Systems Genetics Study of Gut Microbiome-Host Interactions in Mice

By

Qijun Zhang

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

2023

Date of final oral examination: 12/14/2023

The dissertation is approved by the following members of the Final Oral Committee:

Federico E. Rey, Associate Professor, Bacteriology

Karl W. Broman, Professor, Biostatistics and Medical Informatics

Bo Liu, Professor, Cell and Regenerative Biology

Garret Suen, Professor, Bacteriology

Zhengzheng Tang, Associate Professor, Biostatistics and Medical Informatics

ABSTRACT

The microbes inhabiting in the mammalian gut have profound effects on host biology and health. Alterations in the intestinal microbiome have been associated with a broad spectrum of human health conditions, including metabolic and cardiovascular diseases. Environmental factors, including diet and host genetic variation influence the gut microbiome and subsequently affect its interactions with the host. However, the molecular bases of how host genetic variation impacts the gut microbiome remain largely unknown. In this thesis, I investigated the effects of host genetic variation on gut microbiome composition and function in two genetically diverse mouse cohorts. I performed the shotgun metagenomic analyses to characterize gut microbiome. I applied systems genetics approaches and integrated gut microbiome with other omics datasets, including lipidome, transcriptome, and host genotypes, to evaluate how host genetics impact gut microbiome-host interactions.

In **Chapter 1**, I provided an overview of gut microbiome roles to human health and disease and discuss how host genetics associate with the gut microbiome in human and mouse studies.

In **Chapter 2**, I performed quantitative trait locus (QTL) mapping in the Diversity Outbred (DO) mouse cohort to identify genetic loci that are associated with variations in the gut microbiome, cecal lipidome, and intestinal transcriptome. I found overlapping QTL for the abundance of *Akkermansia muciniphila* and cecal levels of ornithine lipids (OL). I discovered that *A. muciniphila* is a major source of OL in the gut, provided evidence that OL have immunomodulatory effects and identified intestinal transcripts co-regulated with these traits including *Atf3*, which encodes for a transcription factor that plays vital roles in modulating metabolism and immunity.

In **Chapter 3**, I performed genome-wide association study (GWAS) in 90 inbred hyperlipidemic mouse strains from the Hybrid Mouse Diversity Panel (HMDP). I identified genomic loci that were associated with microbial enterotypes in the gut. I discovered genetic variants of *Amy1* gene that were associated with alterations in the abundance of taxa in the Firmicutes

(*Lachnospiraceae* family) and Bacteroidetes (*Muribaculaceae* family). Interestingly, these taxa encode distinct starch and sugar metabolism functions. Additionally, I applied Mendelian randomization to reveal that host physiology phenotypes, including liver fibrosis and plasma HDL-cholesterol levels, were causally associated with gut microbiome enterotypes.

Together, this thesis demonstrates the application of systems genetics to the study of gut microbiome-host interactions in mice and provides a foundation for the future mechanism studies and therapeutic opportunities.

ACKNOWLEDGEMENTS

Once upon a time, there was a little boy reading Stephen Hawking's *The Universe in a Nutshell* in his bedroom, pondering the beauty of science. Thanks to my ten-year-old self. Your curiosity and passion for science back then have paved the way for my accomplishments today. May that initial sense of wonder in exploring science accompany you as always.

I still remember the day six years ago when I embarked on a journey far away from home to pursue my graduate study here, brimming with excitement. It's hard to say that this incredible journey has now reached its final chapter. I'm thankful for the encounter with my Ph.D. advisor, Federico Rey, on the CMP orientation day. It was through you that I delved into the fascinating world of the microbiome. Thank you for your guidance and for nurturing me into an independent scientist. Your support in allowing me the freedom to choose research topics and delve into computational biology has been invaluable. I remember every time you come to the office and asked, 'Have you discovered anything exciting?' to me. While our achievements may not yet rival Nobel Prize levels, but I'm proud of what I learnt and uncovered during these years. Your passions for science everyday become the motivation to me as well. I'm grateful and fortunate to be your student.

Thank you, my Ph.D. committee members Karl Broman, Bo Liu, Garret Suen, and Zhengzheng Tang, for your expertise and insightful discussions throughout the years. I'd also like to thank my past and present lab members, as well as my collaborators. I'm grateful to be part of this remarkable scientific community. Teamwork is the cornerstone of achievement of science. My special thanks to Lindsay Traeger. You are a mentor during my rotation and early years of my Ph.D. I had limited knowledge of bioinformatics when I joined the lab, and you were like a guiding beacon, illuminating the path allowing me to pursue my interests in the computation biology field.

Graduate studies are challenging, and life would be boring without friends. Thanks to my friends who supported me along this journey: Zhenxuan Chen, Bo Dong, Shuai Shao, Linquan

Ma, Jiatong Li, Yuhui Li, and Dongxue Du. That happiness hanging out together and shared moments made me not alone after tough days.

Last but certainly not least, I want to express my deep gratitude to my partner, Jieru Zhu, and my parents, for their unwavering support and love throughout this entire journey.

To my parents and my partner

TABLE OF CONTENTS

CHAPTER 1: Introduction

The mammalian intestinal tract harbors trillions of diverse microorganisms including bacteria, archaea, fungi, and viruses, which together constitute the gut microbiota and their aggregate genomes constitute as gut microbiome (Lozupone et al., 2012). These symbiotic microorganisms, forming a complex ecology system, are indispensable to host physiology. Gut microbes have the ability to breakdown indigestible dietary compounds, harvest essential nutrients and provide genetic and metabolic attributes that are absent in the mammalian genome (Bäckhed et al., 2005). Given the diverse functional repertoire, the gut microbiota plays a critical role in host immune maturation (Wu & Wu, 2012), harvesting energy (Canfora et al., 2015), protecting host from pathogens (Kamada et al., 2013), regulating neurologic signaling (Yano et al., 2015) and maintaining gut homeostasis (Lee et al., 2022). It has been demonstrated that changes in gut microbiota composition and microbial functions are associated with a broad spectrum of human health including chronic disorders such as metabolic (Turnbaugh et al., 2006), cardiovascular (Z. Wang et al., 2011) and neurodegenerative diseases (Vogt et al., 2017). Therefore, the gut microbiota represents a target with potential to improve human health. Understanding the mechanisms underlying microbiome-host interactions may facilitate therapeutic developments.

The gut microbiome carries at least 10 times more unique genes than the host genome (Ley, Peterson, et al., 2006). Advances in sequencing technologies and development of computational methods provide the opportunity to characterize and profile the microbial communities with unprecedented resolution. Gene amplicon sequencing using the amplification and sequencing of the variable regions of the 16S ribosomal RNA (rRNA) genes can profile the taxonomic features of bacteria and archaea in a sample (Eckburg et al., 2005), while in the shotgun metagenomic approach, all genomic content of a sample is sequenced, enabling analysis of the entire microbiome including microbial functions and metabolic pathways (Turnbaugh et al., 2007). These techniques facilitate the observation of variations in gut microbiome that are driven by a variety of variables. Because of the contributions of the gut microbiome to various diseases,

it has become increasingly important to identify and understand how specific factors influence the gut microbiome. In this introduction chapter, I first summarize the current understanding of how the gut microbiome modulate health and disease, and discuss the importance of bacterial-derived metabolites for gut microbiome-host interactions, using lipids and bile acids as examples. Then I review how host genetic variation is associated with gut microbiome in human and mouse studies. Additionally, I discuss how systems genetics can help to generate new hypotheses relating different omics datasets in genetically diverse populations and to investigate mechanisms of gut microbiome-host interactions.

1.1 Gut Microbiome in Human Health and Disease

Contributions to host physiology

Gut microbes colonize in mucosal tissues early in life and have profound influence on the host immune system. Exposure to microbes begins in utero and expands rapidly after birth (Rackaityte et al., 2020). After birth, the establishment of the infant gut microbiome is derived from different sources including maternal milk and the environment, such as other family members and pets (Enav et al., 2022). Early life colonization influences host immune system education and the stability of the microbiota during early life impacts the host resistance (or susceptibility) to disease in later life. The germ-free (GF) animals allowed for experiment development to study the microbiota influences on the host physiology. Conventional mice, compared with GF mice, showed activation and de novo generation of colonic regulatory T cells, a specialized population of $CD4^+$ T cells that restrict immune activation, and maintained intestinal homeostasis (Geuking et al., 2011). Therefore, the immune maturation is influenced by the presence of commensal microbes.

An important function of the gut microbiome is to aid the host in nutrient digestion. Dietary compounds such as complex carbohydrates and plant polysaccharides can’t be digested by human enzymes. Instead, the gut microbiota can ferment these non-digestible carbohydrates

such as inulin and resistant starch to yield energy source for microbial growth in the gut. On the other hand, the end products of these no-digestible carbohydrates such as short-chain fatty acids (SCFAs), mainly acetate, propionate and butyrate, have profound effects on host health (Canfora et al., 2015). Butyrate is a major energy source for colonic epithelium, regulator of glucose homeostasis and modulator of inflammation.

Metabolic disorders such as obesity, insulin resistance and type 2 diabetes are associated with systemic inflammation (Osborn & Olefsky, 2012). The gut microbiota contains high levels of lipopolysaccharide (derived from gram negative bacteria) and peptidoglycan (higher levels in gram positives, but present in all bacteria), which can cause inflammation. Colonization of germ-free mice with *Escherichia coli* promoted macrophage infiltration of adipose tissue and leaded to expression of pro-inflammatory cytokines (Caesar et al., 2012). Patients with type 2 diabetes have increased level of plasma lipopolysaccharide, and feeding lipopolysaccharide to mice increased adipose tissue inflammation and reduce insulin sensitivity (Cani et al., 2007). Lipopolysaccharide molecules bind to Toll-like receptor 4 (TLR4), and peptidoglycan fragments are recognized by nucleotide-binding oligomerization domain (NOD) receptors, both of which activates proinflammatory signaling cascades. It is well accepted in the field that the gut microbiota can affect host metabolism by altering tissue inflammation.

Factors influencing the gut microbiome

Diet modulates the composition of gut microbiota in humans and mice. Long-term dietary habits have a considerable effect on the human gut microbiota. Diet supplemented with resistant starch elicit an increase in fecal levels of *Ruminococcus bromii* and *Eubacterium rectale* which are both associated with higher fiber fermentation in humans (Walker et al., 2011). Mice fed on high-fat diets had reduced members of Bacteroidetes and increased members of Firmicutes and Proteobacteria (Turnbaugh et al., 2008). In another study, mice fed on high-fat diet containing lard had lower abundance of the beneficial bacteria *Akkermansia muciniphila*, *Lactobacillus* and

Bifidobacterium, compared with mice fed on high-fat diet containing fish oil (Caesar et al., 2015). The lard diet also reduced insulin sensitivity and increased inflammation in white adipose tissue (WAT) through activation of Toll-like receptor 4 (TLR4) signaling.

The effects of antibiotics on gut microbiota are relative larger than other factors. The gut microbiota are not resilient to repeated antibiotic administration in human (Dethlefsen & Relman, 2011). The antibiotics taken early in life have a profound effect on the gut microbiome that can result in later development of obesity, asthma, inflammatory bowel disease and other disorders (Trasande et al., 2013).

Lifestyle is another important environmental factor that influences the gut microbiota. Exercise produces gut microbial community changes by reducing inflammation, which are correlated with changes in cytokine profile (Clarke et al., 2014). Sleep deprivation is associated with gut microbiota changes, a study showed that the sleep loss is associated with higher level of Firmicutes and lower of Bacteroidetes (Benedict et al., 2016). Stress increases intestinal permeability and is associated with increased level of Firmicutes and decreased level of Bacteroidetes, which are also corresponding to the shifts of inflammation markers (Karl et al., 2017).

Dynamics and dysbiosis of gut microbiome

The human microbiome demonstrates extremely robustness and plasticity over the long timescale and in response to many types of perturbations such as short-term dietary change and traveling (host geography) (David et al., 2014). The composition of gut microbiota is also influenced by circadian rhythm. Disturbance of gut microbiome composition and function can lead to the disruption of host circadian rhythms, which can specifically alter hormone regulation in mice (Leone et al., 2015). Association studies showed the gut dysbiosis is linked with a broad range of diseases in humans and mice. For example, in both humans and mice, the gut microbiome changes are associated with obesity. Genetically obese *ob/ob* mice had more Firmicutes and

fewer Bacteroidetes than the lean wide-type littermates mice (Ley et al., 2005). This phenotype was transferred to germ-free mice that were colonized with the microbiota from the obese donors mice (Turnbaugh et al., 2006). Similar changes have been observed in humans, i.e., Bacteroidetes levels increase when weight is reduced, suggesting Bacteroidetes may be responsive to calorie intake (Ley, Turnbaugh, et al., 2006). Type 2 diabetes (T2D) patients showed a decline in butyrate-producing bacteria, which in preclinical models show metabolic benefits, and increased levels of different opportunistic pathogens (J. Qin et al., 2012). Studies also suggested that T2D patients were likely to have a functional dysbiosis in the gut instead of having specific microbial species that have a direct association with T2D pathophysiology.

Gut microbiome also plays an important role in cardiovascular diseases, including atherosclerosis. Patients with symptomatic atherosclerosis (myocardial infarction or cerebrovascular events) had increased level of genus *Collinsella* and decreased level of genus *Eubacterium* and *Roseburia* compared with health control people (Karlsson et al., 2012). Patients with coronary artery disease had increased number of Lactobacillales and the ratio of Firmicutes to Bacteroidetes compared with control (Kazemian et al., 2020). Gut microbial derived lipopolysaccharides (LPS) can trigger host immune system response via Toll-like receptors (TLRs), incuding a low-grade inflammatory state and aggravate the progression of atherosclerosis. The gut microbiota also regulates host metabolism such as cholesterol and plasma lipids levels to induce atherosclerosis(Velagapudi et al., 2010).

Gut dysbiosis influences liver health and disease. Patients of nonalcoholic fatty liver disease (NAFLD) with advanced fibrosis had shifted microbiota such as lower fecal *Ruminococcus obeum* and *Eubacterium rectale* (Loomba et al., 2017). NAFLD can progress to non-alcoholic steatohepatitis (NASH) in 20–30% of cases with its sequelae of liver scarring, cirrhosis and liver cancer. Patients with NASH show decreased intestinal viral diversity and a reduction of phages (Lang et al., 2020). Gut dysbiosis induces intestinal inflammation, which contributes to intestinal barrier dysfunction and translocation of microorganism associated

molecular patterns (MAMPs) to the liver and cause liver damage. The dysbiosis is also associated with functional metabolic consequences such as microbiome-derived ethanol, whose concentration was higher of individuals with NAFLD or NASH (Meijnikman et al., 2022).

Therapeutic opportunities

Large-scale and longitudinal clinical studies have revealed associations between gut microbiome and the various disease states discussed above. In some cases, these associations have been replicated in animal models, leading to a mechanistic understanding of how gut microbes influence disease (Fischbach, 2018). Introduction of a group selected bacteria containing selected functions (e.g., probiotics), associated with healthy status, can provide health benefits to the host. Direct supplementation with beneficial proteins or metabolites derived from gut microbes can also improve host health (Sorbara & Pamer, 2022). Changes in dietary alteres the gut microbiome in humans. A better understanding of the nutrient requirements and metabolism of beneficial microbes could inform food choices or prebiotics that support these species through targeted supplementation of preferred growth substrates, and further inducing the beneficial effects to the host.

Gut microbiome has also been explored in oncology. Selected gut bacteria can modulate the tumor microenvironment and anticancer therapies (Iida et al., 2013) and certain microbes enhance the efficacy of cancer immunotherapy (Sivan et al., 2015). Engineering of exogenous bacterial and viral agents can be used for cancer therapy, particularly as powerful immunotherapy options or neoadjuvants (Sepich-Poore et al., 2021).

1.2 Metabolites in Microbiome-Host Interactions

The gut microbiome generates a massive number of small molecules through *de novo* biosynthesis or modification of host and dietary substrates. Because the complex diets digested by the host and gut microbes creating the intricate molecular trafficking, the metabolites in the

body are highly dynamic. These metabolites include dietary compounds, gut microbial metabolites produced from dietary and host substrates, host-derived metabolites, drug compounds and microbiome-modified drugs (Koppel et al., 2017). These broad categories of metabolites regulate host metabolism, play important roles for gut microbiome-host communications, and thus it is informative to study the underlying mechanisms that regulate their abundance.

Bile acids

Bile acids (BA) are potent “digestive surfactants” that promote absorption of lipids in the small intestine as well as regulate cholesterol homeostasis. The synthesis of BA occurs exclusively in the liver in a series of enzymatic reactions in the hepatocyte that convert cholesterol into more water-soluble amphipathic compounds, called primary bile acids; these include cholic acid (CA) and chenodeoxycholic acid (CDCA). Primary BA are then conjugated to taurine or glycine. Upon consumption of a meal, BA are secreted into the small intestine, where they aid with fat absorption and taken up (reabsorbed) in the distal ileum and stored in the gallbladder. BA that are not reabsorbed by the host are metabolized by bacteria from small intestine. Gut bacteria deconjugate these bile acids by removal of the glycine or taurine conjugate and are further metabolized in the large intestine into secondary bile acids, including lithocholic acid (LCA) from CDCA and deoxycholic acid (DCA) from CA (Wahlström et al., 2016). Bile acid deconjugation is carried out by bacterial bile salt hydrolase (BSH), which are wildly present in gut microbial community (Song et al., 2019). Deconjugated primary bile acids conversion into secondary bile acids is carried out by bacterial 7-dehydroxylation, a series of reactions by bacterial bile acid-inducible (*bai*) genes (Funabashi et al., 2020). Germ-free mice had less diverse bile acids profiles compared with conventionally raised mice (Swann et al., 2011). Bile acids function as signaling molecules that can act on cellular receptors such as G protein-coupled receptor TGR5, which promotes glucose homeostasis (C. Thomas et al., 2009). Therefore, host metabolism can be

affected through microbial modifications of bile acids, which lead to altered bile acid receptors signaling.

Lipids

Microbial lipids encompass a variety of structures and functions. Lipids are not only important for their structural roles such as in cell membranes, also for their signaling activities to host (Zhang & Rock, 2008). Microbial-derived lipids can be sensed by the host to modulate innate and adaptive immune pathways and to regulate metabolic pathways that in turn influence the progression of chronic inflammation, metabolic syndrome and cardiovascular disease (Yoon et al., 2021). The major lipid classes found in bacterial membranes include phospholipids, such as phosphoethanolamine (PE), phosphoserine (PS), phosphocholine (PC), phosphoinositol (PI) and phospho- glycerol (PG); glycerolipids, such as diacylglycerol (DAG) and triacylglycerol (TAG); cardiolipins (CL); saccharolipids, such as lipopolysaccharides (LPS); and sphingolipids (Heaver et al., 2018). Each lipid class has unique architectures, structural features, and functions.

Many bacterial lipids can be sensed by host pattern recognition receptors, such as Toll-like receptors (TLRs), NOD-like receptors (NLRs) and G protein-coupled receptors (GPCRs) (Miller et al., 2005). For example, a diacyl phosphatidylethanolamine with two branched chains (a15:0-i15:0 PE) produced by gut bacteria *Akkermansia muciniphila* activates host pattern recognition receptor (PRR) heterodimer consisting of TLR2-TLR1 and induces host inflammatory cytokines (Bae et al., 2022). The best characterized saccharolipids are lipopolysaccharides (LPS) from Gram-negative bacteria (Raetz & Whitfield, 2002). Bacterial LPS are recognized by the Toll-like receptor 4 (TLR4) on host cells to initiate pro-inflammatory responses (Park et al., 2009). The alerted structures of Lipid A from LPS induce host immune responses differently, for example, *Bacteroides* LPS is structurally distinct from *E. coli* LPS and inhibits innate immune signaling and endotoxin tolerance (Vatanen et al., 2016). Bacterial sphingolipids also impact host inflammatory and metabolic pathways. A recent study showed that the colonization of germ-free mice with

sphingolipid-deficient bacteria resulted in gut inflammation and changes in host ceramide pools (Brown et al., 2019). The sphingolipids from the outer membrane of *Bacteroides* act as agonists for TLR2 signaling in macrophages and are important in limiting inflammatory signaling (Rocha et al., 2021). Altogether these results suggest that bacterial lipids are key molecules mediating microbe-host interactions.

1.3 Effects of Host Genetics on Gut Microbiome

The human microbiome is unique in each individual and the interpersonal differences are larger than the differences within a person over time ("The Integrative Human Microbiome Project," 2014). Identical twins had more similar gut microbial composition and structure than the nonidentical twins (Goodrich et al., 2014). These suggest the genetic variation influences gut microbial composition. Although previous evidences suggests that the environment including diets dominate over host genetics in shaping the gut microbiome in humans and mice (Carmody et al., 2015; Rothschild et al., 2018). With the increasing statistic powers, recent population genetics studies have identified enormous genomic loci associated with gut microbiome, which provide the bases to study molecular mechanisms of gut microbiome-host interactions.

Host genetic associations in humans

Recent advances in genotyping methods and sequencing technologies have enabled large-scale genome-wide association studies (GWAS) for population genetics in humans. Studies analyzing the effect of host genetics on the human microbiome in larger cohort (> 1,000 individuals) showed that a proportion of gut microorganisms were substantial heritable, ~10% of taxa having heritability (h^2) greater than 0.20, in TwinsUK cohort (Goodrich et al., 2016), Canadian (GEM Project Research Consortium et al., 2016) and Dutch cohort (Lopera-Maya et al., 2020). These taxa are in the range of the heritability of many human common complex traits such as

fasting glucose levels (h^2 = 0.31), insulin levels (h^2 = 0.25) and blood pressure (h^2 = 0.15) (DIAGRAM Consortium et al., 2010).

Among the identified genomic loci and heritable taxa, some have been consistently confirmed across multiple studies. The abundance of *Bifidobacterium* genus was associated with the functional genetic variant near the lactase gene (*LCT*). This association was reported at genome wide significance in UK (Goodrich et al., 2016), Canadian (GEM Project Research Consortium et al., 2016), Dutch (Lopera-Maya et al., 2020) and Finnish (Y. Qin et al., 2020) cohorts and a lower significance level in German cohort (J. Wang et al., 2016). The *Bifidobacterium* are the most heritable taxa reported in the UK and Dutch cohorts. The strong genomic associations for the *Bifidobacterium* were the functional variant rs4988235 and its proxies near *LCT* gene. The G allele of rs4988235 corresponds to the phenotype of decreased ability to metabolize lactose, that is lactase non-persistence. The selective pressure due to consumption of milk after weaning increases of lactase-persistence alleles frequency (Suzuki & Ley, 2020). The lactase non-persistence genotype is associated with higher gut abundance of *Bifidobacterium*, which can degrade lactose, and this association is dependent on milk consumption (Bonder et al., 2016).

Another consistent associations are variants at gene *ABO* locus. In the German cohort, two independent SNPs in the *ABO* locus were associated with the abundance of *Faecalibacterium* (rs3758348) and *Bacteroides* (rs8176632) (Rühlemann et al., 2021). In the Finnish cohort, variants in partial linkage disequilibrium (LD) near *ABO* were associated with the abundance of *Faecalicatena lactaris* and *Collinsella* (Y. Qin et al., 2020). In the Dutch cohort, the same and other variants in LD are associated with *Bifidobacterium* abundance, *Collinsella* abundance and the lactose-degradation pathway (Lopera-Maya et al., 2020). In these three studies, the variant rs601338 from *FUT2* gene was associated with *ABO* antigens expression. The individuals with G allele at rs601338 do not express or expose the A or B antigens of *ABO*. The associations between gut microbiome and variants at ABO locus were dependent on *ABO* antigens secretor

status (A/B/AB-oligosaccharides on H-antigen), blood types (O group or A/B/AB groups) and fiber intake (Y. Qin et al., 2020).

Host genetic associations in mice

Mouse models are instrumental to study disease biology, as they enable precise control of the environment and genetic composition, allowing more rigorous observations and reproducible measurements. The external environment of lab mice can be well controlled and monitored, which also facilitates the study of gene by environment interactions. Mouse genetic populations are equivalent to the genetic diversity of human populations and are adequate for genetic association analyses (Li et al., 2018). Many genetic determinants of complex traits have been identified using mouse populations and verified in human cohorts. In human studies, it is challenging to assess critical environmental factors influencing disease development, thus limiting the ability to study the underlying genetic determinants of complex traits and diseases, as well as the gene by environment interactions (D. Thomas, 2010). Mouse models can be used to examine the influence of genome in the response to nutrients or drugs. The most extensively used mouse strain in biomedical research is C57BL/6J, which however only carries the minor alleles for 19% of the high-impact variants among other inbred mouse strains (Keane et al., 2011). Thus, use of other genetically diverse mouse cohorts is needed for large-scale genomic association studies. There are different resources for this purpose. This includes the BXD cohort, which was derived from the C57BL/6J (B6 or B) and DBA/2J (D2 or D) strains that have different response to drugs and diet-induced obesity (Ashbrook et al., 2021) and the collaborative cross (CC) and diversity outbred (DO) cohorts, which are derived from eight parental strains including three wild derived strains (Saul et al., 2019). The CC/DO founder strains capture around 90% (versus ~13% in BXD) of the common genetic variations in *Mus musculus* strains, approximating human genetic diversity (The Complex Trait Consortium, 2004). Another mouse cohort frequently used by scientists comprises more than 100 inbred strains of mice, which is hybrid mouse diversity panel (HMDP)

(Bennett et al., 2010). Many associations between mouse genomic loci and gut microbiome were identified from these resources.

Studies identified quantitative trait loci (QTLs) for genus level of *Bacteroides*, *Oscillibacter* and *Bifidobacterium* and family levels of Rikenellaceae and Prevotellaceae in BXD cohorts, where the candidate genes involved in Toll-like receptor pathways and cytokine genes (McKnite et al., 2012; Perez-Munoz et al., 2019). A study used HMDP cohort identified associations for species of *Akkermansia muciniphila*, *Ruminococcus gnavus*, *Roseburia spp.*, *Oscillospira spp.*, and *Turicibacter spp.* (Org et al., 2015). Furthermore, these taxa were associated with metabolic and cardiovascular traits, and *A. muciniphila* improved obesity and metabolic parameters in mice fed a high-fat/high-sucrose diet. A study used CC cohort identified hundreds of loci and genes enriched in gastrointestinal cancer, inflammatory responses and lipid metabolism that controlling abundance taxa such as *Lactobacillus* (Snijders et al., 2017). A study used DO cohort identified genetic determinants of gut microbiota composition and bile acid profiles and an overlapping QTL for *Turicibacter sp.* and plasma cholic acid, which mapped to a locus containing the gene for the ileal bile acid transporter, *Slc10a2* (Kemis et al., 2019). All these studies used gut microbiota composition (16S rRNA sequencing) as traits. A major limitation of these studies is lack of microbial function and pathway. Thus, identifying host genomic associations with gut microbial functions and metabolic pathways is necessary for a better understanding of gut microbiome-host interactions.

