cecal levels of 12-ketolithocholic acid (12-KLCA), deoxycholic acid (DCA), allocholic acid (ACA) and tauro-omega-muricholic (TωMCA) overlap. Dashed line denotes LOD threshold of

5.5. (B) Estimated founder allele effects for cecal ACA, (C) cecal 12-KLCA, and (D) cecal DCA levels. (E) Protein coding genes under chr 13 QTL hotspot. (F) Scan of chr 18 bile acid QTL hotspot including plasma taurocholic acid (TCA), plasma taurodeoxycholic acid (TDCA), cecal ursodeoxycholic acid (UDCA), cecal DCA, cecal tauro-muricholic acid (TMCA), and cecal TωMCA.