1.4 Systems Genetics

Mendel's experiments over a century ago explained how specific genetic alleles influence single trait or phenotype, such as color or size of peas. Recently developed technologies can comprehensively dissect the genetic architecture of complex traits with a better understanding of links from genotypes to phenotypes. This new area has been termed as systems genetics (Civelek & Lusis, 2014). From single phenotype to phenomics, the quantitative and detailed

measurements on a broad spectrum of phenotypes, systems genetics aims to integrate genome-wide data across many different levels. In a broader sense, systems genetics uses a genome-wide analysis with many quantitative phenotypes at molecular or organismal level, in different conditions or environments to understand the flow of biological information that underlies complex traits, including human diseases. It also combines a range of both experimental and statistical methods to integrate intermediate phenotypes, such as transcripts, proteins, and metabolites, in genetically diverse populations including humans or model organisms. Systems genetics studies provide global view of the molecular architecture of complex traits and are useful to identify the causal genes, pathways and networks underlying common human diseases.

A linkage or an association is the link between a phenotype with a genotype in a population of individuals. Usually, associations refer to the natural populations of unrelated individuals and linkages refer to the families or groups of families from experimental populations. Such associations and linkages can be identified by the mapping of quantitative trait loci (QTL). A QTL is a region of genome that modulate a phenotype of interest, which include gene expression QTL (eQTL), protein QTL (pQTL), metabolite QTL (mQTL), microbiome QTL (mbQTL) and QTL for other molecular phenotypes. When a genomic region is associated with traits, the causal genetic variants that are responsible for these complex traits can be determined by evaluation of their likely function based on publicly available genomic annotation and prior knowledge. Usually, the cost-efficient genotyping microarrays containing a large number of SNPs which may not be the causal SNPs. In this case the causal SNPs are those highly correlated in the neighborhood region with a large amount of linkage disequilibrium (LD). In human GWAS, fine-mapping strategies seek to determine the causal variants using statistical methods (Schaid et al., 2018). These efforts generate the hypotheses between the natural genetic variants with complex traits. Systems genetics extend from these hypotheses to explore how the information flow from DNA to phenotypes with complementary experiments including transgenic mice model and small interfering RNA (siRNA) and other complementary data sets.

In a systems genetics study, genetic mapping between different molecular layers can provide evidence of a relationship by co-mapping or colocalization. This is based on the idea that if two traits map to the same genomic region, one trait could be the cause to the other one. For example, a study in DO mice performed genetic mapping of gut microbiome and bile acids profiles and found the overlapping QTL for *Turicibacter sp.* and plasma cholic acid, which mapped to a locus containing the gene for the ileal bile acid transporter *Slc10a2* (Kemis et al., 2019). Beyond the co-mapping strategy, mediation analysis seeks to determine whether a QTL has separate effects on two traits, or if it affects one trait through its effect on another trait, in which case the intermediate trait is called a mediator. For example, a study investigated the consequences of natural genetic diversity on the proteome and identified the second protein or transcript as the causal mediator of distant pQTL in DO mice (Chick et al., 2016). Given the dynamic nature of gut microbiome, leveraging the systems genetics would deconvolute the complexity between gut microbiome with other traits and facilitate to understand the mechanisms underlaying the gut microbiome-host interactions.

1.5 References

Ashbrook, D. G., Arends, D., Prins, P., Mulligan, M. K., Roy, S., Williams, E. G., Lutz, C. M., Valenzuela, A., Bohl, C. J., Ingels, J. F., McCarty, M. S., Centeno, A. G., Hager, R., Auwerx, J., Lu, L., & Williams, R. W. (2021). A platform for experimental precision medicine: The extended BXD mouse family. *Cell Systems*, *12*(3), 235-247.e9. https://doi.org/10.1016/j.cels.2020.12.002

Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., & Gordon, J. I. (2005). Host-Bacterial Mutualism in the Human Intestine. *Science*, *307*(5717), 1915–1920. https://doi.org/10.1126/science.1104816

Bae, M., Cassilly, C. D., Liu, X., Park, S.-M., Tusi, B. K., Chen, X., Kwon, J., Filipčík, P., Bolze, A. S., Liu, Z., Vlamakis, H., Graham, D. B., Buhrlage, S. J., Xavier, R. J., & Clardy, J. (2022). Akkermansia muciniphila phospholipid induces homeostatic immune responses. *Nature*. https://doi.org/10.1038/s41586-022-04985-7

Benedict, C., Vogel, H., Jonas, W., Woting, A., Blaut, M., Schürmann, A., & Cedernaes, J. (2016). Gut microbiota and glucometabolic alterations in response to recurrent partial sleep deprivation in normal-weight young individuals. *Molecular Metabolism*, *5*(12), 1175–1186. https://doi.org/10.1016/j.molmet.2016.10.003

Bennett, B. J., Farber, C. R., Orozco, L., Min Kang, H., Ghazalpour, A., Siemers, N., Neubauer, M., Neuhaus, I., Yordanova, R., Guan, B., Truong, A., Yang, W. -p., He, A., Kayne, P., Gargalovic, P., Kirchgessner, T., Pan, C., Castellani, L. W., Kostem, E., … Lusis, A. J. (2010). A high-resolution association mapping panel for the dissection of complex traits in mice. *Genome Research*, *20*(2), 281–290. https://doi.org/10.1101/gr.099234.109

Bonder, M. J., Kurilshikov, A., Tigchelaar, E. F., Mujagic, Z., Imhann, F., Vila, A. V., Deelen, P., Vatanen, T., Schirmer, M., Smeekens, S. P., Zhernakova, D. V., Jankipersadsing, S. A., Jaeger, M., Oosting, M., Cenit, M. C., Masclee, A. A. M., Swertz, M. A., Li, Y., Kumar,

V., … Zhernakova, A. (2016). The effect of host genetics on the gut microbiome. *Nature Genetics*, *48*(11), 1407–1412. https://doi.org/10.1038/ng.3663

Brown, E. M., Ke, X., Hitchcock, D., Jeanfavre, S., Avila-Pacheco, J., Nakata, T., Arthur, T. D., Fornelos, N., Heim, C., Franzosa, E. A., Watson, N., Huttenhower, C., Haiser, H. J., Dillow, G., Graham, D. B., Finlay, B. B., Kostic, A. D., Porter, J. A., Vlamakis, H., … Xavier, R. J. (2019). Bacteroides-Derived Sphingolipids Are Critical for Maintaining Intestinal Homeostasis and Symbiosis. *Cell Host & Microbe*, *25*(5), 668-680.e7. https://doi.org/10.1016/j.chom.2019.04.002

Caesar, R., Reigstad, C. S., Bäckhed, H. K., Reinhardt, C., Ketonen, M., Östergren Lundén, G., Cani, P. D., & Bäckhed, F. (2012). Gut-derived lipopolysaccharide augments adipose macrophage accumulation but is not essential for impaired glucose or insulin tolerance in mice. *Gut*, *61*(12), 1701–1707. https://doi.org/10.1136/gutjnl-2011-301689

Caesar, R., Tremaroli, V., Kovatcheva-Datchary, P., Cani, P. D., & Bäckhed, F. (2015). Crosstalk between Gut Microbiota and Dietary Lipids Aggravates WAT Inflammation through TLR Signaling. *Cell Metabolism*, *22*(4), 658–668. https://doi.org/10.1016/j.cmet.2015.07.026

Canfora, E. E., Jocken, J. W., & Blaak, E. E. (2015). Short-chain fatty acids in control of body weight and insulin sensitivity. *Nature Reviews Endocrinology*, *11*(10), 577–591. https://doi.org/10.1038/nrendo.2015.128

Cani, P. D., Amar, J., Iglesias, M. A., Poggi, M., Knauf, C., Bastelica, D., Neyrinck, A. M., Fava, F., Tuohy, K. M., Chabo, C., Waget, A., Delmée, E., Cousin, B., Sulpice, T., Chamontin, B., Ferrières, J., Tanti, J.-F., Gibson, G. R., Casteilla, L., … Burcelin, R. (2007). Metabolic Endotoxemia Initiates Obesity and Insulin Resistance. *Diabetes*, *56*(7), 1761–1772. https://doi.org/10.2337/db06-1491

Carmody, R. N., Gerber, G. K., Luevano, J. M., Gatti, D. M., Somes, L., Svenson, K. L., & Turnbaugh, P. J. (2015). Diet Dominates Host Genotype in Shaping the Murine Gut

Microbiota. *Cell Host & Microbe*, *17*(1), 72–84. https://doi.org/10.1016/j.chom.2014.11.010

Chick, J. M., Munger, S. C., Simecek, P., Huttlin, E. L., Choi, K., Gatti, D. M., Raghupathy, N., Svenson, K. L., Churchill, G. A., & Gygi, S. P. (2016). Defining the consequences of genetic variation on a proteome-wide scale. *Nature*, *534*(7608), 500–505. https://doi.org/10.1038/nature18270

Civelek, M., & Lusis, A. J. (2014). Systems genetics approaches to understand complex traits. *Nature Reviews Genetics*, *15*(1), 34–48. https://doi.org/10.1038/nrg3575

Clarke, S. F., Murphy, E. F., O'Sullivan, O., Lucey, A. J., Humphreys, M., Hogan, A., Hayes, P., O'Reilly, M., Jeffery, I. B., Wood-Martin, R., Kerins, D. M., Quigley, E., Ross, R. P., O'Toole, P. W., Molloy, M. G., Falvey, E., Shanahan, F., & Cotter, P. D. (2014). Exercise and associated dietary extremes impact on gut microbial diversity. *Gut*, *63*(12), 1913–1920. https://doi.org/10.1136/gutjnl-2013-306541

David, L. A., Materna, A. C., Friedman, J., Campos-Baptista, M. I., Blackburn, M. C., Perrotta, A., Erdman, S. E., & Alm, E. J. (2014). Host lifestyle affects human microbiota on daily timescales. *Genome Biology*, *15*(7), R89. https://doi.org/10.1186/gb-2014-15-7-r89

Dethlefsen, L., & Relman, D. A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proceedings of the National Academy of Sciences*, *108*(supplement_1), 4554–4561. https://doi.org/10.1073/pnas.1000087107

DIAGRAM Consortium, GIANT Consortium, Global BPgen Consortium, Anders Hamsten on behalf of Procardis Consortium, the MAGIC investigators, Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A. U., Wheeler, E., Glazer, N. L., Bouatia-Naji, N., Gloyn, A. L., Lindgren, C. M., Mägi, R., Morris, A. P., Randall, J., … Wilson, J. F. (2010). New genetic loci implicated in fasting glucose homeostasis and

their impact on type 2 diabetes risk. *Nature Genetics*, *42*(2), 105–116. https://doi.org/10.1038/ng.520

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., & Relman, D. A. (2005). Diversity of the Human Intestinal Microbial Flora. *Science*, *308*(5728), 1635–1638. https://doi.org/10.1126/science.1110591

Enav, H., Bäckhed, F., & Ley, R. E. (2022). The developing infant gut microbiome: A strain-level view. *Cell Host & Microbe*, *30*(5), 627–638. https://doi.org/10.1016/j.chom.2022.04.009

Fischbach, M. A. (2018). Microbiome: Focus on Causation and Mechanism. *Cell*, *174*(4), 785–790. https://doi.org/10.1016/j.cell.2018.07.038

Funabashi, M., Grove, T. L., Wang, M., Varma, Y., McFadden, M. E., Brown, L. C., Guo, C., Higginbottom, S., Almo, S. C., & Fischbach, M. A. (2020). A metabolic pathway for bile acid dehydroxylation by the gut microbiome. *Nature*, *582*(7813), 566–570. https://doi.org/10.1038/s41586-020-2396-4

GEM Project Research Consortium, Turpin, W., Espin-Garcia, O., Xu, W., Silverberg, M. S., Kevans, D., Smith, M. I., Guttman, D. S., Griffiths, A., Panaccione, R., Otley, A., Xu, L., Shestopaloff, K., Moreno-Hagelsieb, G., Paterson, A. D., & Croitoru, K. (2016). Association of host genome with intestinal microbial composition in a large healthy cohort. *Nature Genetics*, *48*(11), 1413–1417. https://doi.org/10.1038/ng.3693

Geuking, M. B., Cahenzli, J., Lawson, M. A. E., Ng, D. C. K., Slack, E., Hapfelmeier, S., McCoy, K. D., & Macpherson, A. J. (2011). Intestinal Bacterial Colonization Induces Mutualistic Regulatory T Cell Responses. *Immunity*, *34*(5), 794–806. https://doi.org/10.1016/j.immuni.2011.03.021

Goodrich, J. K., Davenport, E. R., Beaumont, M., Jackson, M. A., Knight, R., Ober, C., Spector, T. D., Bell, J. T., Clark, A. G., & Ley, R. E. (2016). Genetic Determinants of the Gut Microbiome in UK Twins. *Cell Host & Microbe*, *19*(5), 731–743. https://doi.org/10.1016/j.chom.2016.04.017

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., Spector, T. D., Clark, A. G., & Ley, R. E. (2014). Human Genetics Shape the Gut Microbiome. *Cell*, *159*(4), 789–799. https://doi.org/10.1016/j.cell.2014.09.053

Heaver, S. L., Johnson, E. L., & Ley, R. E. (2018). Sphingolipids in host–microbial interactions. *Current Opinion in Microbiology*, *43*, 92–99. https://doi.org/10.1016/j.mib.2017.12.011

Iida, N., Dzutsev, A., Stewart, C. A., Smith, L., Bouladoux, N., Weingarten, R. A., Molina, D. A., Salcedo, R., Back, T., Cramer, S., Dai, R.-M., Kiu, H., Cardone, M., Naik, S., Patri, A. K., Wang, E., Marincola, F. M., Frank, K. M., Belkaid, Y., … Goldszmid, R. S. (2013). Commensal Bacteria Control Cancer Response to Therapy by Modulating the Tumor Microenvironment. *Science*, *342*(6161), 967–970. https://doi.org/10.1126/science.1240527

Kamada, N., Chen, G. Y., Inohara, N., & Núñez, G. (2013). Control of pathogens and pathobionts by the gut microbiota. *Nature Immunology*, *14*(7), 685–690. https://doi.org/10.1038/ni.2608

Karl, J. P., Margolis, L. M., Madslien, E. H., Murphy, N. E., Castellani, J. W., Gundersen, Y., Hoke, A. V., Levangie, M. W., Kumar, R., Chakraborty, N., Gautam, A., Hammamieh, R., Martini, S., Montain, S. J., & Pasiakos, S. M. (2017). Changes in intestinal microbiota composition and metabolism coincide with increased intestinal permeability in young adults under prolonged physiological stress. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *312*(6), G559–G571. https://doi.org/10.1152/ajpgi.00066.2017

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. *Nature Communications*, *3*(1), 1245. https://doi.org/10.1038/ncomms2266

Kazemian, N., Mahmoudi, M., Halperin, F., Wu, J. C., & Pakpour, S. (2020). Gut microbiota and cardiovascular disease: Opportunities and challenges. *Microbiome*, *8*(1), 36. https://doi.org/10.1186/s40168-020-00821-0

Keane, T. M., Goodstadt, L., Danecek, P., White, M. A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., Furlotte, N. A., Eskin, E., Nellåker, C., Whitley, H., Cleak, J., Janowitz, D., Hernandez-Pliego, P., Edwards, A., Belgard, T. G., … Adams, D. J. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*, *477*(7364), 289–294. https://doi.org/10.1038/nature10413

Kemis, J. H., Linke, V., Barrett, K. L., Boehm, F. J., Traeger, L. L., Keller, M. P., Rabaglia, M. E., Schueler, K. L., Stapleton, D. S., Gatti, D. M., Churchill, G. A., Amador-Noguez, D., Russell, J. D., Yandell, B. S., Broman, K. W., Coon, J. J., Attie, A. D., & Rey, F. E. (2019). Genetic determinants of gut microbiota composition and bile acid profiles in mice. *PLOS Genetics*, *15*(8), e1008073. https://doi.org/10.1371/journal.pgen.1008073

Koppel, N., Maini Rekdal, V., & Balskus, E. P. (2017). Chemical transformation of xenobiotics by the human gut microbiota. *Science*, *356*(6344), eaag2770. https://doi.org/10.1126/science.aag2770

Lang, S., Demir, M., Martin, A., Jiang, L., Zhang, X., Duan, Y., Gao, B., Wisplinghoff, H., Kasper, P., Roderburg, C., Tacke, F., Steffen, H.-M., Goeser, T., Abraldes, J. G., Tu, X. M., Loomba, R., Stärkel, P., Pride, D., Fouts, D. E., & Schnabl, B. (2020). Intestinal Virome Signature Associated With Severity of Nonalcoholic Fatty Liver Disease. *Gastroenterology*, *159*(5), 1839–1852. https://doi.org/10.1053/j.gastro.2020.07.005

Lee, J.-Y., Tsolis, R. M., & Bäumler, A. J. (2022). The microbiome and gut homeostasis. *Science*, *377*(6601), eabp9960. https://doi.org/10.1126/science.abp9960

Leone, V., Gibbons, S. M., Martinez, K., Hutchison, A. L., Huang, E. Y., Cham, C. M., Pierre, J. F., Heneghan, A. F., Nadimpalli, A., Hubert, N., Zale, E., Wang, Y., Huang, Y., Theriault, B., Dinner, A. R., Musch, M. W., Kudsk, K. A., Prendergast, B. J., Gilbert, J. A., &

Chang, E. B. (2015). Effects of Diurnal Variation of Gut Microbes and High-Fat Feeding on Host Circadian Clock Function and Metabolism. *Cell Host & Microbe*, *17*(5), 681–689. https://doi.org/10.1016/j.chom.2015.03.006

Ley, R. E., Bäckhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences*, *102*(31), 11070–11075. https://doi.org/10.1073/pnas.0504978102

Ley, R. E., Peterson, D. A., & Gordon, J. I. (2006). Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine. *Cell*, *124*(4), 837–848. https://doi.org/10.1016/j.cell.2006.02.017

Ley, R. E., Turnbaugh, P. J., Klein, S., & Gordon, J. I. (2006). Human gut microbes associated with obesity. *Nature*, *444*(7122), Article 7122. https://doi.org/10.1038/4441022a

Li, H., Wang, X., Rukina, D., Huang, Q., Lin, T., Sorrentino, V., Zhang, H., Bou Sleiman, M., Arends, D., McDaid, A., Luan, P., Ziari, N., Velázquez-Villegas, L. A., Gariani, K., Kutalik, Z., Schoonjans, K., Radcliffe, R. A., Prins, P., Morgenthaler, S., … Auwerx, J. (2018). An Integrated Systems Genetics and Omics Toolkit to Probe Gene Function. *Cell Systems*, *6*(1), 90-102.e4. https://doi.org/10.1016/j.cels.2017.10.016

Loomba, R., Seguritan, V., Li, W., Long, T., Klitgord, N., Bhatt, A., Dulai, P. S., Caussy, C., Bettencourt, R., Highlander, S. K., Jones, M. B., Sirlin, C. B., Schnabl, B., Brinkac, L., Schork, N., Chen, C.-H., Brenner, D. A., Biggs, W., Yooseph, S., … Nelson, K. E. (2017). Gut Microbiome-Based Metagenomic Signature for Non-invasive Detection of Advanced Fibrosis in Human Nonalcoholic Fatty Liver Disease. *Cell Metabolism*, *25*(5), 1054-1062.e5. https://doi.org/10.1016/j.cmet.2017.04.001

Lopera-Maya, E. A., Kurilshikov, A., van der Graaf, A., Hu, S., Andreu-Sánchez, S., Chen, L., Vich Vila, A., Gacesa, R., Sinha, T., Collij, V., Klaassen, M. A. Y., Bolte, L. A., Brandao Gois, M. F., Neerincx, P. B. T., Swertz, M. A., LifeLines Cohort Study, Harmsen, H. J. M., Wijmenga, C., Fu, J., … Sanna, S. (2020). Effect of host genetics on the gut

microbiome in 7,738 participants of the Dutch Microbiome Project. *Nature Genetics*. https://doi.org/10.1101/2020.12.09.417642

Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K., & Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature*, *489*(7415), 220–230. https://doi.org/10.1038/nature11550

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., Williams, R. W., Benson, A. K., Peterson, D. A., & Ciobanu, D. C. (2012). Murine Gut Microbiota Is Defined by Host Genetics and Modulates Variation of Metabolic Traits. *PLoS ONE*, *7*(6), e39191. https://doi.org/10.1371/journal.pone.0039191

Meijnikman, A. S., Davids, M., Herrema, H., Aydin, O., Tremaroli, V., Rios-Morales, M., Levels, H., Bruin, S., De Brauw, M., Verheij, J., Kemper, M., Holleboom, A. G., Tushuizen, M. E., Schwartz, T. W., Nielsen, J., Brandjes, D., Dirinck, E., Weyler, J., Verrijken, A., … Nieuwdorp, M. (2022). Microbiome-derived ethanol in nonalcoholic fatty liver disease. *Nature Medicine*, *28*(10), 2100–2106. https://doi.org/10.1038/s41591-022-02016-6

Miller, S. I., Ernst, R. K., & Bader, M. W. (2005). LPS, TLR4 and infectious disease diversity. *Nature Reviews Microbiology*, *3*(1), 36–46. https://doi.org/10.1038/nrmicro1068

Org, E., Parks, B. W., Joo, J. W. J., Emert, B., Schwartzman, W., Kang, E. Y., Mehrabian, M., Pan, C., Knight, R., Gunsalus, R., Drake, T. A., Eskin, E., & Lusis, A. J. (2015). Genetic and environmental control of host-gut microbiota interactions. *Genome Research*, *25*(10), 1558–1569. https://doi.org/10.1101/gr.194118.115

Osborn, O., & Olefsky, J. M. (2012). The cellular and signaling networks linking the immune system and metabolism in disease. *Nature Medicine*, *18*(3), 363–374. https://doi.org/10.1038/nm.2627

Park, B. S., Song, D. H., Kim, H. M., Choi, B.-S., Lee, H., & Lee, J.-O. (2009). The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature*, *458*(7242), 1191–1195. https://doi.org/10.1038/nature07830

Perez-Munoz, M. E., McKnite, A. M., Williams, E. G., Auwerx, J., Williams, R. W., Peterson, D. A., & Ciobanu, D. C. (2019). Diet modulates cecum bacterial diversity and physiological phenotypes across the BXD mouse genetic reference population. *PLOS ONE*, *14*(10), e0224100. https://doi.org/10.1371/journal.pone.0224100

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., … Wang, J. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*, *490*(7418), 55–60. https://doi.org/10.1038/nature11450

Qin, Y., Havulinna, A. S., Liu, Y., Jousilahti, P., Ritchie, S. C., Tokolyi, A., Sanders, J. G., Valsta, L., Brożyńska, M., Zhu, Q., Tripathi, A., Vazquez-Baeza, Y., Loomba, R., Cheng, S., Jain, M., Niiranen, T., Lahti, L., Knight, R., Salomaa, V., … Méric, G. (2020). Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. *Nature Genetics*. https://doi.org/10.1101/2020.09.12.20193045

Rackaityte, E., Halkias, J., Fukui, E. M., Mendoza, V. F., Hayzelden, C., Crawford, E. D., Fujimura, K. E., Burt, T. D., & Lynch, S. V. (2020). Viable bacterial colonization is highly limited in the human intestine in utero. *Nature Medicine*, *26*(4), 599–607. https://doi.org/10.1038/s41591-020-0761-3

Raetz, C. R. H., & Whitfield, C. (2002). Lipopolysaccharide Endotoxins. *Annual Review of Biochemistry*, *71*(1), 635–700. https://doi.org/10.1146/annurev.biochem.71.110601.135414

Rocha, F. G., Ottenberg, G., Eure, Z. G., Davey, M. E., & Gibson, F. C. (2021). Sphingolipid-Containing Outer Membrane Vesicles Serve as a Delivery Vehicle To Limit Macrophage

Immune Response to Porphyromonas gingivalis. *Infection and Immunity*, *89*(4), e00614-20. https://doi.org/10.1128/IAI.00614-20

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P. I., Godneva, A., Kalka, I. N., Bar, N., Shilo, S., Lador, D., Vila, A. V., Zmora, N., Pevsner-Fischer, M., Israeli, D., Kosower, N., Malka, G., Wolf, B. C., … Segal, E. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature*, *555*(7695), 210–215. https://doi.org/10.1038/nature25973

Rühlemann, M. C., Hermes, B. M., Bang, C., Doms, S., Moitinho-Silva, L., Thingholm, L. B., Frost, F., Degenhardt, F., Wittig, M., Kässens, J., Weiss, F. U., Peters, A., Neuhaus, K., Völker, U., Völzke, H., Homuth, G., Weiss, S., Grallert, H., Laudes, M., … Franke, A. (2021). Genome-wide association study in 8,956 German individuals identifies influence of ABO histo-blood groups on gut microbiome. *Nature Genetics*, *53*(2), 147–155. https://doi.org/10.1038/s41588-020-00747-1

Saul, M. C., Philip, V. M., Reinholdt, L. G., & Chesler, E. J. (2019). High-Diversity Mouse Populations for Complex Traits. *Trends in Genetics*, *35*(7), 501–514. https://doi.org/10.1016/j.tig.2019.04.003

Schaid, D. J., Chen, W., & Larson, N. B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nature Reviews Genetics*, *19*(8), 491–504. https://doi.org/10.1038/s41576-018-0016-z

Sepich-Poore, G. D., Zitvogel, L., Straussman, R., Hasty, J., Wargo, J. A., & Knight, R. (2021). The microbiome and human cancer. *Science*, *371*(6536), eabc4552. https://doi.org/10.1126/science.abc4552

Sivan, A., Corrales, L., Hubert, N., Williams, J. B., Aquino-Michaels, K., Earley, Z. M., Benyamin, F. W., Man Lei, Y., Jabri, B., Alegre, M.-L., Chang, E. B., & Gajewski, T. F. (2015). Commensal Bifidobacterium promotes antitumor immunity and facilitates anti–

PD-L1 efficacy. *Science*, *350*(6264), 1084–1089. https://doi.org/10.1126/science.aac4255

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., Karpen, G. H., Celniker, S. E., Brown, J. B., Borenstein, E., Jansson, J. K., Metz, T. O., & Mao, J.-H. (2017). Influence of early life exposure, host genetics and diet on the mouse gut microbiome and metabolome. *Nature Microbiology*, *2*(2). https://doi.org/10.1038/nmicrobiol.2016.221

Song, Z., Cai, Y., Lao, X., Wang, X., Lin, X., Cui, Y., Kalavagunta, P. K., Liao, J., Jin, L., Shang, J., & Li, J. (2019). Taxonomic profiling and populational patterns of bacterial bile salt hydrolase (BSH) genes based on worldwide human gut microbiome. *Microbiome*, *7*(1), 9. https://doi.org/10.1186/s40168-019-0628-3

Sorbara, M. T., & Pamer, E. G. (2022). Microbiome-based therapeutics. *Nature Reviews Microbiology*, *20*(6), 365–380. https://doi.org/10.1038/s41579-021-00667-9

Suzuki, T. A., & Ley, R. E. (2020). The role of the microbiota in human genetic adaptation. *Science*, *370*(6521), eaaz6827. https://doi.org/10.1126/science.aaz6827

Swann, J. R., Want, E. J., Geier, F. M., Spagou, K., Wilson, I. D., Sidaway, J. E., Nicholson, J. K., & Holmes, E. (2011). Systemic gut microbial modulation of bile acid metabolism in host tissue compartments. *Proceedings of the National Academy of Sciences*, *108*(supplement_1), 4523–4530. https://doi.org/10.1073/pnas.1006734107

The Complex Trait Consortium. (2004). The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nature Genetics*, *36*(11), 1133–1137. https://doi.org/10.1038/ng1104-1133

The Integrative Human Microbiome Project: Dynamic Analysis of Microbiome-Host Omics Profiles during Periods of Human Health and Disease. (2014). *Cell Host & Microbe*, *16*(3), 276–289. https://doi.org/10.1016/j.chom.2014.08.014

Thomas, C., Gioiello, A., Noriega, L., Strehle, A., Oury, J., Rizzo, G., Macchiarulo, A., Yamamoto, H., Mataki, C., Pruzanski, M., Pellicciari, R., Auwerx, J., & Schoonjans, K. (2009). TGR5-Mediated Bile Acid Sensing Controls Glucose Homeostasis. *Cell Metabolism*, *10*(3), 167–177. https://doi.org/10.1016/j.cmet.2009.08.001

Thomas, D. (2010). Gene–environment-wide association studies: Emerging approaches. *Nature Reviews Genetics*, *11*(4), 259–272. https://doi.org/10.1038/nrg2764

Trasande, L., Blustein, J., Liu, M., Corwin, E., Cox, L. M., & Blaser, M. J. (2013). Infant antibiotic exposures and early-life body mass. *International Journal of Obesity*, *37*(1), 16–23. https://doi.org/10.1038/ijo.2012.132

Turnbaugh, P. J., Bäckhed, F., Fulton, L., & Gordon, J. I. (2008). Diet-Induced Obesity Is Linked to Marked but Reversible Alterations in the Mouse Distal Gut Microbiome. *Cell Host & Microbe*, *3*(4), 213–223. https://doi.org/10.1016/j.chom.2008.02.015

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The Human Microbiome Project. *Nature*, *449*(7164), 804–810. https://doi.org/10.1038/nature06244

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*(7122), 1027–1031. https://doi.org/10.1038/nature05414

Vatanen, T., Kostic, A. D., d'Hennezel, E., Siljander, H., Franzosa, E. A., Yassour, M., Kolde, R., Vlamakis, H., Arthur, T. D., Hämäläinen, A.-M., Peet, A., Tillmann, V., Uibo, R., Mokurov, S., Dorshakova, N., Ilonen, J., Virtanen, S. M., Szabo, S. J., Porter, J. A., … Xavier, R. J. (2016). Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. *Cell*, *165*(4), 842–853. https://doi.org/10.1016/j.cell.2016.04.007

Velagapudi, V. R., Hezaveh, R., Reigstad, C. S., Gopalacharyulu, P., Yetukuri, L., Islam, S., Felin, J., Perkins, R., Borén, J., Orešič, M., & Bäckhed, F. (2010). The gut microbiota

modulates host energy and lipid metabolism in mice. *Journal of Lipid Research*, *51*(5), 1101–1112. https://doi.org/10.1194/jlr.M002774

Vogt, N. M., Kerby, R. L., Dill-McFarland, K. A., Harding, S. J., Merluzzi, A. P., Johnson, S. C., Carlsson, C. M., Asthana, S., Zetterberg, H., Blennow, K., Bendlin, B. B., & Rey, F. E. (2017). Gut microbiome alterations in Alzheimer's disease. *Scientific Reports*, *7*(1), 13537. https://doi.org/10.1038/s41598-017-13601-y

Wahlström, A., Sayin, S. I., Marschall, H.-U., & Bäckhed, F. (2016). Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism. *Cell Metabolism*, *24*(1), 41–50. https://doi.org/10.1016/j.cmet.2016.05.005

Walker, A. W., Ince, J., Duncan, S. H., Webster, L. M., Holtrop, G., Ze, X., Brown, D., Stares, M. D., Scott, P., Bergerat, A., Louis, P., McIntosh, F., Johnstone, A. M., Lobley, G. E., Parkhill, J., & Flint, H. J. (2011). Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *The ISME Journal*, *5*(2), 220–230. https://doi.org/10.1038/ismej.2010.118

Wang, J., Thingholm, L. B., Skiecevičienė, J., Rausch, P., Kummen, M., Hov, J. R., Degenhardt, F., Heinsen, F.-A., Rühlemann, M. C., Szymczak, S., Holm, K., Esko, T., Sun, J., Pricop-Jeckstadt, M., Al-Dury, S., Bohov, P., Bethune, J., Sommer, F., Ellinghaus, D., … Franke, A. (2016). Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nature Genetics*, *48*(11), 1396–1406. https://doi.org/10.1038/ng.3695

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*(7341), 57–63. https://doi.org/10.1038/nature09922

Wu, H.-J., & Wu, E. (2012). The role of gut microbiota in immune homeostasis and autoimmunity. *Gut Microbes*, *3*(1), 4–14. https://doi.org/10.4161/gmic.19320

Yano, J. M., Yu, K., Donaldson, G. P., Shastri, G. G., Ann, P., Ma, L., Nagler, C. R., Ismagilov, R. F., Mazmanian, S. K., & Hsiao, E. Y. (2015). Indigenous Bacteria from the Gut Microbiota Regulate Host Serotonin Biosynthesis. *Cell*, *161*(2), 264–276. https://doi.org/10.1016/j.cell.2015.02.047

Yoon, H., Shaw, J. L., Haigis, M. C., & Greka, A. (2021). Lipid metabolism in sickness and in health: Emerging regulators of lipotoxicity. *Molecular Cell*, *81*(18), 3708–3730. https://doi.org/10.1016/j.molcel.2021.08.027

Zhang, Y.-M., & Rock, C. O. (2008). Membrane lipid homeostasis in bacteria. *Nature Reviews Microbiology*, *6*(3), 222–233. https://doi.org/10.1038/nrmicro1839

CHAPTER 2: Genetic mapping of microbial and host traits reveals production of immunomodulatory lipids by *Akkermansia muciniphila* in the murine gut

The work presented in this chapter has been published:

Zhang Q[*]*, Linke V*[*]*, Overmyer KA, Traeger LL, Kasahara K, Miller IJ, Manson DE, Polaske TJ, Kerby RL, Kemis JH, Trujillo EA, Reddy TR, Russell JD, Schueler KL, Stapleton DS, Rabaglia ME, Seldin M, Gatti DM, Keele GR, Pham DT, Gerdt JP, Vivas EI, Lusis AJ, Keller MP, Churchill GA, Blackwell HE, Broman KW, Attie AD, Coon JJ, Rey FE*[+]

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

Data and supplementary information available online

2.1 Abstract

The molecular bases of how host genetic variation impacts the gut microbiome remain largely unknown. Here we used a genetically diverse mouse population and applied systems genetics strategies to identify interactions between host and microbe phenotypes including microbial functions, using faecal metagenomics, small intestinal transcripts and caecal lipids that influence microbe–host dynamics. Quantitative trait locus (QTL) mapping identified murine genomic regions associated with variations in bacterial taxa; bacterial functions including motility, sporulation and lipopolysaccharide production and levels of bacterial- and host-derived lipids. We found overlapping QTL for the abundance of *Akkermansia muciniphila* and caecal levels of ornithine lipids. Follow-up in vitro and in vivo studies revealed that *A. muciniphila* is a major source of these lipids in the gut, provided evidence that ornithine lipids have immunomodulatory effects and identified intestinal transcripts co-regulated with these traits including *Atf3*, which encodes for a transcription factor that plays vital roles in modulating metabolism and immunity. Collectively, these results suggest that ornithine lipids are potentially important for *A. muciniphila*–host interactions and support the role of host genetics as a determinant of responses to gut microbes.

2.2 Introduction

The gut microbiome plays fundamental roles in mammalian physiology and human health (Tremaroli & Bäckhed, 2012; Turnbaugh et al., 2006; Wen et al., 2008). Environmental exposures and host genetic variation modulate gut microbiota composition (Ley et al., 2006; Rey et al., 2013; Yatsunenko et al., 2012) and contribute to the large degree of interpersonal variation observed in human gut microbial communities. Recent advances in sequencing technologies and analytical pipelines have fuelled progress in our understanding of the impact of host genetics and the gut microbiome on health. Population studies have revealed host genetic-gut microbial trait associations in human (Bonder et al., 2016; GEM Project Research Consortium et al., 2016; Hughes et al., 2020; Kurilshikov et al., 2021; J. Wang et al., 2016) and mouse cohorts (Kemis et al., 2019; Org et al., 2015). Additionally, studies leveraging host genetic information and Mendelian randomization have highlighted connections between the gut microbiome and other molecular complex traits including faecal levels of short-chain fatty acids (Sanna et al., 2019), plasma proteins (LifeLines cohort study et al., 2018) and ABO histo-blood group type (Rühlemann et al., 2021) in humans. However, most of these studies have focused on microbial organismal composition and there is currently a major gap in our understanding of the impact of host genetic variation on the functional capacity of the gut microbiome.

Microbial metabolites are critical nodes of communication between microbes and the host. These include small molecules derived from dietary components (for example, Trimethylamine N-oxide) (Z. Wang et al., 2011) or de novo synthesized by microbes such as vitamins (Kjer-Nielsen et al., 2012) and lipids (Brown et al., 2019). Lipids including eicosanoids, phospholipids, sphingolipids and fatty acids act as signalling molecules to control many cellular processes (Baxter et al., 2015; de Carvalho & Caramujo, 2018; Dennis & Norris, 2015). Gut microbes not only modulate absorption of dietary lipids via regulation of bile acid production and metabolism but are also a major source of lipids and precursor metabolites for lipids produced by the host (Kindt et al., 2018; Schoeler & Caesar, 2019). Bacterial cell membrane-associated lipids are also

important for microbe–host interactions (Brown et al., 2019; Kim et al., 2018), although our understanding of their roles in these dynamics is only emerging for gut bacteria.

Defining the general principles that govern microbe–host interactions in the gut ecosystem is a daunting task. Systems genetic studies can generate hypotheses that invoke processes and molecules that have no precedent, which can be used for the identification of genes, pathways and networks underlying these interactions. To investigate the connections between gut microbes, intestinal lipids and host genetic variation, we leveraged the Diversity Outbred (DO) mouse cohort, a genetically diverse population derived from eight founder strains: C57BL/6J (B6), A/J (A/J), 129S1/SvImJ (129), NOD/ShiLtJ (NOD), NZO/HLtJ (NZO), CAST/EiJ (CAST), PWK/PhJ (PWK) and WSB/EiJ (WSB) (Churchill et al., 2012; Svenson et al., 2012). These eight strains harbour distinct gut microbial communities and exhibit disparate metabolic responses to diet-induced metabolic disease (Kreznar et al., 2017). The DO population is maintained by an outbreeding strategy aimed at maximizing the power and resolution of genetic mapping. We characterized the faecal metagenome, intestinal transcriptome and caecal lipidome in DO mice and performed quantitative trait locus (QTL) analysis to identify host genetic loci associated with these traits. We integrated microbiome QTL (mbQTL) and caecal lipidome QTL (clQTL) to uncover microbe–lipid associations and identified candidate genes expressed in the distal small intestine associated with these co-mapping traits. These datasets represent a valuable resource for interrogating the molecular mechanisms underpinning interactions between the host and the gut microbiome.

2.3 Results

Gut microbial features are associated with host genetics

We characterized the faecal microbiome from 264 DO mice fed a high-fat high-sucrose (HF/HS) diet for ~22 weeks (Supplementary Figure 2.1). We and others previously showed that this diet elicits a wide range of metabolic responses in the eight founder strains that are associated with microbiome changes, and identified loci associated with variation in abundance of bacterial taxa in the gut (Kreznar et al., 2017; O'Connor et al., 2014); here we examine the role of host genetics in influencing gut microbiome traits with a focus on gut bacterial functions. Metagenomic analysis revealed ~1.9 million unique predicted microbial open reading frames (that is, metagenes), 2,803 bacterial functions (KEGG orthologues, KOs) and 187 bacterial taxa across all mice. We also performed metagenomic binning to obtain metagenome-assembled genomes (MAGs), corresponding to species-level bacterial genomes (Supplementary Figure 2.2, Supplementary Tables 1-4 and Supplementary Note 1).

We next used QTL analysis to identify regions of the mouse genome associated with the abundance of these traits. We detected 760 associations for KOs (logarithm of odds (LOD) > 6.87, $P_{\text{genome-wide-adj}} < 0.2$), 200 of which were genome-wide significant (LOD > 7.72, $P_{\text{genome-wide-adj}} < 0.05$) and 45 associations for bacterial taxa (LOD > 6.87, $P_{\text{genome-wide-adj}} < 0.2$), 15 of which were genome-wide significant (LOD > 7.72, $P_{\text{genome-wide-adj}} < 0.05$) (Figure 2.1a and Supplementary Tables 5 and 6). We identified a QTL hotspot on chromosome 15 at 63–64 Mbp; this genomic region was associated with 154 microbial traits with LOD score > 6 (Supplementary Table 7). We estimated DO founder allele effects as best linear unbiased predictors for the traits that mapped to this locus. Among these, we detected two clear groups of traits that exhibited opposite allele effects: a group of KOs and taxa showing positive association with the 129 allele, and another group of KOs and taxa that were negatively associated with the 129 allele (Supplementary Figure 2.3). As detailed below, the two most abundant gut bacterial phyla, Firmicutes and Bacteroidetes, mapped to this locus with opposite allele effects.

Pathway enrichment analysis showed that bacterial ‘motility proteins’ and ‘cell growth’ functional categories were significantly enriched in the group of KOs associated most strongly with 129 alleles (Figure 2.1b,c). More specifically, abundances of 14 sporulation functions were negatively associated with 129 alleles (Figure 2.1d). Further investigation of the KO distribution across all MAGs revealed that all bacterial sporulation KOs were only present in MAGs belonging to Firmicutes, whereas most of KOs that showed positive 129 allele effects were present in MAGs belonging to Bacteroidetes (Supplementary Figure 2.4a). To assess whether the allele effects observed from QTL mapping corresponded to the trait patterns in the DO founder strains, we examined previously published 16S ribosomal RNA gene data from age-matched mice from the eight founder strains, also fed an HF/HS diet (Kemis et al., 2019). Consistent with these findings, we found that the 129 mouse strain had higher levels of Bacteroidetes and the highest Bacteroidetes/Firmicutes ratio (Supplementary Figure 2.4b). Interestingly, we detected a significant positive correlation between the number of sporulation KOs in Firmicutes MAGs mapping at this locus and the LOD scores for these MAGs (Figure 2.1e). Importantly, Firmicutes MAGs commonly detected in our dataset that do not contain sporulation KOs (for example, *Lactobacillus*, *Lactococcus*) did not exhibit significant association to this QTL. These results support the notion that host genetic variation affects gut community structure in part by modulating the abundance of sporulating bacteria.

Single nucleotide polymorphism (SNP) association analysis within the Chr15 QTL hotspot identified six significant SNPs: two intron variants, SNP rs582880514 in the *Gsdmc* gene and SNP rs31810445 in the *Gsdmc2* gene, both with LOD scores of 8.0; four SNPs that were intergenic variants (Supplementary Figure 2.4c). Gasdermins (Gsdm) are a family of pore-forming proteins that cause membrane permeabilization and pyroptosis (Shi et al., 2015), an inflammatory form of programmed cell death that is triggered by intra- and extracellular pathogens (Liu et al., 2021). These results indicate that host genetic variation in Gsdmc/Gsdmc2 is associated with

abundance of gut bacterial functions and raises the hypothesis that these host proteins could modulate the abundance of bacterial taxa harbouring motility and/or sporulation functions.

Caecal lipids are associated with gut microbes and host genetics

We employed a broad discovery strategy to agnostically detect lipid actors potentially relevant to gut microbiome–host interactions. We used liquid chromatography coupled with tandem mass spectrometry (LC–MS/MS) to characterize the caecal lipidome of 381 DO mice, including all mice used for the metagenomic analysis. We identified 1,048 lipid species representing 35 lipid classes (Figure 2.2a,b) and the four major lipid categories: (1) fatty acyls, (2) phospholipids, (3) sphingolipids and (4) glycerolipids. The highest numbers of lipids were recorded for the classes of triglycerides (TG) and phosphatidylcholines (PC), species known to be abundant in the mammalian host (Jain et al., 2014). Of the 3,384 lipid species detected in DO caecum, 547 (16.2%) were detected at higher levels in the caecum of conventionally raised mice compared with caecum of germ-free animals (fold-change >10-fold, adjusted $P < 0.05$). Phosphatidylglycerols (PG), for example, which represent the second largest phospholipid class in our data, are known to be a major component of the bacterial lipidome (Sohlenkamp & Geiger, 2016). In mammals, on the other hand, PG are only a minor component. Similarly, among glycerolipids, monogalactosyldiacylglycerols (MGDG) account for the second highest number of lipids detected in this class. While they are found at high levels in bacteria and plants, these lipids are only minor components of animal tissue (Parsons & Rock, 2013). These findings suggest that our analysis of the caecal lipidome captures components of the host and the gut microbiome. Correlation analysis between MAGs and caecal lipids abundance, plus comparison of the caecal lipidome of conventionally raised vs germ-free mice identified taxa that potentially modulate the abundance of lipids in the gut (Supplementary Figure 2.5a,b, Supplementary Tables 8-10 and Supplementary Note 2). Furthermore, QTL mapping identified 399 significant QTL associations for caecal lipid features (LOD > 7.60, $P_{\text{genome-wide-adj}} < 0.05$) (Figure 2.2c, Supplementary

Table 11 and Supplementary Note 3). Altogether these associations provide a wealth of information offering potential molecular descriptors of the genetic regulation of the microbiome.

Mediation analysis reveals bacteria–caecal lipids connections

To identify causal links between gut microbial traits and caecal lipid traits, we performed mediation analysis between individual gut microbial metagenes and lipid features that co-map (Methods). Mediation analysis seeks to determine whether a QTL has separate effects on two traits, or if it affects one trait through its effect on another trait, in which case the intermediate trait is called a mediator. Figure 3a shows gut microbial metagenes mediating the QTL effect on a caecal lipid trait. We reasoned that if a microbial trait influenced a caecal lipid that was independent from the caecal lipid's QTL, its inclusion as a covariate would be unlikely to affect the caecal lipid QTL signal significantly. However, for microbial traits that mediate the QTL effect on the caecal lipid, there would be a large drop in the original caecal lipid QTL LOD score. Interestingly, we found three caecal lipid features with QTL that were mediated by microbial metagenes. Most of these mediating microbial traits were genes belonging to the bacterium *Akkermansia muciniphila*. It is important to note that the direction of the causal effect between microbial trait and caecal lipid cannot be directly inferred from the data. These results suggest that *A. muciniphila* levels and the abundance of these lipid species in the gut are modulated by the same loci and that the two traits are potentially connected (Figure 2.3b,c).

We further tested whether these caecal lipids and *A. muciniphila* mapped to the same loci. Mapping of the 46 reconstructed *A. muciniphila* MAGs to the host genome revealed multiple QTL including Chr1: 92.9 Mbp, Chr2: 79.4 Mbp, Chr7: 129.8 Mbp, Chr12: 59.4 Mbp, and Chr15: 75.9 Mbp (Figure 2.3d). Interestingly, the three caecal lipids also showed QTL at the same loci and exhibited similar founder allele effect patterns (Figure 2.3e). These founder allele effects on *A. muciniphila* abundance are consistent with a previous study of gut bacterial abundance in the DO founder strains (Kemis et al., 2019). Although these lipid features were not initially identified by

our lipidomic analysis pipeline, they appeared to be closely related to each other. Further analysis of their fragmentation spectra suggested that these unidentified features were ornithine lipids (OL) (Figure 2.3b and Supplementary Note 4). This was confirmed with a synthetic OL (see below). The three features would have the sum compositions of OL 30:0, OL 31:0 and OL 32:0, detected as [M+H]+ ions. In OL, a 3-hydroxy fatty acid is connected via an amide linkage to the ornithine amino acid that serves as the headgroup. A second fatty acid is then connected to the first via an ester linkage (Vences-Guzmán et al., 2012). OL are bacteria-specific non-phosphorus glycolipids that are found in the outer membranes of selected Gram-negative bacteria (Geiger et al., 2010; López-Lara et al., 2003).

A. muciniphila produces OL in the mouse and human gut

A. muciniphila is a Gram-negative bacterium that has been associated with many beneficial effects on host metabolic health (Depommier, 2019; Everard et al., 2013). While previous research suggests that OL are important for microbe–host interactions (Diercks et al., 2015; Kim et al., 2018), the occurrence of these lipids in gut bacteria was not known. To test whether *A. muciniphila* produces OL, we first profiled lipids in *A. muciniphila* and two other Gram-negative species, *Bacteroides thetaiotaomicron* and *Escherichia coli* grown under anaerobic conditions. We found similarly high levels of all three targeted OL species in extracts from *A. muciniphila* but not in the other species, which were indistinguishable from the solvent blank (Figure 2.4a). Since phosphate limitation triggers production of OL in some bacterial species (Kim et al., 2018), in follow-up experiments we tested whether phosphate levels modulated abundance of OL in *A. muciniphila* grown in vitro. We examined three different levels of phosphate (0.02 mM (growth limiting), 0.2 mM (adequate) and 2 mM (excess)). LC–MS/MS analysis confirmed that OL are a dominant lipid species detected in *A. muciniphila* cell extracts regardless of the phosphate levels included in the growth media (Supplementary Figure 2.6a,b). Furthermore, OL were detected in extracellular vesicles isolated from *A. muciniphila* grown in vitro (Supplementary

Figure 2.6c and Supplementary Note 6). These results suggest that OL are probably localized in the *A. muciniphila* outer membranes and provide insights into how these lipids may interact with the host.

We further profiled lipids produced by *A. muciniphila* colonizing the gut of gnotobiotic mice. Five groups of adult germ-free B6 mice were mono-colonized with each of the species mentioned above, bi-associated with *E. coli* and *A. muciniphila* or kept germ-free ($n = 3–5$ per group). Mice were maintained in the same HF/HS diet used for the DO study for two weeks after inoculation. LC–MS/MS analysis of caecal contents from these mice showed that only mice colonized with *A. muciniphila* had detectable levels of OL in their caecum (Figure 2.4b). Altogether, these results confirm that *A. muciniphila* gut colonization is causally linked with high levels of OL.

We examined whether *A. muciniphila* colonization is associated with the presence of OL in the human gut. We analysed lipid content in a subset of faecal samples from a previously characterized cohort of old adults (Dill-McFarland et al., 2019) spanning a wide range of *A. muciniphila* relative abundances (not detectable to 39.8%). LC–MS/MS analysis of these human faecal samples detected a broader range of OL species than axenic cultures or mice colonized with *A. muciniphila*, but the levels of the three previously identified OL 15:0_15:0, OL 16:0_15:0 and OL 17:0_15:0 were all significantly correlated with *A. muciniphila* levels (Figure 2.4c). Together, these results suggest that *A. muciniphila* is a major producer of OL in the mouse and human gut.

OL modulate lipopolysaccharide (LPS)-induced cytokine responses

To test whether *A. muciniphila*-derived OL elicit immune responses on the host, we first chemically synthesized the most abundant OL detected in the DO mouse gut, that is, OL_15:0_15:0. Because of the generally beneficial effects of *A. muciniphila* on host health as previously documented in both human and mouse studies, and on the basis of the structural similarity between OL and lipid A from LPS, we speculated that the OL might function as

antagonists of lipid A. We examined the effects of the OL preparation in the absence and presence of LPS on cytokine production by bone-marrow-derived-macrophages (BMDM). Treatment with LPS induced a significant increase in the production of TNF-α and IL-6 by BMDM obtained from B6 and 129 mice (Supplementary Figure 2.7a). In contrast, treatment with OL preparation did not stimulate significant production of TNF-α and IL-6 by these cells (Supplementary Figure 2.7b), except for a modest increase at 500 ng ml^{-1} and 1,000 ng ml^{-1}. However, we observed that pretreatment of macrophages with OL had an inhibitory effect on LPS-induced TNF-α and IL-6 in both B6 and 129 mice without causing significant changes in cell viability (Supplementary Figure 2.7c,d). These results suggest that *A. muciniphila*-derived OL can prevent LPS-induced inflammation response. Furthermore, we measured other cytokines secreted by LPS-treated BMDM and observed that the OL preparation inhibited the production of IL-1β, MCP-1, MIP-1α, GM-CSF, IL-12 and RANTES (Figure 2.5), although there were differences in the responses to LPS and OL as a function of BMDM genetic background. In addition, OL increased the levels of anti-inflammatory cytokine IL-10 in these cells (Figure 2.5), suggesting that OL may modulate inflammation by altering the levels of both pro-inflammatory and anti-inflammatory cytokines. Interestingly, production of IL-12 in the presence of LPS was more than ten times higher in 129 mice than in B6 mice, and OL had a larger inhibitory effect in these mice (Figure 2.5). These results indicate that *A. muciniphila*-derived OL may influence host innate immune responses and their effects may vary as a function of host genetics.

Intestinal genes co-map with A. muciniphila and OL QTL

We sought to generate regulatory maps of gene expression regulation in the small intestine and to identify overlapping SNPs associated with gut microbiome. We reasoned that identifying genes whose expression demonstrate shared genetic architecture with bacterial taxa/genes/lipids would not only narrow the list of candidate genes at each locus but would also provide invaluable insights into the biology underlying the microbe–host interactions. Furthermore,

the power of expression QTL (eQTL) mapping to connect genetic polymorphism and complex traits has been well documented by others (Gusev et al., 2016; Tian et al., 2015). We profiled transcript levels in the distal small intestines of 234 DO mice using RNA-seq. We detected 8,137 transcripts with a minimum of ten counts per million (CPM) in at least 10% of DO mice. We identified 4,462 local eQTL with an average LOD score of 21.2 and 10,894 distal eQTL with an average LOD score of 7.1 (Supplementary Table 12). By comparing eQTL allele effects with those for the co-mapping mbQTL and clQTL, we identified gut microbial features and caecal lipids that were potentially co-regulated with intestinal transcripts (Supplementary Figure 2.8 and Supplementary Note 7).

We searched the support intervals for the five co-mapping QTL regions for *A. muciniphila* and OL (Chr1, Chr2, Chr7, Chr12 and Chr15) for candidate host genes of interest using the eQTL data. By comparing the allele effects between co-mapping eQTL and the *A. muciniphila*/OL QTL, we identified several candidate host genes whose eQTL allele effects were correlated with *A. muciniphila*/OL (Figure 2.6, Supplementary Figure 2.9 and Supplementary Table 13). At the Chr1 QTL region, there were four candidate genes: (1) Gene Activating transcription factor 3 (*Atf3*) had a distal eQTL at Chr1: 92.96 Mbp with QTL LOD score of 6.55. ATF3 plays an important role during host immune response events by negatively regulating the transcription of pro-inflammatory cytokines induced by the activation of toll-like receptor 4 (Gilchrist et al., 2006). (2) The gene TRAF-interacting protein with a forkhead-associated domain (*Tifa*) had a distal eQTL at Chr1: 90.95 Mbp with LOD score of 6.19. TIFA has been reported to sense bacterial-derived heptose-1,7-bisphosphate—an intermediate in the synthesis of LPS—via a cytosolic surveillance pathway triggering the NF-kB response (Gaudet et al., 2015; Zhou et al., 2018). Additionally, TIFA interacts with TRAF6 to mediate host innate immune responses. (3) The gene Jumonji domain-containing protein 8 (*Jmjd8*) had a distal eQTL at Chr1: 92.14 Mbp with LOD score of 6.72. JMJD8 functions as a positive regulator of TNF-induced NF-kB signalling (Yeo et al., 2016). A recent study showed that JMJD8 is required for LPS-mediated inflammation and

insulin resistance in adipocytes (You et al., 2021). (4) The gene *Gcg* had a distal eQTL at Chr1: 92.36 Mbp with LOD score of 7.11. *Gcg* encodes multiple peptides including glucagon, glucagon-like peptide-1(GLP-1). GLP-1 levels are induced by a variety of inflammatory stimuli, including endotoxin, IL-1β and IL-6 (Kahles et al., 2014). The finding that these genes with distal eQTL that co-map with *A. muciniphila* and OL QTL on Chr1 are involved in host immune responses to microbial-associated molecular patterns (MAMPs) such as LPS suggests that expression of these genes contributes to the regulation of host responses to OL and/or potentially modulates the abundance of *A. muciniphila*.

Dissecting the link between A. muciniphila and Atf3

We investigated whether the co-mapping between *A. muciniphila*/OL QTL and *Atf3* gene eQTL could be explained by ATF3 impacting the abundance of these traits. To address this question, we measured the abundance of this taxon in wild-type (WT) mice and animals lacking the *Atf3* gene consuming HF/HS diet for four weeks. We observed that *Atf3*$^{-/-}$ and WT mice had comparable levels of *A. muciniphila* in faeces as detected by qPCR. Abundance of *A. muciniphila* was ~15% lower in faecal samples from *Atf3*$^{-/-}$ mice compared with wild type ($n = 7$ per genotype), yet the differences did not reach significance (Supplementary Figure 2.10a). These results suggest that *Atf3* does not play a major role in *A. muciniphila* fitness. It might also act in combination with other factors, which would align with the observation that the abundance of gut *A. muciniphila* is a polygenic trait.

An alternative explanation for the observed co-mapping is that *A. muciniphila*/OL modulate expression of *Atf3*. To examine this idea, we assessed expression profiles of B6 and 129 BMDM stimulated with LPS or a combination of the OL preparation and LPS. DESeq2 analysis identified 674 genes differentially expressed in cells from B6 mice treated with OL (420 genes were upregulated and 254 genes downregulated), whereas 384 genes (304 genes were upregulated and 80 genes downregulated) were impacted by OL in BMDM derived from 129 mice. While

differences in gene expression of some of the cytokines discussed above (Supplementary Figure 2.10b) were consistent between genotypes, the overall overlap of differentially expressed genes between genotypes was relatively low (Supplementary Figure 2.10c) and the responses to the OL varied significantly by genotype (Supplementary Figure 2.10e). As mentioned above, ATF3 is a negative regulator of TLR4 signalling. We observed that OL upregulated *Atf3* expression for both B6 and 129 BMDMs (Supplementary Figure 2.10d). Furthermore, a previous study (Labzin et al., 2015) identified 30 genes downregulated by ATF3 in BMDMs (B6 background). Consistent with this result, we found that OL downregulated the expression of these genes in BMDM derived from B6 mice. In contrast, we found that 18 out of these 30 genes were upregulated by OL in BMDM from 129 mice (Supplementary Figure 2.10f). These results suggest that the observed co-mapping between *A. muciniphila*/OL QTL and *Atf3* eQTL could be explained by the effect of OL on *Atf3* gene expression and that increased expression of this gene may trigger distinct programmes as a function of host genotype potentially impacting immune and metabolic responses differently.

Altogether, the work supports the notion that *A. muciniphila* is the major producer of caecal OL in the distal gut and that *A. muciniphila*-produced OL can negatively regulate host LPS-induced inflammation by upregulating *Atf3* expression.

2.4 Discussion

We applied a systems genetics approach to identify relationships between gut microbes, their encoded functions, caecal lipids and host intestinal gene expression. We identified bacterial functions influenced by host genetic variation and discovered that the bacterium *A. muciniphila* produces immunoactive OL that are detected in faecal samples from humans and mice colonized with this bacterium. *A. muciniphila* has been previously associated with host genetic variation at several loci in both mice and humans (Benson et al., 2010; Kurilshikov et al., 2021; Leamy et al., 2014; Org et al., 2015); however, environmental conditions including diet, which is a major known determinant of microbiome composition, differ dramatically among these studies. The associations described in the present study differ from the ones previously reported in other mouse studies using different diets (Benson et al., 2010; Org et al., 2015). We also examined whether gut microbiome traits acted as mediator to previously published metabolic traits for the same cohort of DO mice (Keller et al., 2018); however, no significant mediation was detected, possibly due to the limited statistical power of our study to infer the influence of the gut microbiome on complex metabolic traits.

Previous work suggested that some Gram-negative bacteria produce OL under phosphate-limiting conditions (Dees & Shively, 1982; Kawai et al., 1988; Vences-Guzmán et al., 2015). In contrast, we observed that OL levels were consistently high across a 100-fold phosphate level range, suggesting that phosphate is not a major driver of OL synthesis in *A. muciniphila*. Notably, a recent study showed that increased OL production by the bacterial pathogen *Pseudomonas aeruginosa* makes its cellular surface more hydrophobic, and resulted in lower virulence and higher resistance to antimicrobials and host immune defences (Kim et al., 2018). *A. muciniphila* consumes host glycans present in the mucus layer, which is in proximity to the host epithelium. While mucin carbohydrates and amino acids serve as substrates for *A. muciniphila*, there are also soluble host defence molecules trapped in this layer that prevent invasion of microbes to the underlying mucosal epithelial cells. We speculate that membrane OL

impact interactions of *A. muciniphila* with the intestinal milieu and may represent an adaptation critical to its niche and important for its interactions with the host. Development of tools to genetically manipulate *A. muciniphila* will be needed to test these hypotheses.

The inhibitory effects of OL on LPS-induced cytokines that we and others have observed (Kawai et al., 1991; Kawai & Akagawa, 1989) may represent an important aspect of how *A. muciniphila* impact host physiology. Previous studies identified both natural and synthetic molecules that can inhibit TLR4-mediated LPS signalling—compounds that prevent septic shock, and have anti-inflammatory and anti-neuropathic pain activities in vivo (Peri et al., 2010). One group of LPS antagonist molecules targeting CD14 shares structural features with *A. muciniphila* OL including a glucose unit linked to two hydrophobic chains and a basic nitrogen on C-6 (Piazza et al., 2009), supporting the potential anti-inflammatory effects of OL. Although the precise mechanisms of how OL inhibit LPS signalling are unknown, our study suggests that *A. muciniphila*-derived OL may modulate inflammatory responses.

Remarkably, three host innate immunity genes—*Atf3, Tifa* and *Jmjd8*—were co-regulated with *A. muciniphila*. *Tifa* is located in the 'cytokine-dependent colitis susceptibility locus' (*Cdcs1*) region, a critical genetic determinant of colitis susceptibility in 129 and B6 strains (Ryzhakov et al., 2018). TIFA is an important modifier of innate immune signalling through its regulation of TRAF proteins, leading to the activation of NF-κB and inflammation. Considering the importance of TIFA-dependent immunity to Gram-negative bacteria (Gaudet et al., 2015), and the differential effects of OL on LPS-treated BMDM from 129 and B6 strains, our results suggest that this gene could be a key player in *A. muciniphila*-OL–host interactions. Previous studies suggested that ATF3 modulates inflammatory responses by suppressing the expression of TLR4 or CCL4 in macrophages (Gilchrist et al., 2006; Khuu et al., 2007) and revealed a critical role of microbiota in ATF3-mediated gut homoeostasis (Cao et al., 2020). These studies showed that ATF3 negatively regulates *Il6* and *Il12* gene expression levels (Gilchrist et al., 2006). In line with this, we found that OL negatively influence these cytokines in LPS-treated BMDM, and their

abundance is associated with the same locus that influences *Atf3* expression. Previous studies also showed that ATF3 positively regulates host expression of antimicrobial peptides (Du et al., 2022) and suggested that the production of OL makes the bacterium *P. aeruginosa* more hydrophobic and resistant to cationic antimicrobial peptides (Kim et al., 2018). However, we observe neither co-mapping of *A. muciniphila* with expression of antimicrobial peptides nor pronounced differences in *A. muciniphila* colonization levels between *Atf3*$^{-/-}$ mice and WT littermates. Instead, the co-mapping of *A. muciniphila* and *Atf3* could be explained by our findings suggesting that (1) *A. muciniphila* is a major producer of OL in the gut and (2) OL upregulate expression of this key regulator. Although the molecular mechanisms underlying these observations warrant further investigation, these results suggest that *A. muciniphila* and OL levels are linked to central players of the host immune defence system and support the predominant role of host genetics as a determinant of responses to gut microbes, in particular to *A. muciniphila*.

In summary, the work presented here links the presence of OL in the human and mouse gut with *A. muciniphila* and suggests that these lipids are key players in *A. muciniphila*–host interactions. Our work highlights the importance of bacterial functions and lipids as mediators of the influence of host genetics on the gut microbiome.

2.5 Methods

Animal studies

Animal care and study protocols were approved by the AAALAC-accredited Institutional Animal Care and Use Committee of the College of Agricultural Life Sciences at the University of Wisconsin-Madison (UW-Madison). All experiments with mice were performed under protocols approved by the UW-Madison Animal Care and Use Committee (Protocol number A005821 for the DO mice, Protocol number M00559 for gnotobiotic and *Atf3* KO mice).

DO mouse model

DO mice were obtained from the Jackson Laboratory at ~four weeks of age and maintained in the Department of Biochemistry vivarium at the UW-Madison. DO mice were allocated in waves of 100 animals, each with an equal number of males and females. All mice were maintained in a temperature (22.2 °C) and humidity (60%) controlled environment under a 12 h light/dark cycle (lights on at 6:00 and off at 18:00). All mice were fed an HF/HS diet (TD.08811, Envigo Teklad, 44.6% kcal fat, 34% carbohydrate and 17.3% protein) and received sterilized water ad libitum upon arrival at the facility. Mice were kept in the same vivarium room and were individually housed to monitor food intake and prevent cross-inoculation via coprophagy. DO mice were killed at 22–25 weeks of age. Faecal samples were collected immediately before euthanasia after a four h fast. Caecal contents and additional tissues were collected promptly after killing and all samples were immediately flash frozen in liquid nitrogen and stored at −80 °C until further processing. Other studies have been published with these mice (Keller et al., 2018, 2019; Kemis et al., 2019; Linke et al., 2020).

Gnotobiotic studies

C57BL/6J germ-free mice were bred and housed in the gnotobiotic mouse facility at the UW-Madison. Male mice were used for the ornithine lipid study. All mice were maintained in a

controlled environment (22.2 °C, 60% humidity) in plastic flexible film gnotobiotic isolators under a strict 12 h light/dark cycle (lights on at 6:00 and off at 18:00) on standard chow diet (LabDiet 5021). At eight weeks of age, mice were switched to a western-style HF/HS diet (44.6% kcal fat, 34% carbohydrate and 17.3% protein) from Envigo Teklad (TD.08811) and orally gavaged with 200 µl of bacterial cultures. At two weeks after colonization, mice were euthanized and caecal contents collected.

DO founder mice

C57BL6J (B6) and 129S1/SvImJ (129) male mice (five weeks old) were obtained from the Jackson Laboratory. All mice were maintained in a controlled environment (22.2 °C, 60% humidity) under a strict 12 h light/dark cycle (lights on at 6:00 and off at 18:00). All mice were fed a standard chow diet (LabDiet 5021) and received sterilized water ad libitum for 1 week. At six weeks of age, all mice were euthanized to collect bone marrow cells.

Atf3 mouse studies

Atf3 heterozygous mice (B6.129X1-*Atf3*^{tm1Dron}/HaiMmnc) were obtained from the Mutant Mouse Resource and Research Center at University of North Carolina. Age- and sex-matched littermates of Atf3-deficient whole body knockout mice (*Atf3*$^{-/-}$) and WT mice were generated by crossing *Atf3* heterozygous mice. Mice were maintained in a controlled environment under a strict 12 h light/dark cycle (lights on at 6:00 and off at 18:00) at 22.2 °C and 60% humidity. Animals were fed an HF/HS diet (TD.08811, Envigo Teklad, 44.6% kcal fat, 34% carbohydrate and 17.3% protein) and received sterilized water ad libitum after weaning. Faecal samples were collected at seven weeks of age.

Metagenomic shotgun DNA sequencing

Faecal DNA was extracted from individual pellets collected from DO mice using previously described methods (Kreznar et al., 2017; Turnbaugh et al., 2009). Following DNA extraction, Illumina paired-end (PE) libraries were constructed using a previously described protocol (Faith et al., 2011), with a modification of gel selecting DNA fragments at ~450 bp length. PE reads (2 × 125) were generated using a combination of MiSeq and HiSeq 2500 platforms.

Metagenomic reads processing

Raw reads were preprocessed using Fastx Toolkit (v0.0.13) as follows: (1) for demultiplexing raw samples, fastx_barcode_splitter.pl with –partial 2, mismatch 2 was used; (2) when more than one forward and reverse read file existed for a single sample (due to being run on more than one lane, more than one platform or at more than one time), read files were concatenated into one forward and one reverse read file; (3) barcodes were trimmed to form reads (fastx_trimmer -f 9 -Q 33) and (4) reads were trimmed to remove low-quality sequences (fastq_quality_trimmer -t 20 -l 30 -Q33). Following trimming, unpaired reads were eliminated from the analysis using custom Python scripts. To identify and eliminate host sequences, reads were aligned against the mouse genome (mm10/GRCm38) using bowtie2 (v2.3.4) (Langmead & Salzberg, 2012) with default settings, and microbial DNA reads that did not align with the mouse genome were identified using samtools (v1.3) (samtools view -b -f 4 -f 8).

Metagenomic de novo assembly and gene prediction

After removing low-quality sequences and host contaminating DNA sequences, each metagenomic sample was de novo assembled into longer DNA fragments (contigs) using metaSPAdes (v3.11.1) (Nurk et al., 2017) with multiple *k*-mer sizes (metaspades.py -k 21, 33, 55, 77). Contigs shorter than 500 bp were discarded from further processing. Open reading frames (ORFs) (that is, microbial genes, also called metagenes) were predicted from assembled contigs via Prodigal (v2.6.3) (Hyatt et al., 2010) using Hidden Markov Model (HMM) with default

parameters. All predicted genes shorter than 100 bp were discarded from further processing. To remove redundant genes, all predicted ORFs were compared pairwise using the criterion of 95% identity at the nucleotide level over 90% of the length of the shorter ORFs via CD-HIT (v4.6.8) (W. Li & Godzik, 2006). In each CD-HIT cluster, the longest ORF was selected as representative. This final non-redundant (NR) microbial gene set was defined as the DO gut microbiome NR gene catalogue.

Metagenomic annotation

Gene taxonomic annotation was performed using DIAMOND (v0.9.23) (Buchfink et al., 2015) by aligning genes in the DO gut microbiome NR gene catalogue with the NCBI NR database (downloaded 21 December 2018) using default cutoffs: *e*-value $<1 \times 10^{-3}$ and bit score >50. Taxonomic assignment used the following parameters: '–taxonmap prot.accession2taxid.gz–taxonnodes nodes.dmp' in DIAMOND command and was determined by the lowest common ancestor (LCA) algorithm when there were multiple alignments. Gene functional annotation was done using the KEGG orthology and links annotation (KOALA) method via the KEGG server (https://www.kegg.jp/ghostkoala/), using 2,698,820 prokaryote genus pan-genomes as reference. The bit score cut-off for *K*-number assignment was 60.

Microbiome trait quantification

Quantification of microbial genes was done by aligning clean PE reads from each sample with the DO gut microbiome NR gene catalogue using Bowtie2 (v2.3.4) and default parameters. RSEM (v1.3.1) (B. Li & Dewey, 2011) was used to estimate microbial gene abundance. Relative abundances of microbial gene CPM were calculated using microbial gene expected counts divided by gene effective length, then normalized by the total sum. We focused the taxonomic analysis on bacteria since these represented the vast majority of annotated metagenes. We detected 1,927,034 total metagenes and from these, 1,636,209 were annotated as bacterial

genes, 195 as archaeal genes, 17,372 as eukaryotic genes and 946 as viruses. There were also 272,312 genes that were unclassified. To obtain abundance information for microbial functions, the CPM of genes with the same KO annotation were summed together. In case there were multiple KO annotations for a single gene, we used all KO annotations. To obtain taxonomic abundance, the CPM of genes with the same NCBI taxa annotation were summed together at phylum, order, class, family and genus levels, with a minimum of ten genes per taxon.

MAGs reconstruction

To reconstruct bacterial genomes, we clustered assembled contigs with the density-based algorithm DBSCAN using features of two reduced dimensions of contigs 5-mer frequency and contig coverage. The binning process was performed by the pipeline Autometa (docker image: ijmiller2/autometa:docker_patch) (Miller et al., 2019) and allowed deconvolution of taxonomically distinct microbial genomes from metagenomic sequences. The quality of reconstructed metagenomes was evaluated using CheckM (v1.1.3) (Parks et al., 2015); genome completeness >90% and genome contamination <5% were required to assign high-quality MAGs. MAGs quantification was done by aligning all clean PE reads from each sample with MAGs from the same sample. Genome coverage was calculated using the bedtools (v2.29.2) 'genomecov' command, followed by normalization by library size across all samples. To further remove redundant MAGs, we clustered high-quality MAGs on the basis of whole-genome nucleotide similarity estimation (pairwise average nucleotide identity (ANI)) using Mash software (v2.2) (Ondov et al., 2016) with 90% ANI. From high-quality MAGs, we also annotated predicted ORFs from each MAG against the KEGG database and compared the functional potential encoded among different taxa. *A. muciniphila* MAG IDs are included in Supplementary Table 14.

Sample preparation for caecal lipidomic analysis

Caecal contents (30 ± 7.5 mg) along with 10 µl SPLASH Lipidomix internal standard mixture were aliquoted into a tube with a metal bead and 270 µl methanol (MeOH) were added for protein precipitation. Control samples comprised 30 ± 7.5 mg of bead beat-combined DO founder strain caecum (NZO, PWK, NOD, B6, 129, AJ) extracted with each batch. To each tube, 900 µl methyl tert-butyl ether (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 eight min at 25 Hz. Finally, the mixture was centrifuged for eight min at 11,000 × *g* at 4 °C, after which 240 µl of the lipophilic upper layer were transferred to glass vials and dried by vacuum centrifuge for 60 min.

The dried lipophilic extracts were re-suspended in 200 µl MeOH:toluene (9:1 v/v) per 10 mg dry weight (minimum of 100 µl) to account for varying water content in the samples. The dry weight was determined by drying down the remaining mixture including all solid parts.

LC–MS/MS analysis of DO mouse caecal samples

Sample analysis by LC–MS/MS was performed in randomized order on an Acquity CSH C18 column held at 50 °C (2.1 mm × 100 mm × 1.7 µm particle diameter; Waters) using an Ultimate 3000 RSLC binary pump (400 µl min^{-1} flow rate; Thermo Fisher) or a Vanquish binary pump for validation experiments. Mobile phase A consisted of 10 mM ammonium acetate in acetonitrile/H_2O (70:30 v/v) containing 250 µl l^{-1} acetic acid. Mobile phase B consisted of 10 mM ammonium acetate in isopropanol/acetonitrile (90:10 v/v) with the same additives. Mobile phase B was initially held at 2% for two min and then increased to 30% over three min; further increased to 50% over one min and 85% over 14 min; and then raised to 95% over one min and held for seven min. The column was re-equilibrated for two min before the next injection.

DO lipid extracts (20 µl) were injected by an Ultimate 3000 RSLC autosampler (Thermo Fisher) coupled to a Q Exactive Focus mass spectrometer by a HESI II heated electrospray ionization (ESI) source. Both source and inlet capillary were kept at 300 °C. Sheath gas was set

to 25 units, auxiliary gas to ten units and the spray voltage was set to 5,000 V (+) and 4,000 V (−), respectively. The MS was operated in polarity switching mode, acquiring positive and negative mode MS1 and MS2 spectra (Top2) during the same separation. MS acquisition parameters were 17,500 resolving power, 1×10^6 automatic gain control (AGC) target for MS1 and 1×10^5 AGC target for MS2 scans, 100 ms MS1 and 50 ms MS2 ion accumulation time, 200- to 1,600 Th MS1 and 200- to 2,000 Th MS2 scan range, 1 Th isolation width for fragmentation, stepped HCD collision energy (20, 30, 40 units), 1.0% under fill ratio and ten s dynamic exclusion.

QTL mapping

Genetic QTL mapping was performed using the R/qtl2 (v0.24) package (Broman et al., 2019) which fit a linear mixed effect model that included accounting for overall genetic relationship with a random effect, that is, kinship effect. The leave one chromosome out (LOCO) method was used, which accounts for population structure without reducing QTL mapping power. For each gut microbiome trait and caecal lipidome traits, sex, days on diet and mouse cohort (wave) were used as additive covariates as described previously (Kemis et al., 2019). For gut microbiome traits and caecal lipidome traits, normalized abundance/coverage was transformed to normal quantiles. The mapping statistic reported was the $\log_{10}$ likelihood ratio (LOD score). The QTL support interval was defined using the 95% Bayesian confidence interval (Broman et al., 2019). Significance thresholds for QTL were determined by permutation analysis ($n = 1{,}000$). We included 2,803 gut microbiome function traits, 197 gut microbiome taxon traits and 3,384 caecal lipid feature traits for our QTL mapping. The reported genome-wide *P* values were not adjusted for the multiple phenotypes to avoid overly declaring QTL in the initial analysis. We used genome-wide $P < 0.05$ for significant QTL and used genome-wide $P < 0.2$ to find concordant QTL mapping and hotspots.

Mediation analysis

Mediation analysis was carried out as previously described (Chick et al., 2016). Mediation analysis was used to relate individual gut microbial metagenes and lipid features by scanning all 136,200 identified metagenes with at least ten CPM in 20% of the samples to all 3,963 caecal lipid features. We used the subset of animals for which both gut metagenomic and caecal lipid data were available ($n = 221$). We first defined gut microbial traits with suggestive QTL as the outcome variable; we then included candidate caecal lipid mediators as additive covariates in the suggestive mbQTL mapping model and re-ran the QTL analysis. We performed the same analysis with caecal lipid features as the outcome and gut microbial features as candidate mediators. A mediatory role was supported by a significant decrease in LOD score from the original outcome QTL. Significance of the LOD score drop for a given candidate gut microbial metagene mediator on a given caecal lipid QTL was estimated by *z*-score scaled by LOD score drop, and a conservative $z\text{-score} \leq -6$ was recorded as a potential causal mediator. The mean of fitted distributions for a given gut bacterial taxon, for example all metagenes from *A. muciniphila* gut, was scaled to the corresponding *z*-score to evaluate the mediation significance for this gut bacterial taxon.

Bacterial culturing and bacterial extracellular vesicle isolation

A. muciniphila was grown anaerobically in defined medium (Supplementary Table 15). To test for the effects of phosphate condition, the concentration of phosphate in the medium was adjusted to 0.02, 0.2 or 2 mM. *E. coli* MS200-1 strain was grown in LC medium (10 g l^{-1} bacto-tryptone, 5 g l^{-1} bacto-yeast extract, 5 g l^{-1} NaCl). *B. thetaiotaomicron* strain VPI-5482 was grown in CMM medium. All bacterial strains were grown at 37 °C. Cells for lipid analyses from the three strains were obtained by centrifugation. Isolation of *A. muciniphila* extracellular vesicles used a previously described method (Ashrafian et al., 2019).

Human faecal samples

Stool samples were obtained from a previous study (Dill-McFarland et al., 2019). Samples were collected from participants of the Wisconsin Longitudinal Study. Briefly, participants collected stool samples directly into sterile containers, then samples were kept at ~4 °C until arrival (48 h or less) at the processing laboratory. Upon arrival, sterile straws were filled with the faecal material and stored at −80 °C as previously described (Dill-McFarland et al., 2019). 16S rRNA gene sequencing data for these samples were previously published. The use of the Wisconsin Longitudinal Study faecal samples was approved by the Institutional Review Board at UW-Madison. Consent from participants was obtained via a process involving both verbal and written components by trained interviewers, and records were archived both digitally and physically at UW-Madison. This effort did not include collection of samples from vulnerable populations or from minors.

Sample preparation for OL validation experiments

For caecal contents, 30 ± 6 mg caecal contents were aliquoted into a tube with a metal bead and 280 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 eight min at 25 Hz. For bacterial cultures, ~75 µl of bacterial culture were aliquoted into a tube and 280 µl MeOH were added for protein precipitation. After the mixture was vortexed for 10 s, 900 µl MTBE were added as extraction solvent and the mixture was vortexed for ten s and mixed on an orbital shaker for six min. Phase separation was induced by adding 225 µl of water followed by 20 s of vortexing. All steps were performed at 4 °C on ice. Finally, each mixture was centrifuged for eight min at 11,000 × *g* at 4 °C, after which 240 µl of the lipophilic upper layer were transferred to glass vials and dried in a vacuum centrifuge for 60 min. The dried lipophilic extracts were re-suspended in 200 µl MeOH:toluene (9:1 v/v).

LC–MS/MS analysis of OL validation experiments

Sample analysis by LC–MS/MS was performed in randomized order on an Acquity CSH C18 column held at 50 °C (2.1 mm × 100 mm × 1.7 µm particle diameter; Waters) using an Ultimate 3000 RSLC binary pump (400 µl min^{-1} flow rate; Thermo Fisher) or a Vanquish binary pump. The same mobile phase and gradient as for the DO samples were used.

For the validation experiments, 10 µl of caecal or culture extract were injected by a Vanquish Split Sampler HT autosampler (Thermo Fisher) coupled to a Q Exactive HF mass spectrometer by a HESI II heated ESI source. Both source and inlet capillary were kept at 350 °C (Thermo Fisher). Sheath gas was set to 25 units, auxiliary gas to 15 units and spare gas to five units, while the spray voltage was set to 3,500 V and the S-lens RF level to 90. The MS was operated in polarity switching dd-MS2 mode (Top2), acquiring positive and negative mode MS1 and MS2 spectra during the same separation. MS acquisition parameters were 30,000 resolution, 1×10^6 AGC target for MS1 and 5×10^5 AGC target for MS2 scans, 100 ms MS1 and 50 ms MS2 ion accumulation time, 200 to 2,000 Th MS1 scan range, 1.0 Th isolation width for fragmentation and stepped HCD collision energy (20, 30, 40 units).

Lipidomic analysis

All resulting LC–MS lipidomics raw files were converted to mgf files via MSConvertGUI (ProteoWizard, Dr Parag Mallick, Stanford University) and processed using LipiDex (Hutchins et al., 2018) and Compound Discoverer 2.0 or 2.1.0.398 (Thermo Fisher) for DO and validation experiments, respectively. All raw files were loaded into Compound Discoverer with blanks marked as such to generate two result files using the following workflow processing nodes: Input Files, Select Spectra, Align Retention Times, Detect Unknown Compounds, Group Unknown Compounds, Fill Gaps and Mark Background Compounds for the so called ‘Aligned’ result and solely Input Files, Select Spectra and Detect Unknown Compounds for an ‘Unaligned’ Result. Under Select Spectra, the retention time limits were set between 0.4 and 21 min, MS order as well as unrecognized MS order replacements were set to MS1. Further replacements were set to

FTMS Mass Analyzer and HCD Activation Type. Under Align Retention Times, the mass tolerance was set to ten ppm and the maximum shift according to the data set to 0.6 min for the DO and 0.5 min for the validation experiments. Under Detect Unknown Compounds, the mass tolerance was also set to ten ppm, with an S/N threshold of five (DO) or three (validation), and a minimum peak intensity of 5×10^6 (DO) or 1×10^5 (validation).

For the DO samples, [M+H]+1 and [M−H]−1 were selected as ions and a maximum peak width of 0.75 min as well as a minimum number of scans per peak equalling seven were set. For the validation samples, [M+H]+1 and [M−H+TFA]−1 were selected as ions and a maximum peak width of 0.75 min as well as a minimum number of scans per peak equalling five were set. Lastly, for Group Unknown Compounds as well as Fill Gaps, mass tolerance was set to ten ppm and retention time tolerance to 0.2 min. For best compound selection, rules #1 and #2 were set to unspecified, while MS1 was selected for preferred MS order and [M+H]+1 as the preferred ion. For everything else, the default settings were used. Resulting peak tables were exported as Excel files in three levels of Compounds, Compound per File and Features (just Features for the 'Unaligned') and later saved as csv. In LipiDex' Spectrum Searcher 'LipiDex_HCD_Acetate', 'LipiDex_HCD_Plants', 'LipiDex_Splash_ISTD_Acetate', LipiDex_HCD_ULCFA' and 'Ganglioside_20171205' were selected as libraries for the DO, and 'Coon_Lab_HCD_Acetate_20171229', 'Ganglioside_20171205' and 'Ornithine-Lipids_20180404' for the validation experiments. For all searches, the defaults of 0.01 Th for MS1 and MS2 search tolerances, a maximum of one returned search result and an MS2 low mass cut-off of 61 Th were kept. Under the Peak Finder tab, Compound Discoverer was chosen as peak table type, and its 'Aligned' and 'Unaligned' results, as well as the MS/MS results from Spectrum Researcher were uploaded. Features had to be identified in a minimum of one file while keeping the defaults of a minimum of 75% of lipid spectral purity, an MS2 search dot product of at least 500 and reverse dot product of at least 700, as well as a multiplier of 2.0 for FWHM window, a maximum of 15 ppm mass difference, adduct/dimer and in-source fragment (and adduct and dimer) filtering and a

maximum RT M.A.D Factor of 3.5. As post-processing in the DO, all features that were only found in one file and had no ID were deleted, and duplicates were also deleted. Peak areas of the three targeted ornithine lipid species were obtained via TraceFinder v3.3.350.0 (Thermo Fisher). Details of the lipid classes searched for in these databases with their respective adducts are shown in Supplementary Table 15. Lipids ID matching was performed at <±5 ppm between runs.

OL synthesis

Chemicals and methods

All chemicals were obtained from Chem-Impex, Sigma-Aldrich, Agros Organics or TCI America. All reagents and solvents were used without further purification except for hexane, ethyl acetate and dichloromethane, which were distilled before use. Analytical thin-layer chromatography (TLC) was performed on 250 µm glass-backed silica plates with F-254 fluorescent indicator from Silicycle. Visualization was performed using UV light and iodine.

General instrumentation information

Nuclear magnetic resonance (NMR) spectra were recorded in deuterated solvents at 400 MHz on a Bruker-Avance spectrometer equipped with a BFO probe, and at 500 MHz on a Bruker-Avance spectrometer equipped with a DCH cryoprobe. Chemical shifts are reported in parts per million using residual solvent peaks or tetramethylsilane (TMS) as a reference. Couplings are reported in hertz (Hz). ESI–exact mass measurement (ESI–EMM) mass spectrometry data were collected on a Waters LCT instrument.

OL synthesis

Tridecanoic acid (compound 1, 3.2 g, 15 mmol) was dissolved in dichloromethane (150 ml, 0.1 M) in a round-bottom flask equipped with a stir bar. 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochloride (EDC-HCl) (4.3 g, 22.5 mmol), 4-dimethylaminopyridine (DMAP)

(273 mg, 2.25 mmol) and Meldrum's acid (3.2 g, 22.5 mmol) were added to the flask, and the reaction was stirred overnight at room temperature. The next day, the reaction mixture was washed with 1 M HCl (3 × 75 ml), saturated $NaHCO_3$ (3 × 75 ml) and brine (3 × 75 ml). The mixture was then dried over magnesium sulfate and concentrated under reduced pressure. The resultant oil was then dissolved in benzene (19 ml) in a round-bottom flask with a stir bar, and benzyl alcohol (45 mmol, 4.7 ml) was added. The reaction was heated to 95 °C for three hours and then concentrated under reduced pressure. The crude reaction mixture was purified by silica gel flash chromatography (5–10% ethyl acetate in hexane as eluent), yielding 3.6 g of compound 2 as an oil (69% yield over two steps).

Compound 2 (3.6 g, 10.4 mmol) was added to a round-bottom flask equipped with a stir bar and dissolved in a 2:1 mixture of tetrahydrofuran (16 ml) and ethanol (8 ml). The round-bottom flask was cooled in an ice bath, and sodium cyanoborohydride (1.6 g, 26 mmol) was added to the mixture. One M aqueous HCl (26 ml, 26 mmol) was added via addition funnel, and the reaction was allowed to stir to room temperature and monitored by TLC. Upon consumption of starting material, the aqueous portion of the reaction was extracted with dichloromethane (3 × 20 ml) and combined with the organic portion. The combined organic portions were washed with brine (3 × 20 ml), dried over $MgSO_4$ and concentrated under reduced pressure to yield 3.26 g of compound 3 (93% crude). The material was used without further purification.

Pentadecanoic acid (1.93 g, 9 mmol) was added to a round-bottom flask equipped with a stir bar and dissolved in dichloromethane (80 ml). To the flask was added EDC-HCl (2.68 g, 14 mmol), DMAP (974 mg, 8 mmol) and compound 3 (2.78 g, 8 mmol). The reaction mixture was allowed to stir overnight at room temperature. The next day, the mixture was washed with 1 M HCl (3 × 50 ml), saturated $NaHCO_3$ (3 × 50 ml) and saturated brine (3 × 50 ml). The mixture was then dried over magnesium sulfate and concentrated under reduced pressure. The crude material was purified by silica gel flash chromatography (5–10% ethyl acetate in hexane as eluent), yielding 4.3 g of compound 4 (94% isolated yield).

To a flame-dried round-bottom flask equipped with a stir bar was added Pd/C (798 mg, 0.75 mmol Pd). Dry dichloromethane was added to the flask to make a slurry, and the atmosphere was exchanged for nitrogen. Compound 4 (4.3 g, 7.5 mmol) was dissolved in anhydrous methanol and added to the reaction vessel. The atmosphere was then exchanged for hydrogen (balloon pressure), and the reaction was allowed to proceed overnight. The next day, the reaction was diluted with ethyl acetate and filtered over celite. The mixture was concentrated under reduced pressure to yield compound 5 as a white solid (3.5 g, 97% crude yield). The material was used without further purification.

Compound 5 (256 mg 0.5 mmol) was added to a round-bottom flask equipped with a stir bar and dissolved in dimethylformamide (DMF) (5 ml). To the flask was added N,N-Diisopropylethylamine (DIPEA) (277 µl, 1.6 mmol) and hexafluorophosphate azabenzotriazole tetramethyl uronium (HATU) (216 mg, 5.5 mmol), and the mixture was stirred for 15 min. Protected ornithine (250 mg, 0.6 mmol) was added to the mixture, which was stirred at room temperature and monitored by TLC. When starting material was no longer observed by TLC, the mixture was diluted in diethyl ether (20 ml) and washed with 1 M HCl (3 × 20 ml), saturated $NaHCO_3$ (3 × 20 ml) and brine (3 × 20 ml). The mixture was dried over magnesium sulfate and concentrated under reduced pressure to yield a white solid (376 mg crude). This sample was combined with an additional sample of the same crude material that appeared identical by ^1H NMR analysis and was then purified by silica gel flash chromatography (25% ethyl acetate in hexane as eluent) to yield 131 mg of compound 6.

To a flame-dried round-bottom flask equipped with a stir bar was added Pd/Cn (17.0 mg, 0.16 mmol Pd). Dry dichloromethane was added to the flask to make a slurry, and the atmosphere was exchanged for nitrogen. The protected ornithine lipid (compound 6, 131 mg, 0.160 mmol) was dissolved in a mixture of 4 ml anhydrous methanol/dichloromethane (DCM) (1:1) and added to the reaction vessel. The atmosphere was then exchanged for hydrogen (balloon pressure), and the reaction was allowed to proceed overnight. The next day, the reaction was filtered over celite.

The mixture was concentrated under reduced pressure to yield OL as an off-white solid (82.2 mg, 86% crude yield). Deprotected OL was identified using LC and ESI-EMM ([M]+ calculated 597.5207, measured 597.5188, 0.002 ppm) in the resultant mixture and the material was used without further purification in the experiments described herein.

RNA-seq and eQTL analysis

Samples of flash-frozen distal ileum from DO mice were homogenized with Qiagen Tissuelyser (two step two min at 25 Hz, with flipping plate homogenization with five min ice incubation). Total RNA was extracted from homogenized samples using Qiagen 96 universal kit (Qiagen). RNA clean-up was performed using Qiagen RNeasy mini kit (Qiagen). DNA was removed by on-column DNase digestion (Qiagen). Purified RNA was quantified using a Nanodrop 2000 spectrophotometer and RNA fragment analyzer (Agilent). Library preparation was performed using the TruSeq Stranded mRNA sample preparation guide (Illumina). IDT unique dual indexes (UDIs), Illumina UDIs or NEXTflex UDIs were used as barcodes for each library sample. RNA sequencing was performed on an Illumina NovaSeq 6000 platform. Raw RNA-seq reads quality control was performed using Trimmomatic (v0.39) (Bolger et al., 2014) with default parameters. Genotype-free genome reconstruction and allele specific expression quantification were performed using the GBRS tool (http://churchill-lab.github.io/gbrs/). Genes with at least ten transcripts per million in at least 10% of DO mice were used for downstream analyses. For eQTL mapping, sex, RNA-seq index, RNA-seq wave and mouse cohort (wave) were used as additive covariates. eQTL analysis was otherwise the same as previously described (Keller et al., 2018).

BMDM assay and cell viability measurement

Bone marrow was isolated from femur and tibia from ~six-week-old B6 and 129 mice fed with chow diet. Bone marrow cells were re-suspended into single-cell suspensions and cultured in complete DMEM medium supplemented with 10% fetal calf serum (FCS), 2 mM l-glutamine, 1%

penicillin/streptomycin and 20 ng ml^{-1} mouse macrophage colony stimulating factor (M-CSF) (BioLegend) for the purpose of differentiation. BMDM cells were randomly allocated into treatment groups. BMDMs were collected at day seven and treated with LPS, OL or LPS + OL for 6 hours in media supplemented with 1% fetal bovine serum (FBS), then supernatants were collected for measurement of cytokines. For optimization, cytokine (TNF-α and IL-6) production from LPS- or OL-treated BMDM was performed using mouse TNF-α ELISA MAX Deluxe kit and mouse IL-6 ELISA MAX Deluxe kit (BioLegend), respectively. Follow-up cytokine (IL-1β, IL-6, IL-10, IL-12, MCP-1, TNF-α, MIP-1α, GM-CSF and RANTES) production assays in response to LPS + OL co-cultured BMDM were performed using Q-Plex Mouse Cytokine Screen 16-Plex (Quansys). Cell viability was determined by flow cytometry (Thermo Fisher Attune NxT) after staining with 7-amino-actinomycin D (eBioscience).

RNA-seq of BMDM

Frozen BMDM were homogenized with Qiagen Tissuelyser (two min at 20 Hz) and total RNA was extracted using Qiagen 96 universal kit (Qiagen). RNA clean-up was performed using Qiagen RNeasy mini kit (Qiagen). DNA was removed by on-column DNase digestion (Qiagen). Library preparation was performed using the TruSeq Stranded mRNA sample preparation guide (Illumina). RNA sequencing was performed on an Illumina NovaSeq 6000 platform. Raw RNA-seq reads quality control was performed using Trimmomatic v0.39) (Bolger et al., 2014) with default parameters. Gene quantification was performed using RSEM (v1.3.1) (B. Li & Dewey, 2011). DESeq2 (v1.26.0) (Love et al., 2014) was used to identify differentially expressed genes between groups.

***Akkermansia*-specific qPCR for mouse faecal samples**

To quantify *Akkermansia* abundance in mouse faecal samples, previously validated primers specific for *A. muciniphila* were used (forward CAGCACGTGAAGGTGGGGAC and

reverse CTTGCGGTTGGCTTCAGAT) (Collado et al., 2007). *A. muciniphila* genomic DNA isolated from a pure culture was used to generate a standard curve encompasing seven points (range: 1 ng μl^{-1}–0.015625 ng μl^{-1}). The PCR reaction contained SsoAdvanced Universal SYBR Green Supermix (Bio-Rad). Faecal *A. muciniphila* abundance was normalized by faecal weight.

Data analysis and statistical analysis

All data integration and statistical analysis were performed in R (v3.6.3). Data collection and analysis were not performed blind to the conditions of the experiments. No data were excluded from the analysis. No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those reported in previous publications (Kemis et al., 2019). Differences between groups were evaluated using unpaired two-tailed Welch's *t*-test. Enrichment analysis was performed with Fisher's exact test using a custom R function. Correlation analysis was performed with two-sided Spearman's correlation using the R function 'cor.test()'. For multiple testing, Benjamini-Hochberg false discovery rate (FDR) procedure was used to adjust *P* values. Data integration was performed using R packages dplyr (v1.0.6), tidyr (v1.1.3), reshape2 (v1.4.4) and data.table (v1.14.0). Heat maps were plotted using the R package pheatmap (v1.0.12). Other plots were created using the R packages ggplot2 (v3.3.3), gridExtra (v2.3), RcolorBrewer (v1.1-2) and ggsci (v2.9).

Data availability

DO metagenomic WGS data are available from the Sequence Read Archive (SRA) under accession PRJNA744213. RNA-seq data are available from the Sequence Read Archive (SRA) under accession numbers PRJNA772743 and PRJNA896574. Mass spectrometry data files are available on Chorus (chorusproject.org) under accession with project ID 1681 (direct links to DO caecum lipidomics: https://chorusproject.org/anonymous/download/experiment/10cb106716da44cd924a

3c73ac30083d and founder strains caecum lipidomics: https://chorusproject.org/anonymous/download/experiment/ad7566e8f45942d2ba0f579857629b55). Genotypes data and additional phenotype data associated with DO mouse are available at Dryad (https://doi.org/10.5061/dryad.pj105). SNP associations data cc_variants.sqlite are available at https://ndownloader.figshare.com/files/18533342 and mouse genes data mouse_genes_mgi.sqlite used for QTL mapping are available at https://ndownloader.figshare.com/files/17609252. Source data are provided with this paper.

Code availability

All code used in this study is available in GitHub (https://github.com/qijunz/Zhang_DO_paper) or in the corresponding software package websites.

2.6 Contributions

F.E.R., M.P.K. and A.D.A. conceived the study. Q.Z., V.L., L.L.T., A.D.A., J.J.C. and F.E.R. designed experiments. K.L.S., D.S.S. and M.E.R. assisted with mouse sample collection. L.L.T. and J.H.K. contributed to sample processing for DNA sequencing. Q.Z., L.L.T. and K.W.B. performed metagenomic and QTL analysis. V.L., K.A.O., E.A.T., T.R.R. and J.D.R. collected lipidomic data. V.L., I.J.M., M.P.K., D.M.G., G.R.K., D.T.P. and G.A.C. analysed lipidome and lipidome QTL data. D.E.M., T.J.P., J.P.G. and H.E.B. synthetized OL. R.L.K. performed bacterial culture experiments. Q.Z. and K.K. performed cell culture studies. E.I.V. assisted with gnotobiotic mouse experiments. Q.Z., M.S. and A.J.L. assisted with intestine RNA-seq. Q.Z., V.L. and F.E.R. wrote the manuscript. All authors approved the final manuscript.

2.7 Acknowledgements

We thank the University of Wisconsin Biotechnology Center DNA Sequencing Facility for providing sequencing and support services; the University of Wisconsin Center for High Throughput Computing (CHTC) in the Department of Computer Sciences for providing computational resources, support and assistance; K. Anantharaman from University of Wisconsin-Madison for providing computational resources and support and the University of Wisconsin Carbone Cancer Center Flow Lab for support services. This work was supported by National Institutes of Health (NIH) grants DK108259 (F.E.R.), HL144651 (F.E.R. and A.J.L.), HL148577 (F.E.R. and A.J.L.), DK101573 (A.D.A.), GM131817 (H.E.B.), GM070683 (K.W.B. and G.A.C.); NIH National Center for Quantitative Biology of Complex Systems grant P41108538 (J.J.C.); NIH National Institute of Allergy and Infectious Diseases grant T32AI55397 (J.H.K.); NLM Computation and Informatics in Biology and Medicine Postdoctoral Fellowship 5T15LM007359 (L.L.T.) and T32DK007665 (L.L.T.) and NIH Chemistry-Biology Interface Training Grant T32 GM008505 (T.J.P.). This work was also supported by Fondation Leducq 17CVD01 (F.E.R.). V.L. was supported by the Foundation for Polish Science (MAB/2017/2 and START 064.2022), the European Molecular Biology Organization EMBO (Postdoctoral Fellowship ALTF 474-2021) and the National Science Centre, Poland (SONATINA 5 2021/40/C/NZ3/00283). The ‘Regenerative Mechanisms for Health - ReMedy’ project (MAB/2017/2) is carried out within the International Research Agendas Programme of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund.

2.8 References

Ashrafian, F., Shahriary, A., Behrouzi, A., Moradi, H. R., Keshavarz Azizi Raftar, S., Lari, A., Hadifar, S., Yaghoubfar, R., Ahmadi Badi, S., Khatami, S., Vaziri, F., & Siadat, S. D. (2019). Akkermansia muciniphila-Derived Extracellular Vesicles as a Mucosal Delivery Vector for Amelioration of Obesity in Mice. *Frontiers in Microbiology*, *10*, 2155. https://doi.org/10.3389/fmicb.2019.02155

Baxter, A. A., Hulett, M. D., & Poon, I. K. (2015). The phospholipid code: A key component of dying cell recognition, tumor progression and host–microbe interactions. *Cell Death & Differentiation*, *22*(12), 1893–1905. https://doi.org/10.1038/cdd.2015.122

Benson, A. K., Kelly, S. A., Legge, R., Ma, F., Low, S. J., Kim, J., Zhang, M., Oh, P. L., Nehrenberg, D., Hua, K., Kachman, S. D., Moriyama, E. N., Walter, J., Peterson, D. A., & Pomp, D. (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proceedings of the National Academy of Sciences*, *107*(44), 18933–18938. https://doi.org/10.1073/pnas.1007028107

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bonder, M. J., Kurilshikov, A., Tigchelaar, E. F., Mujagic, Z., Imhann, F., Vila, A. V., Deelen, P., Vatanen, T., Schirmer, M., Smeekens, S. P., Zhernakova, D. V., Jankipersadsing, S. A., Jaeger, M., Oosting, M., Cenit, M. C., Masclee, A. A. M., Swertz, M. A., Li, Y., Kumar, V., … Zhernakova, A. (2016). The effect of host genetics on the gut microbiome. *Nature Genetics*, *48*(11), 1407–1412. https://doi.org/10.1038/ng.3663

Broman, K. W., Gatti, D. M., Simecek, P., Furlotte, N. A., Prins, P., Sen, Ś., Yandell, B. S., & Churchill, G. A. (2019). R/qtl2: Software for Mapping Quantitative Trait Loci with High-

Dimensional Data and Multiparent Populations. *Genetics*, *211*(2), 495–502. https://doi.org/10.1534/genetics.118.301595

Brown, E. M., Ke, X., Hitchcock, D., Jeanfavre, S., Avila-Pacheco, J., Nakata, T., Arthur, T. D., Fornelos, N., Heim, C., Franzosa, E. A., Watson, N., Huttenhower, C., Haiser, H. J., Dillow, G., Graham, D. B., Finlay, B. B., Kostic, A. D., Porter, J. A., Vlamakis, H., … Xavier, R. J. (2019). Bacteroides-Derived Sphingolipids Are Critical for Maintaining Intestinal Homeostasis and Symbiosis. *Cell Host & Microbe*, *25*(5), 668-680.e7. https://doi.org/10.1016/j.chom.2019.04.002

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

Cao, Y., Wang, X., Yang, Q., Deng, H., Liu, Y., Zhou, P., Xu, H., Chen, D., Feng, D., Zhang, H., Wang, H., & Zhou, J. (2020). Critical Role of Intestinal Microbiota in ATF3-Mediated Gut Immune Homeostasis. *The Journal of Immunology*, *205*(3), 842–852. https://doi.org/10.4049/jimmunol.1901000

Chick, J. M., Munger, S. C., Simecek, P., Huttlin, E. L., Choi, K., Gatti, D. M., Raghupathy, N., Svenson, K. L., Churchill, G. A., & Gygi, S. P. (2016). Defining the consequences of genetic variation on a proteome-wide scale. *Nature*, *534*(7608), 500–505. https://doi.org/10.1038/nature18270

Churchill, G. A., Gatti, D. M., Munger, S. C., & Svenson, K. L. (2012). The diversity outbred mouse population. *Mammalian Genome*, *23*(9–10), 713–718. https://doi.org/10.1007/s00335-012-9414-2

Collado, M. C., Derrien, M., Isolauri, E., de Vos, W. M., & Salminen, S. (2007). Intestinal Integrity and Akkermansia muciniphila , a Mucin-Degrading Member of the Intestinal Microbiota Present in Infants, Adults, and the Elderly. *Applied and Environmental Microbiology*, *73*(23), 7767–7770. https://doi.org/10.1128/AEM.01477-07

de Carvalho, C., & Caramujo, M. (2018). The Various Roles of Fatty Acids. *Molecules*, *23*(10), 2583. https://doi.org/10.3390/molecules23102583

Dees, C., & Shively, J. M. (1982). Localization of quantitation of the ornithine lipid of Thiobacillus thiooxidans. *Journal of Bacteriology*, *149*(2), 798–799. https://doi.org/10.1128/jb.149.2.798-799.1982

Dennis, E. A., & Norris, P. C. (2015). Eicosanoid storm in infection and inflammation. *Nature Reviews Immunology*, *15*(8), 511–523. https://doi.org/10.1038/nri3859

Depommier, C. (2019). Supplementation with Akkermansia muciniphila in overweight and obese human volunteers: A proof-of-concept exploratory study. *Nature Medicine*, *25*, 16.

Diercks, H., Semeniuk, A., Gisch, N., Moll, H., Duda, K. A., & Hölzl, G. (2015). Accumulation of Novel Glycolipids and Ornithine Lipids in Mesorhizobium loti under Phosphate Deprivation. *Journal of Bacteriology*, *197*(3), 497–509. https://doi.org/10.1128/JB.02004-14

Dill-McFarland, K. A., Tang, Z.-Z., Kemis, J. H., Kerby, R. L., Chen, G., Palloni, A., Sorenson, T., Rey, F. E., & Herd, P. (2019). Close social relationships correlate with human gut microbiota composition. *Scientific Reports*, *9*(1), 703. https://doi.org/10.1038/s41598-018-37298-9

Du, Y., Ma, Z., Zheng, J., Huang, S., Yang, X., Song, Y., Dong, D., Shi, L., & Xu, D. (2022). ATF3 Positively Regulates Antibacterial Immunity by Modulating Macrophage Killing and Migration Functions. *Frontiers in Immunology*, *13*, 839502. https://doi.org/10.3389/fimmu.2022.839502

Everard, A., Belzer, C., Geurts, L., Ouwerkerk, J. P., Druart, C., Bindels, L. B., Guiot, Y., Derrien, M., Muccioli, G. G., Delzenne, N. M., de Vos, W. M., & Cani, P. D. (2013). Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. *Proceedings of the National Academy of Sciences*, *110*(22), 9066–9071. https://doi.org/10.1073/pnas.1219451110

Faith, J. J., McNulty, N. P., Rey, F. E., & Gordon, J. I. (2011). Predicting a Human Gut Microbiota's Response to Diet in Gnotobiotic Mice. *Science*, *333*(6038), 101–104. https://doi.org/10.1126/science.1206025

Gaudet, R. G., Sintsova, A., Buckwalter, C. M., Leung, N., Cochrane, A., Li, J., Cox, A. D., Moffat, J., & Gray-Owen, S. D. (2015). Cytosolic detection of the bacterial metabolite HBP activates TIFA-dependent innate immunity. *Science*, *348*(6240), 1251–1255. https://doi.org/10.1126/science.aaa4921

Geiger, O., González-Silva, N., López-Lara, I. M., & Sohlenkamp, C. (2010). Amino acid-containing membrane lipids in bacteria. *Progress in Lipid Research*, *49*(1), 46–60. https://doi.org/10.1016/j.plipres.2009.08.002

GEM Project Research Consortium, Turpin, W., Espin-Garcia, O., Xu, W., Silverberg, M. S., Kevans, D., Smith, M. I., Guttman, D. S., Griffiths, A., Panaccione, R., Otley, A., Xu, L., Shestopaloff, K., Moreno-Hagelsieb, G., Paterson, A. D., & Croitoru, K. (2016). Association of host genome with intestinal microbial composition in a large healthy cohort. *Nature Genetics*, *48*(11), 1413–1417. https://doi.org/10.1038/ng.3693

Gilchrist, M., Thorsson, V., Li, B., Rust, A. G., Korb, M., Kennedy, K., Hai, T., Bolouri, H., & Aderem, A. (2006). Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4. *Nature*, *441*(7090), 173–178. https://doi.org/10.1038/nature04768

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W. J. H., Jansen, R., de Geus, E. J. C., Boomsma, D. I., Wright, F. A., Sullivan, P. F., Nikkola, E., Alvarez, M., Civelek, M., Lusis, A. J., Lehtimäki, T., Raitoharju, E., Kähönen, M., Seppälä, I., … Pasaniuc, B. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics*, *48*(3), 245–252. https://doi.org/10.1038/ng.3506

Hughes, D. A., Bacigalupe, R., Wang, J., Rühlemann, M. C., Tito, R. Y., Falony, G., Joossens, M., Vieira-Silva, S., Henckaerts, L., Rymenans, L., Verspecht, C., Ring, S., Franke, A., Wade, K. H., Timpson, N. J., & Raes, J. (2020). Genome-wide associations of human

gut microbiome variation and implications for causal inference analyses. *Nature Microbiology*, *5*(9), 1079–1087. https://doi.org/10.1038/s41564-020-0743-8

Hutchins, P. D., Russell, J. D., & Coon, J. J. (2018). LipiDex: An Integrated Software Package for High-Confidence Lipid Identification. *Cell Systems*, *6*(5), 621-625.e5. https://doi.org/10.1016/j.cels.2018.03.011

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*(1), 119. https://doi.org/10.1186/1471-2105-11-119

Jain, M., Ngoy, S., Sheth, S. A., Swanson, R. A., Rhee, E. P., Liao, R., Clish, C. B., Mootha, V. K., & Nilsson, R. (2014). A systematic survey of lipids across mouse tissues. *American Journal of Physiology-Endocrinology and Metabolism*, *306*(8), E854–E868. https://doi.org/10.1152/ajpendo.00371.2013

Kahles, F., Meyer, C., Mollmann, J., Diebold, S., Findeisen, H. M., Lebherz, C., Trautwein, C., Koch, A., Tacke, F., Marx, N., & Lehrke, M. (2014). GLP-1 Secretion Is Increased by Inflammatory Stimuli in an IL-6-Dependent Manner, Leading to Hyperinsulinemia and Blood Glucose Lowering. *Diabetes*, *63*(10), 3221–3229. https://doi.org/10.2337/db14-0100

Kawai, Y., & Akagawa, K. (1989). Macrophage Activation by an Ornithine-Containing Lipid or a Serine-Containing Lipid. *INFECT. IMMUN.*, *57*(7), 2086–2091.

Kawai, Y., Kaneda, K., Morisawa, Y., & Akagawa, K. (1991). Protection of Mice from Lethal Endotoxemia by Use of an Ornithine-Containing Lipid or a Serine-Containing Lipid. *INFECT. IMMUN.*, *59*(8), 2560–2566.

Kawai, Y., Yano, I., & Kaneda, K. (1988). Various kinds of lipoamino acids including a novel serine-containing lipid in an opportunistic pathogen Flavobacterium. Their structures and biological activities on erythrocytes. *European Journal of Biochemistry*, *171*(1–2), 73–80. https://doi.org/10.1111/j.1432-1033.1988.tb13760.x

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., Kendziorski, C., Broman, K. W., Yandell, B. S., Churchill, G. A., & Attie, A. D. (2018). Genetic Drivers of Pancreatic Islet Function. *Genetics*, *209*(1), 335–356. https://doi.org/10.1534/genetics.118.300864

Keller, M. P., Rabaglia, M. E., Schueler, K. L., Stapleton, D. S., Gatti, D. M., Vincent, M., Mitok, K. A., Wang, Z., Ishimura, T., Simonett, S. P., Emfinger, C. H., Das, R., Beck, T., Kendziorski, C., Broman, K. W., Yandell, B. S., Churchill, G. A., & Attie, A. D. (2019). Gene loci associated with insulin secretion in islets from non-diabetic mice. *Journal of Clinical Investigation*, *129*(10), 4419–4432. https://doi.org/10.1172/JCI129143

Kemis, J. H., Linke, V., Barrett, K. L., Boehm, F. J., Traeger, L. L., Keller, M. P., Rabaglia, M. E., Schueler, K. L., Stapleton, D. S., Gatti, D. M., Churchill, G. A., Amador-Noguez, D., Russell, J. D., Yandell, B. S., Broman, K. W., Coon, J. J., Attie, A. D., & Rey, F. E. (2019). Genetic determinants of gut microbiota composition and bile acid profiles in mice. *PLOS Genetics*, *15*(8), e1008073. https://doi.org/10.1371/journal.pgen.1008073

Khuu, C. H., Barrozo, R. M., Hai, T., & Weinstein, S. L. (2007). Activating transcription factor 3 (ATF3) represses the expression of CCL4 in murine macrophages. *Molecular Immunology*, *44*(7), 1598–1605. https://doi.org/10.1016/j.molimm.2006.08.006

Kim, S.-K., Park, S.-J., Li, X.-H., Choi, Y.-S., Im, D.-S., & Lee, J.-H. (2018). Bacterial ornithine lipid, a surrogate membrane lipid under phosphate-limiting conditions, plays important roles in bacterial persistence and interaction with host: *Role of ornithine lipid in chronic adaptation*. *Environmental Microbiology*, *20*(11), 3992–4008. https://doi.org/10.1111/1462-2920.14430

Kindt, A., Liebisch, G., Clavel, T., Haller, D., Hörmannsperger, G., Yoon, H., Kolmeder, D., Sigruener, A., Krautbauer, S., Seeliger, C., Ganzha, A., Schweizer, S., Morisset, R., Strowig, T., Daniel, H., Helm, D., Küster, B., Krumsiek, J., & Ecker, J. (2018). The gut

microbiota promotes hepatic fatty acid desaturation and elongation in mice. *Nature Communications*, *9*(1), 3760. https://doi.org/10.1038/s41467-018-05767-4

Kjer-Nielsen, L., Patel, O., Corbett, A. J., Le Nours, J., Meehan, B., Liu, L., Bhati, M., Chen, Z., Kostenko, L., Reantragoon, R., Williamson, N. A., Purcell, A. W., Dudek, N. L., McConville, M. J., O'Hair, R. A. J., Khairallah, G. N., Godfrey, D. I., Fairlie, D. P., Rossjohn, J., & McCluskey, J. (2012). MR1 presents microbial vitamin B metabolites to MAIT cells. *Nature*, *491*(7426), 717–723. https://doi.org/10.1038/nature11605

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., Hagenbuch, B., Attie, A. D., & Rey, F. E. (2017). Host Genotype and Gut Microbiome Modulate Insulin Secretion and Diet-Induced Metabolic Phenotypes. *Cell Reports*, *18*(7), 1739–1750. https://doi.org/10.1016/j.celrep.2017.01.062

Kurilshikov, A., Medina-Gomez, C., Bacigalupe, R., Radjabzadeh, D., Wang, J., Demirkan, A., Le Roy, C. I., Raygoza Garay, J. A., Finnicum, C. T., Liu, X., Zhernakova, D. V., Bonder, M. J., Hansen, T. H., Frost, F., Rühlemann, M. C., Turpin, W., Moon, J.-Y., Kim, H.-N., Lüll, K., … Zhernakova, A. (2021). Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nature Genetics*, *53*(2), 156–165. https://doi.org/10.1038/s41588-020-00763-1

Labzin, L. I., Schmidt, S. V., Masters, S. L., Beyer, M., Krebs, W., Klee, K., Stahl, R., Lütjohann, D., Schultze, J. L., Latz, E., & De Nardo, D. (2015). ATF3 Is a Key Regulator of Macrophage IFN Responses. *The Journal of Immunology*, *195*(9), 4446–4455. https://doi.org/10.4049/jimmunol.1500204

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

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., & Pomp, D. (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 Biology*, *15*(12), 1–20. https://doi.org/10.1186/s13059-014-0552-6

Ley, R. E., Turnbaugh, P. J., Klein, S., & Gordon, J. I. (2006). Human gut microbes associated with obesity. *Nature*, *444*(7122), Article 7122. https://doi.org/10.1038/4441022a

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 1–16.

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

LifeLines cohort study, BIOS consortium, Zhernakova, D. V., Le, T. H., Kurilshikov, A., Atanasovska, B., Bonder, M. J., Sanna, S., Claringbould, A., Võsa, U., Deelen, P., Franke, L., de Boer, R. A., Kuipers, F., Netea, M. G., Hofker, M. H., Wijmenga, C., Zhernakova, A., & Fu, J. (2018). Individual variations in cardiovascular-disease-related protein levels are driven by genetics and gut microbiome. *Nature Genetics*, *50*(11), 1524–1532. https://doi.org/10.1038/s41588-018-0224-7

Linke, V., Overmyer, K. A., Miller, I. J., Brademan, D. R., Hutchins, P. D., Trujillo, E. A., Reddy, T. R., Russell, J. D., Cushing, E. M., Schueler, K. L., Stapleton, D. S., Rabaglia, M. E., Keller, M. P., Gatti, D. M., Keele, G. R., Pham, D., Broman, K. W., Churchill, G. A., Attie, A. D., & Coon, J. J. (2020). A large-scale genome–lipid association map guides lipid identification. *Nature Metabolism*, *2*(10), 1149–1162. https://doi.org/10.1038/s42255-020-00278-3

Liu, X., Xia, S., Zhang, Z., Wu, H., & Lieberman, J. (2021). Channelling inflammation: Gasdermins in physiology and disease. *Nature Reviews Drug Discovery*, *20*(5), 384–405. https://doi.org/10.1038/s41573-021-00154-z

López-Lara, I. M., Sohlenkamp, C., & Geiger, O. (2003). Membrane Lipids in Plant-Associated Bacteria: Their Biosyntheses and Possible Functions. *Molecular Plant-Microbe Interactions®*, *16*(7), 567–579. https://doi.org/10.1094/MPMI.2003.16.7.567

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Miller, I. J., Rees, E. R., Ross, J., Miller, I., Baxa, J., Lopera, J., Kerby, R. L., Rey, F. E., & Kwan, J. C. (2019). Autometa: Automated extraction of microbial genomes from individual shotgun metagenomes. *Nucleic Acids Research*, *47*(10), e57–e57. https://doi.org/10.1093/nar/gkz148

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: A new versatile metagenomic assembler. *Genome Research*, *27*(5), 824–834. https://doi.org/10.1101/gr.213959.116

O'Connor, A., Quizon, P. M., Albright, J. E., Lin, F. T., & Bennett, B. J. (2014). Responsiveness of cardiometabolic-related microbiota to diet is influenced by host genetics. *Mammalian Genome*, *25*(11–12), 583–599. https://doi.org/10.1007/s00335-014-9540-0

Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., & Phillippy, A. M. (2016). Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biology*, *17*(1), 132. https://doi.org/10.1186/s13059-016-0997-x

Org, E., Parks, B. W., Joo, J. W. J., Emert, B., Schwartzman, W., Kang, E. Y., Mehrabian, M., Pan, C., Knight, R., Gunsalus, R., Drake, T. A., Eskin, E., & Lusis, A. J. (2015). Genetic and environmental control of host-gut microbiota interactions. *Genome Research*, *25*(10), 1558–1569. https://doi.org/10.1101/gr.194118.115

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and

metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Parsons, J. B., & Rock, C. O. (2013). Bacterial lipids: Metabolism and membrane homeostasis. *Progress in Lipid Research*, *52*(3), 249–276. https://doi.org/10.1016/j.plipres.2013.02.002

Peri, F., Piazza, M., Calabrese, V., Damore, G., & Cighetti, R. (2010). Exploring the LPS/TLR4 signal pathway with small molecules. *Biochemical Society Transactions*, *38*(5), 1390–1395. https://doi.org/10.1042/BST0381390

Piazza, M., Rossini, C., Della Fiorentina, S., Pozzi, C., Comelli, F., Bettoni, I., Fusi, P., Costa, B., & Peri, F. (2009). Glycolipids and Benzylammonium Lipids as Novel Antisepsis Agents: Synthesis and Biological Characterization. *Journal of Medicinal Chemistry*, *52*(4), 1209–1213. https://doi.org/10.1021/jm801333m

Rey, F. E., Gonzalez, M. D., Cheng, J., Wu, M., Ahern, P. P., & Gordon, J. I. (2013). Metabolic niche of a prominent sulfate-reducing human gut bacterium. *Proceedings of the National Academy of Sciences*, *110*(33), 13582–13587. https://doi.org/10.1073/pnas.1312524110

Rühlemann, M. C., Hermes, B. M., Bang, C., Doms, S., Moitinho-Silva, L., Thingholm, L. B., Frost, F., Degenhardt, F., Wittig, M., Kässens, J., Weiss, F. U., Peters, A., Neuhaus, K., Völker, U., Völzke, H., Homuth, G., Weiss, S., Grallert, H., Laudes, M., … Franke, A. (2021). Genome-wide association study in 8,956 German individuals identifies influence of ABO histo-blood groups on gut microbiome. *Nature Genetics*, *53*(2), 147–155. https://doi.org/10.1038/s41588-020-00747-1

Ryzhakov, G., West, N. R., Franchini, F., Clare, S., Ilott, N. E., Sansom, S. N., Bullers, S. J., Pearson, C., Costain, A., Vaughan-Jackson, A., Goettel, J. A., Ermann, J., Horwitz, B. H., Buti, L., Lu, X., Mukhopadhyay, S., Snapper, S. B., & Powrie, F. (2018). Alpha kinase 1 controls intestinal inflammation by suppressing the IL-12/Th1 axis. *Nature Communications*, *9*(1), 3797. https://doi.org/10.1038/s41467-018-06085-5

Sanna, S., van Zuydam, N. R., Mahajan, A., Kurilshikov, A., Vich Vila, A., Võsa, U., Mujagic, Z., Masclee, A. A. M., Jonkers, D. M. A. E., Oosting, M., Joosten, L. A. B., Netea, M. G., Franke, L., Zhernakova, A., Fu, J., Wijmenga, C., & McCarthy, M. I. (2019). Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nature Genetics*, *51*(4), 600–605. https://doi.org/10.1038/s41588-019-0350-x

Schoeler, M., & Caesar, R. (2019). Dietary lipids, gut microbiota and lipid metabolism. *Reviews in Endocrine and Metabolic Disorders*, *20*(4), 461–472. https://doi.org/10.1007/s11154-019-09512-0

Shi, J., Zhao, Y., Wang, K., Shi, X., Wang, Y., Huang, H., Zhuang, Y., Cai, T., Wang, F., & Shao, F. (2015). Cleavage of GSDMD by inflammatory caspases determines pyroptotic cell death. *Nature*, *526*(7575), 660–665. https://doi.org/10.1038/nature15514

Sohlenkamp, C., & Geiger, O. (2016). Bacterial membrane lipids: Diversity in structures and pathways. *FEMS Microbiology Reviews*, *40*(1), 133–159. https://doi.org/10.1093/femsre/fuv008

Svenson, K. L., Gatti, D. M., Valdar, W., Welsh, C. E., Cheng, R., Chesler, E. J., Palmer, A. A., McMillan, L., & Churchill, G. A. (2012). High-Resolution Genetic Mapping Using the Mouse Diversity Outbred Population. *Genetics*, *190*(2), 437–447. https://doi.org/10.1534/genetics.111.132597

Tian, J., Keller, M. P., Oler, A. T., Rabaglia, M. E., Schueler, K. L., Stapleton, D. S., Broman, A. T., Zhao, W., Kendziorski, C., Yandell, B. S., Hagenbuch, B., Broman, K. W., & Attie, A. D. (2015). Identification of the Bile Acid Transporter *Slco1a6* as a Candidate Gene That Broadly Affects Gene Expression in Mouse Pancreatic Islets. *Genetics*, *201*(3), 1253–1262. https://doi.org/10.1534/genetics.115.179432

Tremaroli, V., & Bäckhed, F. (2012). Functional interactions between the gut microbiota and host metabolism. *Nature*, *489*(7415), 242–249. https://doi.org/10.1038/nature11552

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A., Affourtit, J. P., Egholm, M., Henrissat, B., Heath, A. C., Knight, R., & Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature*, *457*(7228), 480–484. https://doi.org/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*(7122), 1027–1031. https://doi.org/10.1038/nature05414

Vences-Guzmán, M. Á., Geiger, O., & Sohlenkamp, C. (2012). Ornithine lipids and their structural modifications: From A to E and beyond. *FEMS Microbiology Letters*, *335*(1), 1–10. https://doi.org/10.1111/j.1574-6968.2012.02623.x

Vences-Guzmán, M. Á., Guan, Z., Escobedo-Hinojosa, W. I., Bermúdez-Barrientos, J. R., Geiger, O., & Sohlenkamp, C. (2015). Discovery of a bifunctional acyltransferase responsible for ornithine lipid synthesis in *S erratia proteamaculans*: A bifunctional ornithine lipid synthase. *Environmental Microbiology*, *17*(5), 1487–1496. https://doi.org/10.1111/1462-2920.12562

Wang, J., Thingholm, L. B., Skiecevičienė, J., Rausch, P., Kummen, M., Hov, J. R., Degenhardt, F., Heinsen, F.-A., Rühlemann, M. C., Szymczak, S., Holm, K., Esko, T., Sun, J., Pricop-Jeckstadt, M., Al-Dury, S., Bohov, P., Bethune, J., Sommer, F., Ellinghaus, D., … Franke, A. (2016). Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. *Nature Genetics*, *48*(11), 1396–1406. https://doi.org/10.1038/ng.3695

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*(7341), 57–63. https://doi.org/10.1038/nature09922

Wen, L., Ley, R. E., Volchkov, P. Yu., Stranges, P. B., Avanesyan, L., Stonebraker, A. C., Hu, C., Wong, F. S., Szot, G. L., Bluestone, J. A., Gordon, J. I., & Chervonsky, A. V. (2008). Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature*, *455*(7216), 1109–1113. https://doi.org/10.1038/nature07336

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., Heath, A. C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J. G., Lozupone, C. A., Lauber, C., Clemente, J. C., Knights, D., … Gordon, J. I. (2012). Human gut microbiome viewed across age and geography. *Nature*, *486*(7402), 222–227. https://doi.org/10.1038/nature11053

Yeo, K. S., Tan, M. C., Wong, W. Y., Loh, S. W., Lam, Y. L., Tan, C. L., Lim, Y.-Y., & Ea, C.-K. (2016). JMJD8 is a positive regulator of TNF-induced NF-κB signaling. *Scientific Reports*, *6*(1), 34125. https://doi.org/10.1038/srep34125

You, D., Jung, B. C., Villivalam, S. D., Lim, H.-W., & Kang, S. (2021). JMJD8 is a Novel Molecular Nexus Between Adipocyte-Intrinsic Inflammation and Insulin Resistance. *Diabetes*, *71*(1), 43–59. https://doi.org/10.2337/db21-0596

Zhou, P., She, Y., Dong, N., Li, P., He, H., Borio, A., Wu, Q., Lu, S., Ding, X., Cao, Y., Xu, Y., Gao, W., Dong, M., Ding, J., Wang, D.-C., Zamyatina, A., & Shao, F. (2018). Alpha-kinase 1 is a cytosolic innate immune receptor for bacterial ADP-heptose. *Nature*, *561*(7721), 122–126. https://doi.org/10.1038/s41586-018-0433-3

2.9 Figures

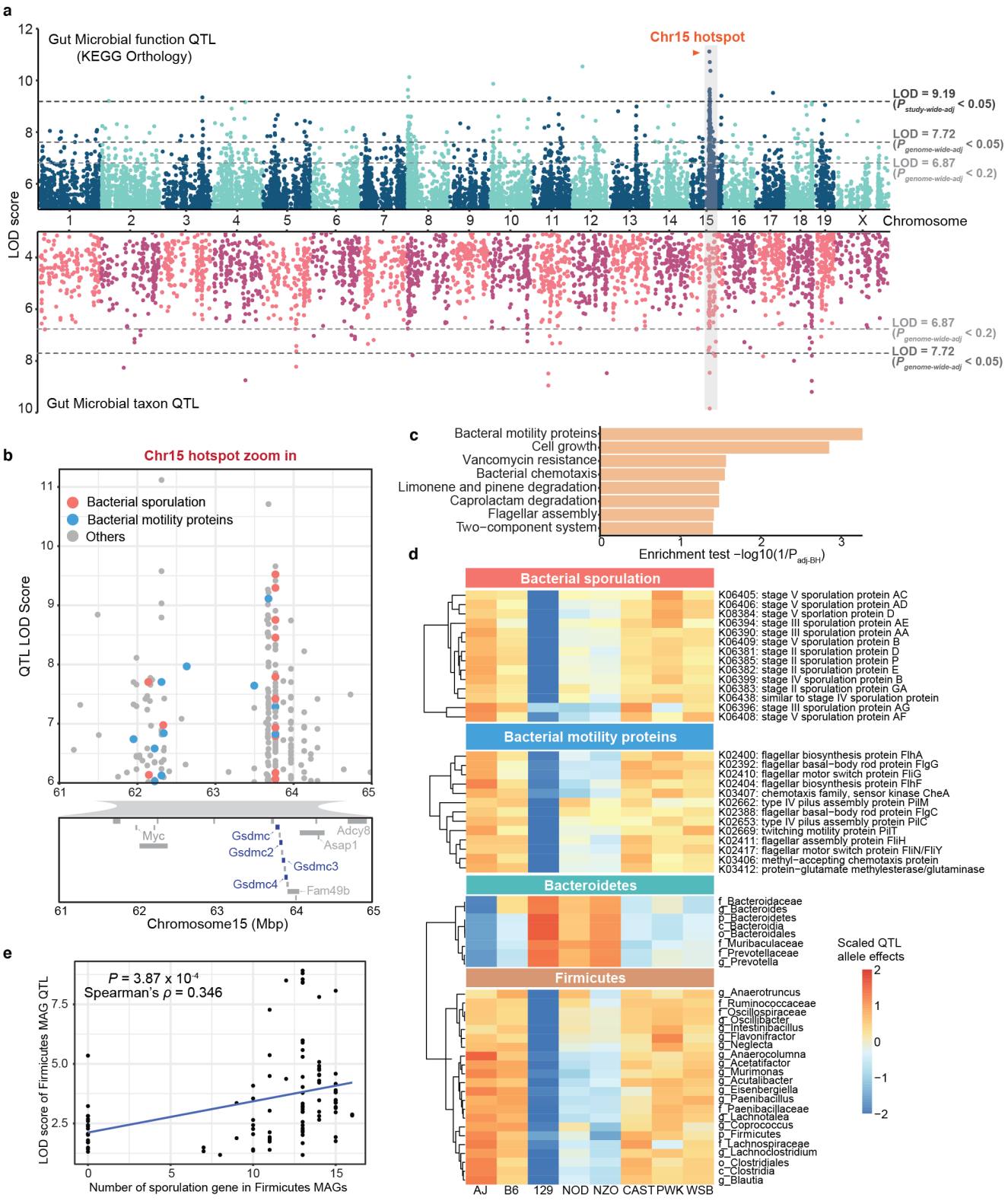

a
Gut Microbial function QTL
(KEGG Orthology)
Chr15 hotspot
LOD = 9.19
(P study-wide-adj < 0.05)
LOD = 7.72
(P genome-wide-adj < 0.05)
LOD = 6.87
(P genome-wide-adj < 0.2)
LOD score
Chromosome
Gut Microbial taxon QTL
b
Chr15 hotspot zoom in
Bacterial sporulation
Bacterial motility proteins
Others
QTL LOD Score
Myc
Gsdmc
Gsdmc2
Gsdmc3
Gsdmc4
Adcy8
Asap1
Fam49b
Chromosome15 (Mbp)
c
Bacteral motility proteins
Cell growth
Vancomycin resistance
Bacterial chemotaxis
Limonene and pinene degradation
Caprolactam degradation
Flagellar assembly
Two−component system
Enrichment test −log10(1/P adj-BH)
d
Bacterial sporulation
K06405: stage V sporulation protein AC
K06406: stage V sporulation protein AD
K08384: stage V sporlation protein D
K06394: stage III sporulation protein AE
K06390: stage III sporulation protein AA
K06409: stage V sporulation protein B
K06381: stage II sporulation protein D
K06385: stage II sporulation protein P
K06382: stage II sporulation protein E
K06399: stage IV sporulation protein B
K06383: stage II sporulation protein GA
K06438: similar to stage IV sporulation protein
K06396: stage III sporulation protein AG
K06408: stage V sporulation protein AF
Bacterial motility proteins
K02400: flagellar biosynthesis protein FlhA
K02392: flagellar basal−body rod protein FlgG
K02410: flagellar motor switch protein FliG
K02404: flagellar biosynthesis protein FlhF
K03407: chemotaxis family, sensor kinase CheA
K02662: type IV pilus assembly protein PilM
K02388: flagellar basal−body rod protein FlgC
K02653: type IV pilus assembly protein PilC
K02669: twitching motility protein PilT
K02411: flagellar assembly protein FliH
K02417: flagellar motor switch protein FliN/FliY
K03406: methyl−accepting chemotaxis protein
K03412: protein−glutamate methylesterase/glutaminase
Bacteroidetes
f_Bacteroidaceae
g_Bacteroides
p_Bacteroidetes
c_Bacteroidia
o_Bacteroidales
f_Muribaculaceae
f_Prevotellaceae
g_Prevotella
Firmicutes
g_Anaerotruncus
f_Ruminococcaceae
f_Oscillospiraceae
g_Oscillibacter
g_Intestinibacillus
g_Flavonifractor
g_Neglecta
g_Anaerocolumna
g_Acetatifactor
g_Murimonas
g_Acutalibacter
g_Eisenbergiella
g_Paenibacillus
f_Paenibacillaceae
g_Lachnotalea
g_Coprococcus
p_Firmicutes
f_Lachnospiraceae
g_Lachnoclostridium
o_Clostridiales
c_Clostridia
g_Blautia
AJ B6 129 NOD NZO CAST PWK WSB
Scaled QTL
allele effects
e
P = 3.87 x 10-4
Spearman's ρ = 0.346
LOD score of Firmicutes MAG QTL
Number of sporulation gene in Firmicutes MAGs

Figure 2.1. Genetic architecture of QTL for microbial traits in the DO mouse cohort. **a**, QTL mapping results for 2,803 gut microbial KO function traits (top panel) and 187 bacterial taxa traits (bottom panel) using sex, days on diet and cohort as covariates. Each dot represents a QTL on the mouse genome for a given trait. Dashed lines represent significance thresholds for QTL determined by permutation tests (LOD > 9.19, $P_{\text{study-wide-adj}} < 0.05$; LOD > 7.72, $P_{\text{genome-wide-adj}} < 0.05$; LOD > 6.87, $P_{\text{genome-wide-adj}} < 0.2$). QTL hotspot at Chromosome 15 is highlighted by grey shading and orange colour text. **b**, Gut microbiome QTL hotspot on Chr15 has multiple bacterial sporulation and motility functions mapping to it. Protein coding genes are displayed for Chr15: 61–65 Mbp region, Gasdermin genes are highlighted in blue. **c**, Enrichment analysis (Fisher's exact test) for functions mapping at hotspot on Chr15. **d**, QTL for microbial functions that mapped to Chromosome 15 hotspot had negative 129S1/SvImJ allele effects. QTL for Firmicutes mapping to Chromosome 15 hotspot had negative 129S1/SvImJ allele effects, whereas QTL for Bacteroidetes mapping to this locus had positive 129S1/SvImJ allele effects. **e**, Spearman correlation analysis between the number of sporulation KOs detected in Firmicutes MAGs mapping at Chromosome 15 QTL hotspot and the LOD scores for these MAGs ($P = 3.87 \times 10^{-3}$, Spearman's $\rho = 0.346$).

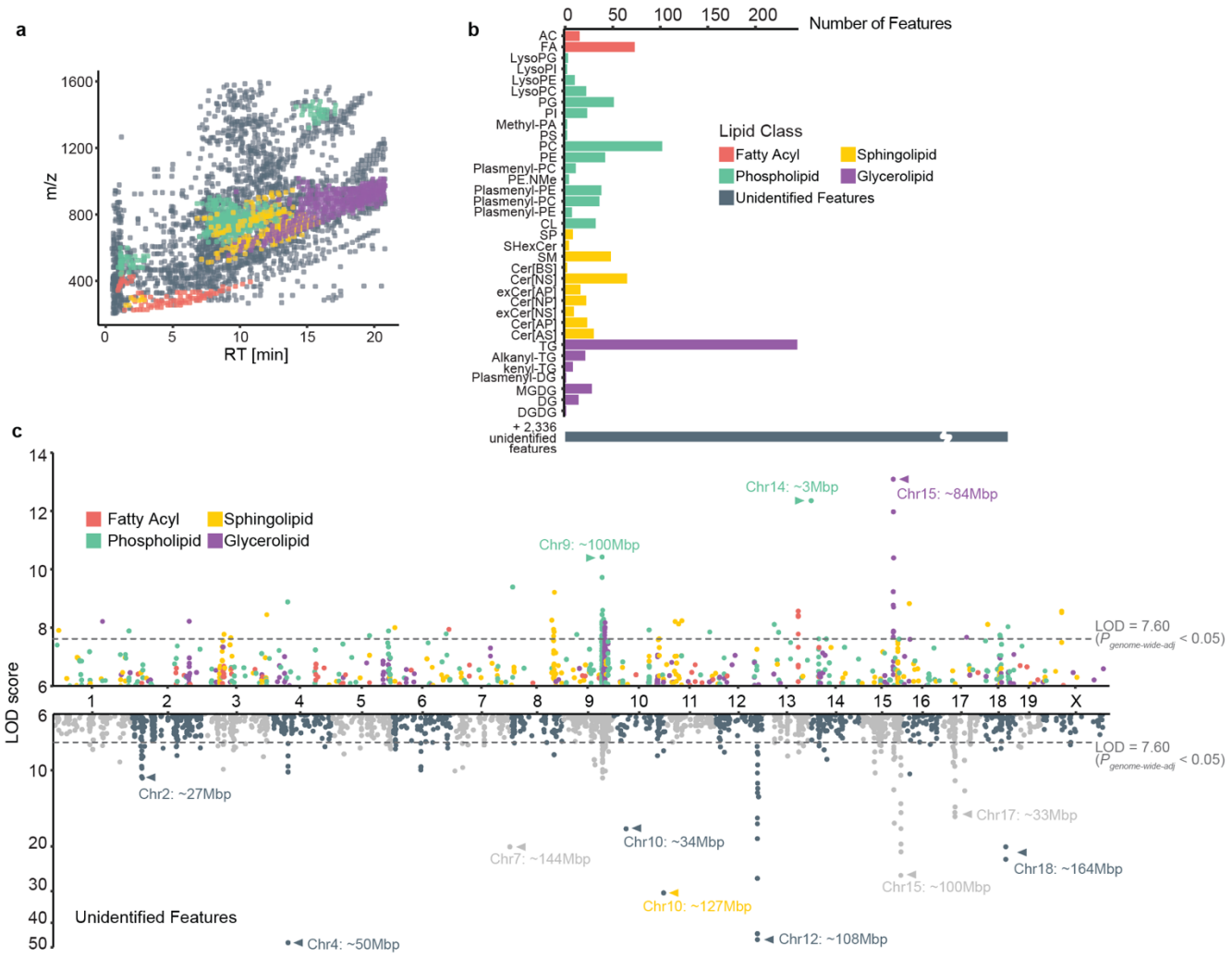


Figure 2.2. Genetic architecture of the caecal lipidome in DO mice. a, A total of 3,384 caecal lipid features were quantified across 381 DO mice, 1,048 of which were identified as lipids from four major classes. Each dot represents a caecal lipid feature. Features of each class occupied characteristic regions in the *m*/*z* – RT space. **b**, Identified lipids belonged to 35 lipid subclasses, with bacteria-associated PG and MGDG as common subclasses. **c**, A total of 3,964 suggestive caecal lipid QTL (LOD > 6, $P_{\text{genome-wide-adj}} < 0.2$) and 12 QTL hotspots were identified. Hotspots are marked with arrows and the corresponding genomic locus indicated. Dashed lines represent significance thresholds for QTL as determined by permutation tests (LOD > 7.60, $P_{\text{genome-wide-adj}} < 0.05$). Of the identified lipids, 68.2% showed a total of 1,162 QTL (top panel), while a similar

portion of 70.1% of unidentified features contributed 2,802 QTL (bottom panel). RT, retention time. For lipid class abbreviations, see Supplementary Table 16.

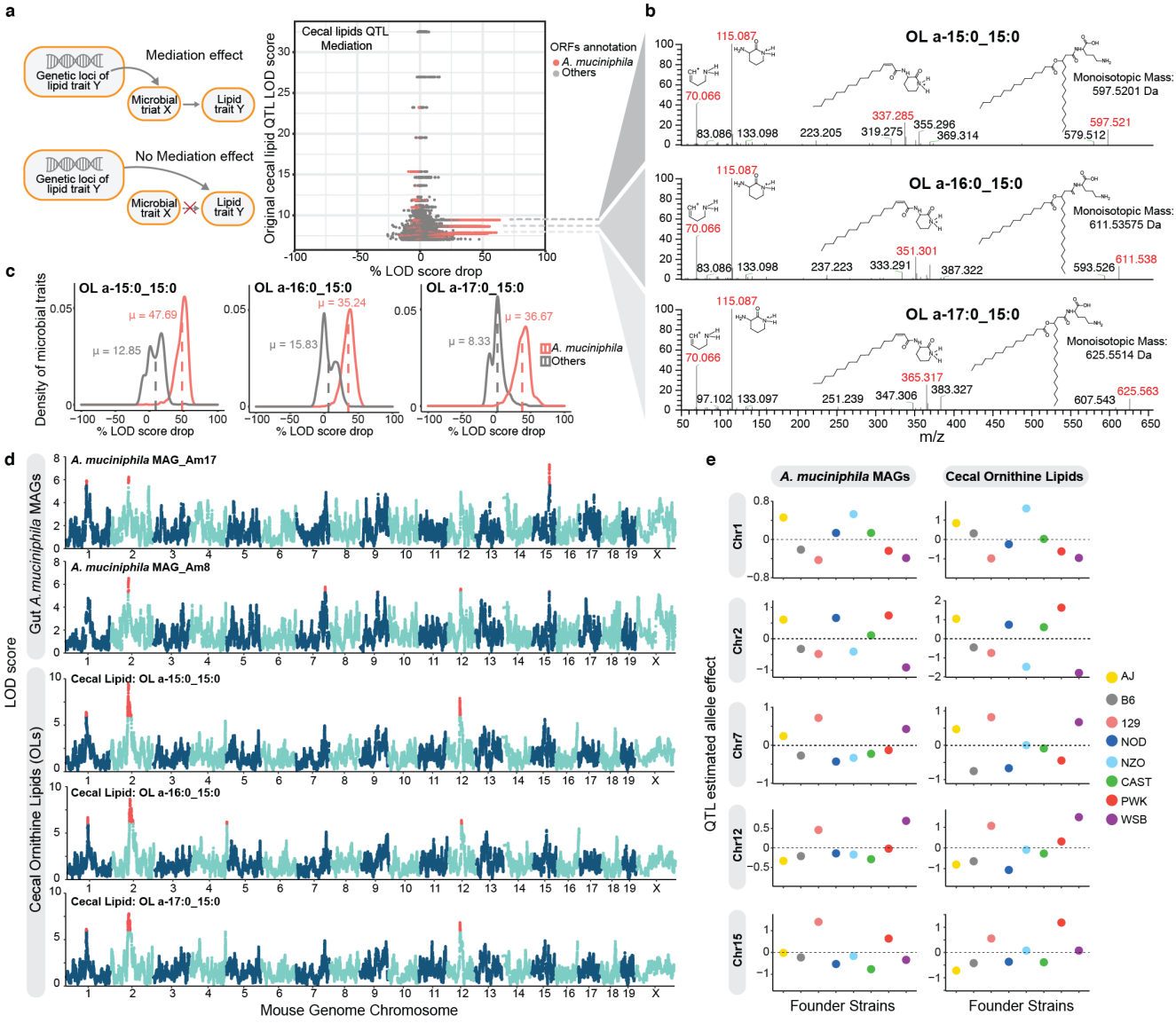


Figure 2.3. Mediation analysis revealed potential causal relationship between *A. muciniphila* and OL. a, Illustration of Mediation effect model and Null model. Each dot in the scatterplot represents the result of the mediation test for a gut microbial metagene–caecal lipid feature pair; *x* axis shows the drop in QTL LOD score for caecal lipid features when adding gut microbial metagenes as covariates to the caecal lipid QTL model; *y* axis shows the original QTL LOD score for each caecal lipid. Dots with the same *y* axis value represent the mediation test of individual metagenes with one caecal lipid feature. A high QTL LOD score drop represents a significant mediation effect of the gut microbial feature to the caecal lipid feature. Association of

three unknown caecal lipid features with the host genome was impacted by *A. muciniphila* genes. This is depicted as multiple red dots (many dots appear as lines) showing significant mediation effects. **b**, Three lipid features mediated by *A. muciniphila* genes were identified as ornithine lipids. The dashed lines connecting **a** and **b** point to the fragmentation patterns of identified ornithine lipids, as shown by the *m*/*z* values; key fragments are shown in red colour together with their respective chemical structures. **c**, Distribution of LOD score drop when adding individual *A. muciniphila* genes as covariates (Mediation model) or adding individual genes not from *A. muciniphila* as covariates (Null model) for three identified ornithine lipids. **d**, Three ornithine lipids species QTL co-mapped at five loci (Chromosome 1, Chromosome 2, Chromosome 7, Chromosome 12, Chromosome 15) with *A. muciniphila* MAGs QTL.QTL with LOD > 5.5 are highlighted by red colour. **e**, Founder allele effects for *A. muciniphila* MAGs and caecal OL were estimated in the DO population from the founder strain coefficients observed for the corresponding QTL at each locus from **d**.

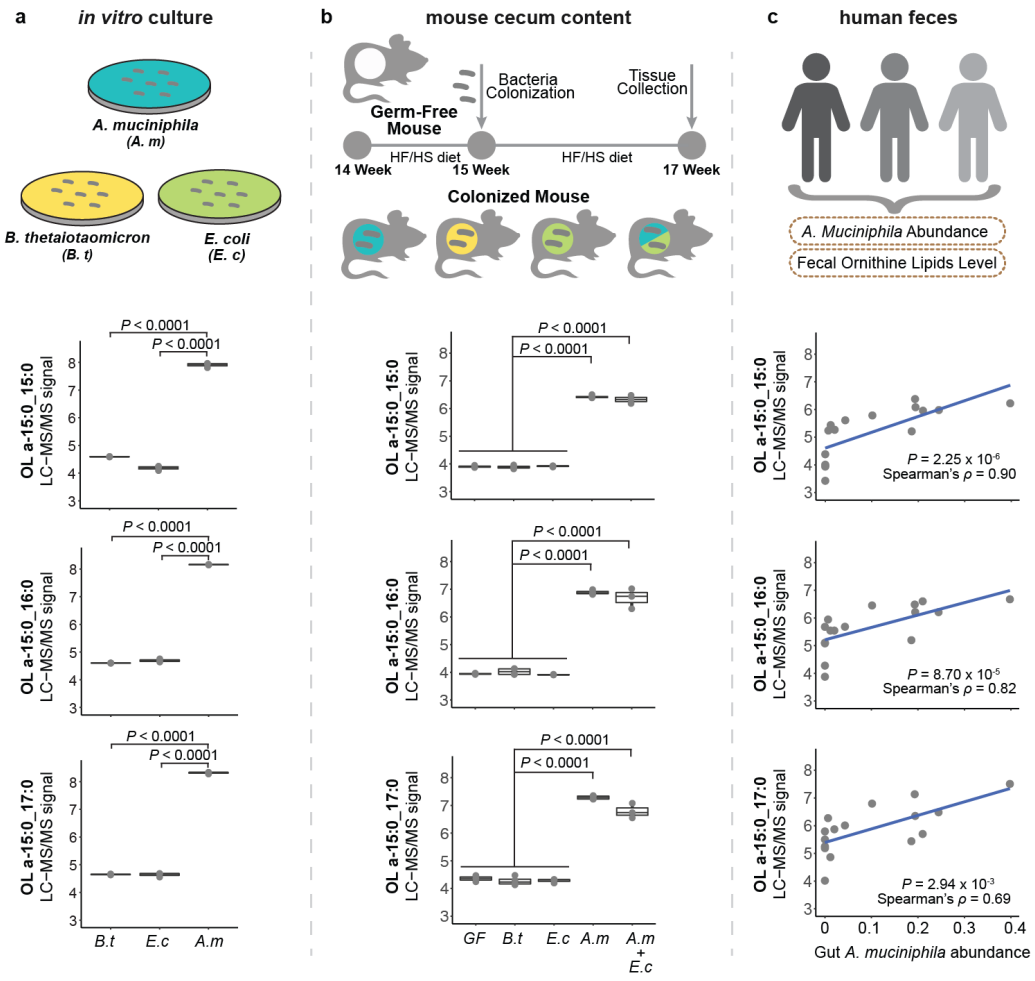


Figure 2.4. *A. muciniphila* produces OL in the mouse and human gut. a, OL abundance for the three major species detected in mice in cell pellets collected from *A. muciniphila* (*A. m*), *B. thetalotamicron* (*B. t*) and *E. coli* (*E. c*) grown in vitro ($n = 3$ biologically independent samples per organism). **b**, OL detected in caecal contents from gnotobiotic mice colonized with *A. muciniphila*, *B. thetaiotaomicron*, *E. coli* and *A. muciniphila* plus *E. coli* for two weeks ($n = 3$–4 mice per treatment). **c**, Detection of prominent OL species in human faecal samples is significantly correlated with *A. muciniphila* abundance as determined by two-sided Spearman correlation ($n = 16$ independent faecal samples). Box and whisker plots denote the interquartile range, median and spread of points within 1.5 times the interquartile range; data beyond the end of the whiskers are plotted individually. Statistical difference between treatment groups was tested by unpaired two-sided Welch's *t*- test.

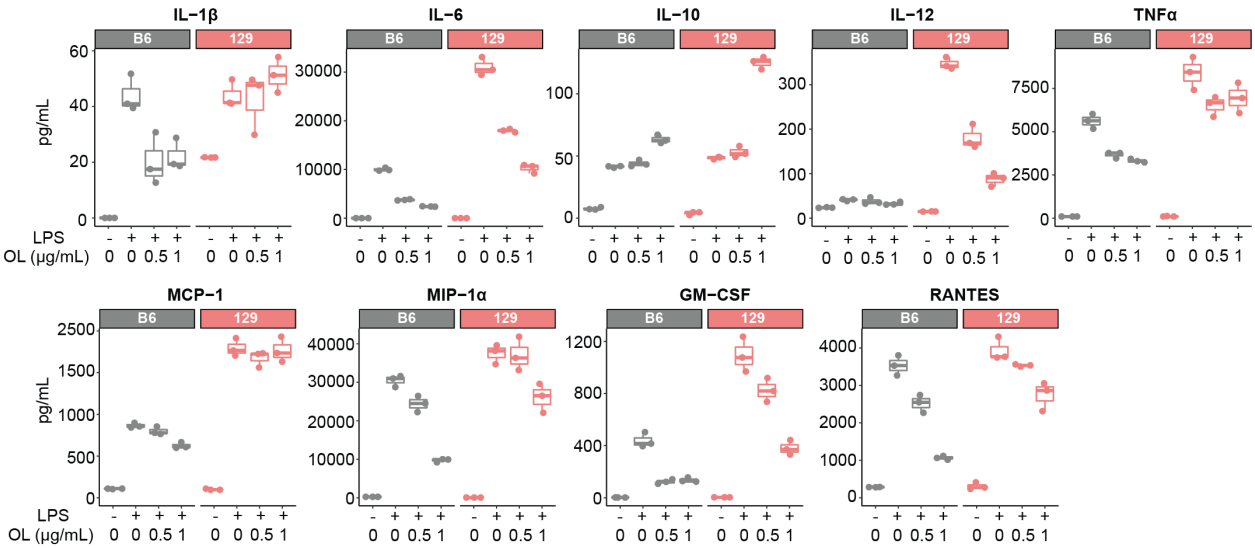


Figure 2.5. OL modulate LPS-induced production of cytokines from BMDM. Levels of IL-1β, IL-6, IL-10, IL-12, TNF-α, MCP-1, MIP-1α, GM-CSF and RANTES detected in supernatants from B6 and 129 mice BMDM stimulated with LPS (10 ng ml^{-1}) and different concentrations of OL. Box and whisker plots denote the interquartile range, median and spread of points within 1.5 times the interquartile range; data beyond the end of the whiskers are plotted individually.

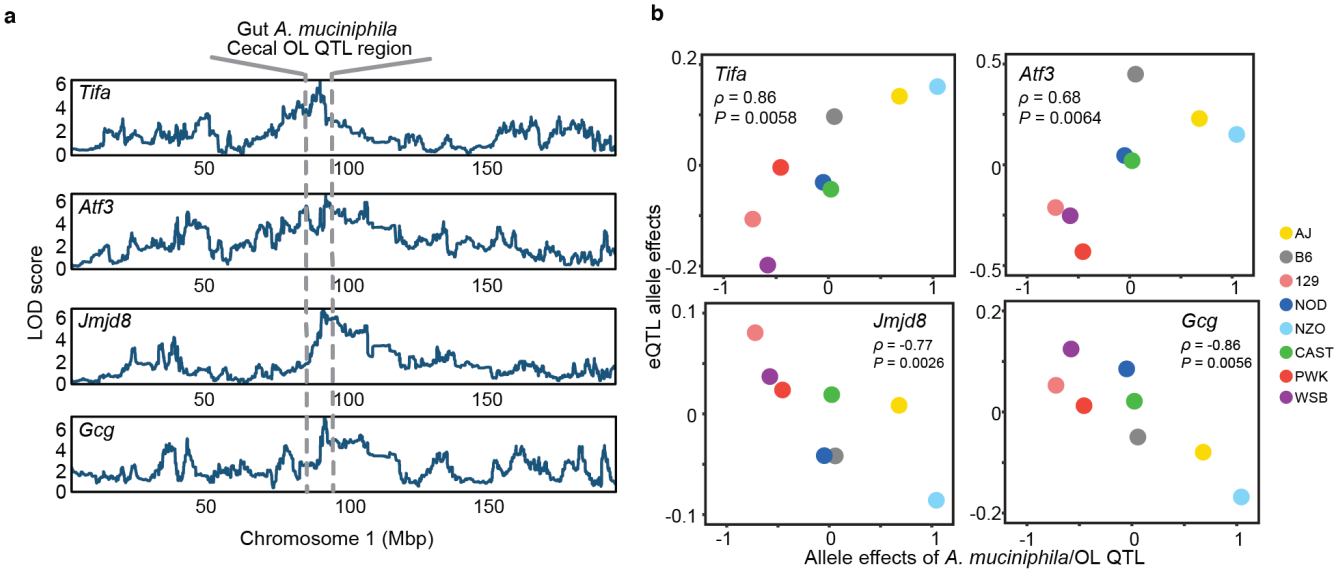
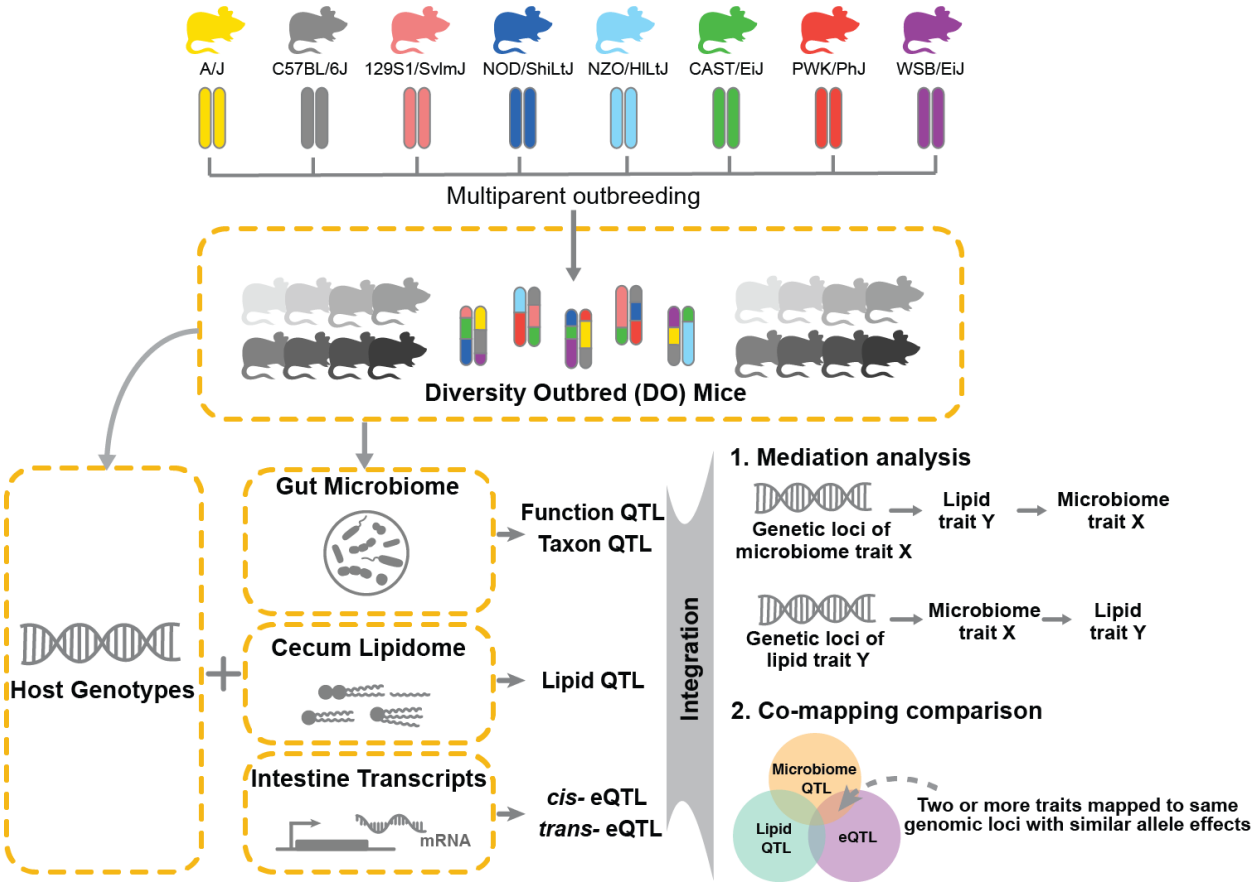
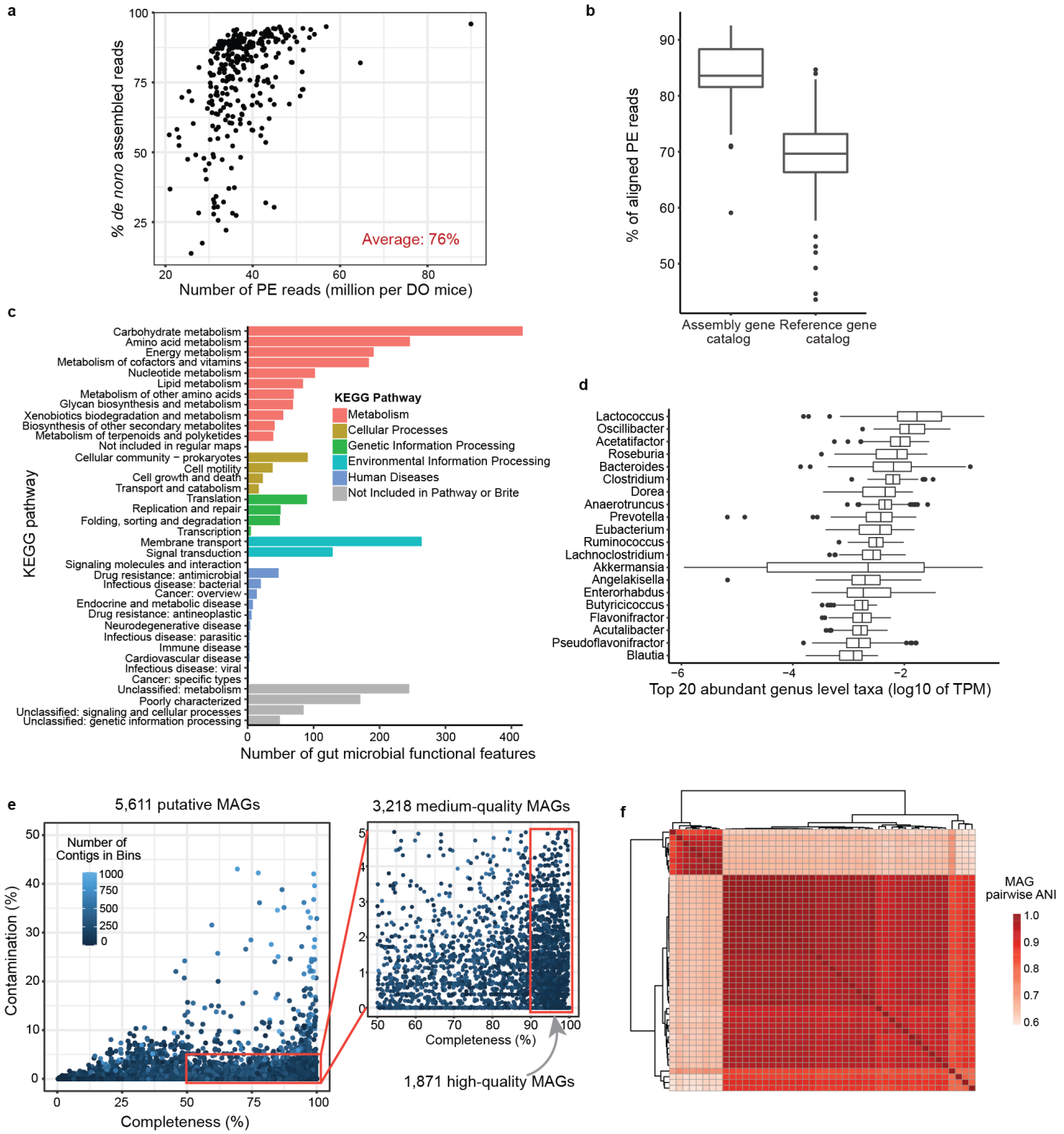


Figure 2.6. eQTL for distal small intestine (ileum) genes that co-map with *A. muciniphila* and caecal OL at Chromosome 1. a, QTL of *A. muciniphila*, caecal OL and eQTL for *Tifa*, *Atf3*, *Jmjd8* and *Gcg* co-map at Chr1: 90–95 Mbp. LOD score in *y* axis represents significance of QTL for each trait. **b**, Spearman correlation of allele effects between *Tifa*, *Atf3*, *Jmjd8* and *Gcg* gene eQTL and *A. muciniphila*/OL QTL.

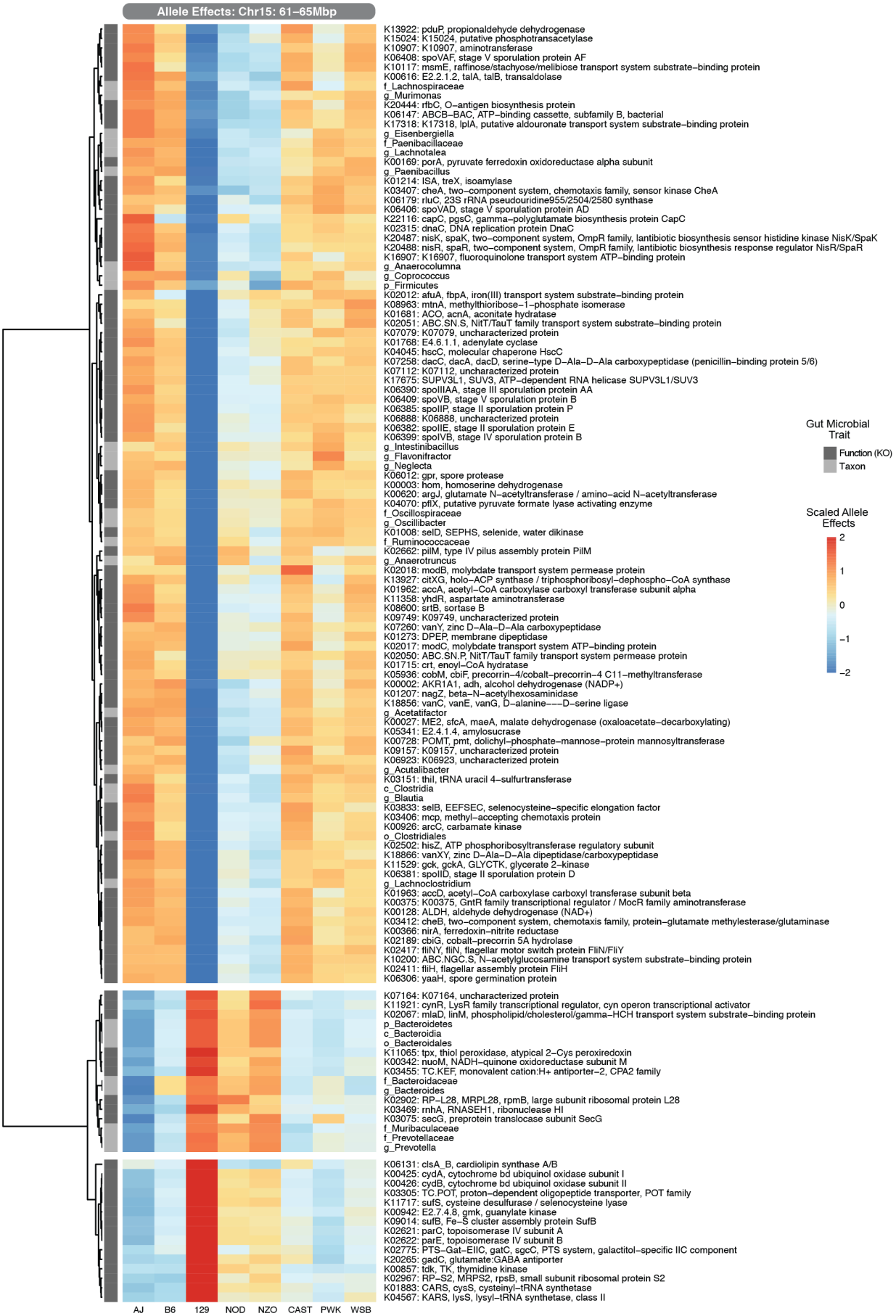


Supplementary Figure 2.1 Overview of the study. Fecal metagenomes (n = 264 animals), caecal lipidomes (n = 381 animals) and distal small intestine transcriptomes (n = 234 animals) were generated from Diversity Outbred mice. Quantitative trait loci (QTL) analysis identified genomic regions associated with variations in bacterial taxa, bacterial functions, levels of bacterial- and host-derived lipids and small intestine transcript levels. Mediation analysis and co-mapping comparisons were used to identify causal links between traits.

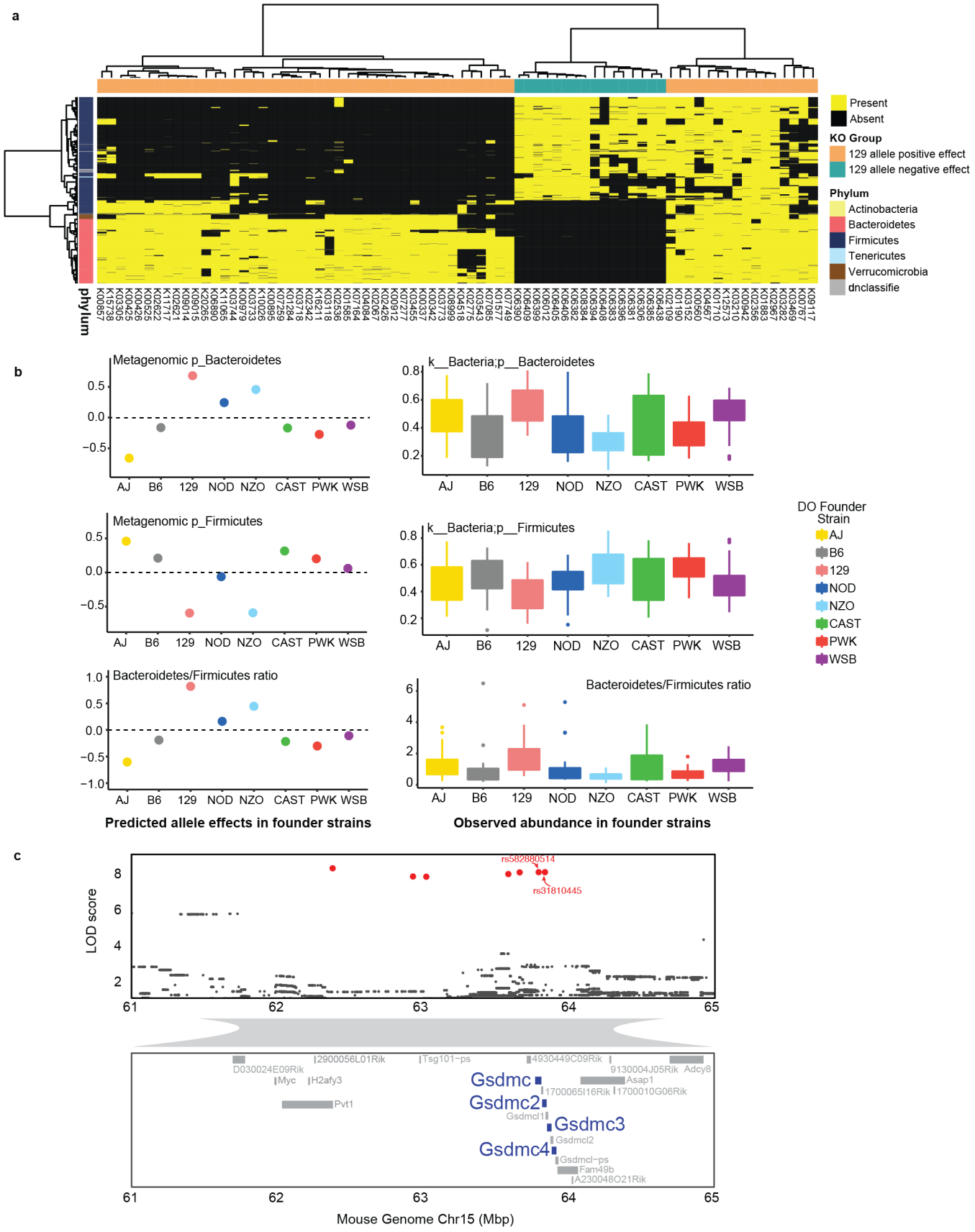


Supplementary Figure 2.2 DO metagenomic analysis. a, Average percent of assembled reads across all samples. **b**, Comparison of percent of reads mapping to our generated assembly *vs*. public database (n = 297 animals). **c**, Microbial functions detected for KEGG pathways across all metagenomes. KEGG Orthology (KO) numbers were identified by annotating predicted ORFs to

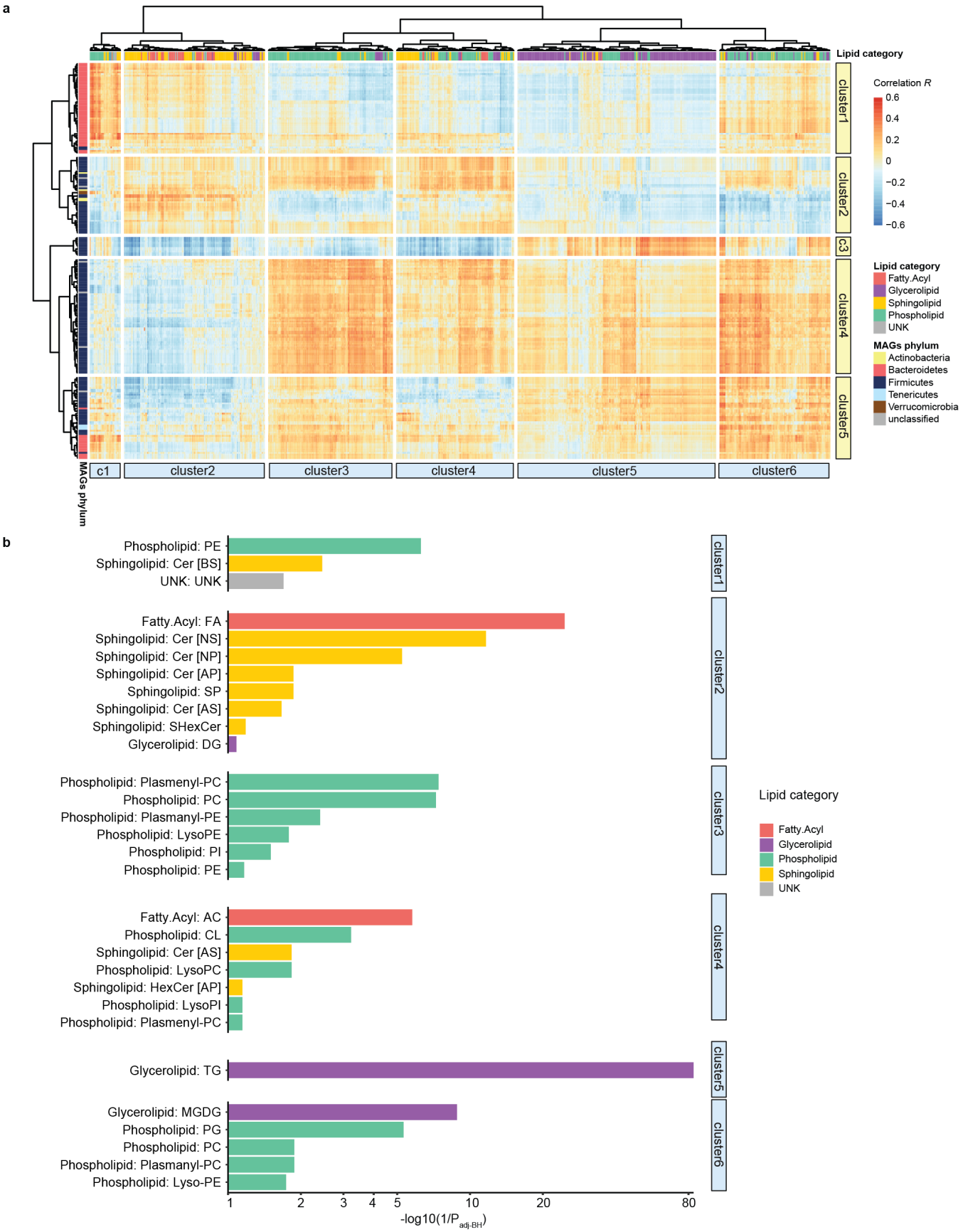
the KEGG database. **d**, Top 20 gut microbial genera detected across all DO mice (n = 264 animals). **e**, Quality of metagenome-assembled genomes. **f**, Two variants of *A. muciniphila* MAGs detected in the DO mice. Box and whisker plots denote the interquartile range, median and spread of points within 1.5 times the interquartile range, data beyond the end of the whiskers are plotted individually.


Allele Effects: Chr15: 61–65Mbp
K13922: pduP, propionaldehyde dehydrogenase
K15024: K15024, putative phosphotransacetylase
K10907: K10907, aminotransferase
K06408: spoVAF, stage V sporulation protein AF
K10117: msmE, raffinose/stachyose/melibiose transport system substrate–binding protein
K00616: E2.2.1.2, talA, talB, transaldolase
f_Lachnospiraceae
g_Murimonas
K20444: rfbC, O–antigen biosynthesis protein
K06147: ABCB–BAC, ATP–binding cassette, subfamily B, bacterial
K17318: K17318, lplA, putative aldouronate transport system substrate–binding protein
g_Eisenbergiella
f_Paenibacillaceae
g_Lachnotalea
K00169: porA, pyruvate ferredoxin oxidoreductase alpha subunit
g_Paenibacillus
K01214: ISA, treX, isoamylase
K03407: cheA, two–component system, chemotaxis family, sensor kinase CheA
K06179: rluC, 23S rRNA pseudouridine955/2504/2580 synthase
K06406: spoVAD, stage V sporulation protein AD
K22116: capC, pgsC, gamma–polyglutamate biosynthesis protein CapC
K02315: dnaC, DNA replication protein DnaC
K20487: nisK, spaK, two–component system, OmpR family, lantibiotic biosynthesis sensor histidine kinase NisK/SpaK
K20488: nisR, spaR, two–component system, OmpR family, lantibiotic biosynthesis response regulator NisR/SpaR
K16907: K16907, fluoroquinolone transport system ATP–binding protein
g_Anaerocolumna
g_Coprococcus
p_Firmicutes
K02012: afuA, fbpA, iron(III) transport system substrate–binding protein
K08963: mtnA, methylthioribose–1–phosphate isomerase
K01681: ACO, acnA, aconitate hydratase
K02051: ABC.SN.S, NitT/TauT family transport system substrate–binding protein
K07079: K07079, uncharacterized protein
K01768: E4.6.1.1, adenylate cyclase
K04045: hscC, molecular chaperone HscC
K07258: dacC, dacA, dacD, serine–type D–Ala–D–Ala carboxypeptidase (penicillin–binding protein 5/6)
K07112: K07112, uncharacterized protein
K17675: SUPV3L1, SUV3, ATP–dependent RNA helicase SUPV3L1/SUV3
K06390: spoIIIAA, stage III sporulation protein AA
K06409: spoVB, stage V sporulation protein B
K06385: spoIIP, stage II sporulation protein P
K06888: K06888, uncharacterized protein
K06382: spoIIE, stage II sporulation protein E
K06399: spoIVB, stage IV sporulation protein B
g_Intestinibacillus
g_Flavonifractor
g_Neglecta
K06012: gpr, spore protease
K00003: hom, homoserine dehydrogenase
K00620: argJ, glutamate N–acetyltransferase / amino–acid N–acetyltransferase
K04070: pflX, putative pyruvate formate lyase activating enzyme
f_Oscillospiraceae
g_Oscillibacter
K01008: selD, SEPHS, selenide, water dikinase
f_Ruminococcaceae
K02662: pilM, type IV pilus assembly protein PilM
g_Anaerotruncus
K02018: modB, molybdate transport system permease protein
K13927: citXG, holo–ACP synthase / triphosphoribosyl–dephospho–CoA synthase
K01962: accA, acetyl–CoA carboxylase carboxyl transferase subunit alpha
K11358: yhdR, aspartate aminotransferase
K08600: srtB, sortase B
K09749: K09749, uncharacterized protein
K07260: vanY, zinc D–Ala–D–Ala carboxypeptidase
K01273: DPEP, membrane dipeptidase
K02017: modC, molybdate transport system ATP–binding protein
K02050: ABC.SN.P, NitT/TauT family transport system permease protein
K01715: crt, enoyl–CoA hydratase
K05936: cobM, cbiF, precorrin–4/cobalt–precorrin–4 C11–methyltransferase
K00002: AKR1A1, adh, alcohol dehydrogenase (NADP+)
K01207: nagZ, beta–N–acetylhexosaminidase
K18856: vanC, vanE, vanG, D–alanine–––D–serine ligase
g_Acetatifactor
K00027: ME2, sfcA, maeA, malate dehydrogenase (oxaloacetate–decarboxylating)
K05341: E2.4.1.4, amylosucrase
K00728: POMT, pmt, dolichyl–phosphate–mannose–protein mannosyltransferase
K09157: K09157, uncharacterized protein
K06923: K06923, uncharacterized protein
g_Acutalibacter
K03151: thiI, tRNA uracil 4–sulfurtransferase
c_Clostridia
g_Blautia
K03833: selB, EEFSEC, selenocysteine–specific elongation factor
K03406: mcp, methyl–accepting chemotaxis protein
K00926: arcC, carbamate kinase
o_Clostridiales
K02502: hisZ, ATP phosphoribosyltransferase regulatory subunit
K18866: vanXY, zinc D–Ala–D–Ala dipeptidase/carboxypeptidase
K11529: gck, gckA, GLYCTK, glycerate 2–kinase
K06381: spoIID, stage II sporulation protein D
g_Lachnoclostridium
K01963: accD, acetyl–CoA carboxylase carboxyl transferase subunit beta
K00375: K00375, GntR family transcriptional regulator / MocR family aminotransferase
K00128: ALDH, aldehyde dehydrogenase (NAD+)
K03412: cheB, two–component system, chemotaxis family, protein–glutamate methylesterase/glutaminase
K00366: nirA, ferredoxin–nitrite reductase
K02189: cbiG, cobalt–precorrin 5A hydrolase
K02417: fliNY, fliN, flagellar motor switch protein FliN/FliY
K10200: ABC.NGC.S, N–acetylglucosamine transport system substrate–binding protein
K02411: fliH, flagellar assembly protein FliH
K06306: yaaH, spore germination protein
K07164: K07164, uncharacterized protein
K11921: cynR, LysR family transcriptional regulator, cyn operon transcriptional activator
K02067: mlaD, linM, phospholipid/cholesterol/gamma–HCH transport system substrate–binding protein
p_Bacteroidetes
c_Bacteroidia
o_Bacteroidales
K11065: tpx, thiol peroxidase, atypical 2–Cys peroxiredoxin
K00342: nuoM, NADH–quinone oxidoreductase subunit M
K03455: TC.KEF, monovalent cation:H+ antiporter–2, CPA2 family
f_Bacteroidaceae
g_Bacteroides
K02902: RP–L28, MRPL28, rpmB, large subunit ribosomal protein L28
K03469: rnhA, RNASEH1, ribonuclease HI
K03075: secG, preprotein translocase subunit SecG
f_Muribaculaceae
f_Prevotellaceae
g_Prevotella
K06131: clsA_B, cardiolipin synthase A/B
K00425: cydA, cytochrome bd ubiquinol oxidase subunit I
K00426: cydB, cytochrome bd ubiquinol oxidase subunit II
K03305: TC.POT, proton–dependent oligopeptide transporter, POT family
K11717: sufS, cysteine desulfurase / selenocysteine lyase
K00942: E2.7.4.8, gmk, guanylate kinase
K09014: sufB, Fe–S cluster assembly protein SufB
K02621: parC, topoisomerase IV subunit A
K02622: parE, topoisomerase IV subunit B
K02775: PTS–Gat–EIIC, gatC, sgcC, PTS system, galactitol–specific IIC component
K20265: gadC, glutamate:GABA antiporter
K00857: tdk, TK, thymidine kinase
K02967: RP–S2, MRPS2, rpsB, small subunit ribosomal protein S2
K01883: CARS, cysS, cysteinyl–tRNA synthetase
K04567: KARS, lysS, lysyl–tRNA synthetase, class II
AJ
B6
129
NOD
NZO
CAST
PWK
WSB
Gut Microbial Trait
Function (KO)
Taxon
Scaled Allele Effects
2
1
0
−1
−2

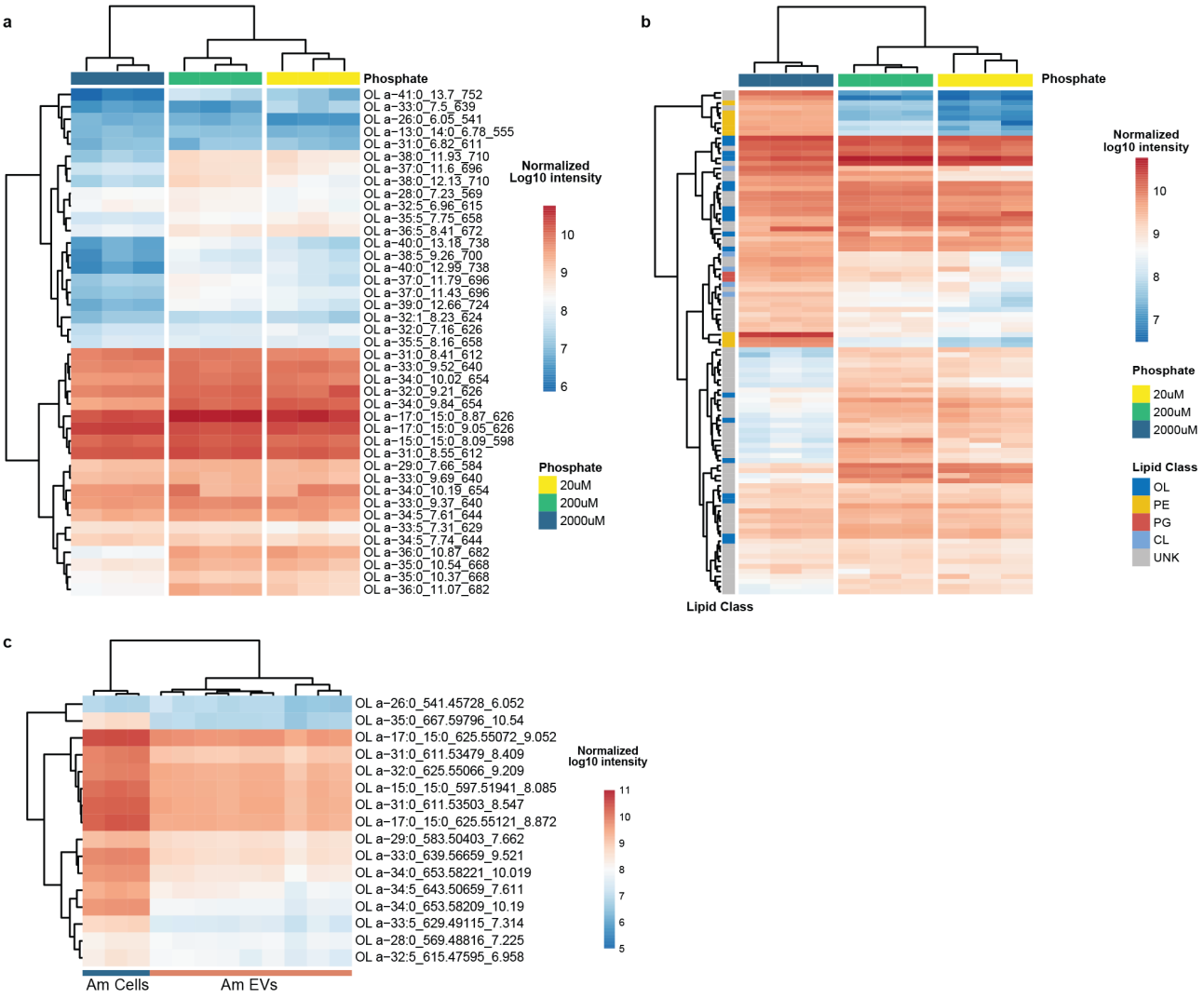
Supplementary Figure 2.3 DO gut microbiome QTL hotspot at Chr15: 61–65Mbp. Founder allele effects of KO and taxa trait QTL at Chr15 hotspot (LOD > 6).


a
Present
Absent
KO Group
129 allele positive effect
129 allele negative effect
Phylum
Actinobacteria
Bacteroidetes
Firmicutes
Tenericutes
Verrucomicrobia
dnclassifie
phylum
b
Metagenomic p_Bacteroidetes
k__Bacteria;p__Bacteroidetes
Metagenomic p_Firmicutes
k__Bacteria;p__Firmicutes
Bacteroidetes/Firmicutes ratio
Bacteroidetes/Firmicutes ratio
AJ B6 129 NOD NZO CAST PWK WSB
DO Founder Strain
AJ
B6
129
NOD
NZO
CAST
PWK
WSB
Predicted allele effects in founder strains
Observed abundance in founder strains
c
LOD score
rs582880514
rs31810445
Gsdmc
Gsdmc2
Gsdmc3
Gsdmc4
Pvt1
Myc
H2afy3
Asap1
Fam49b
Mouse Genome Chr15 (Mbp)

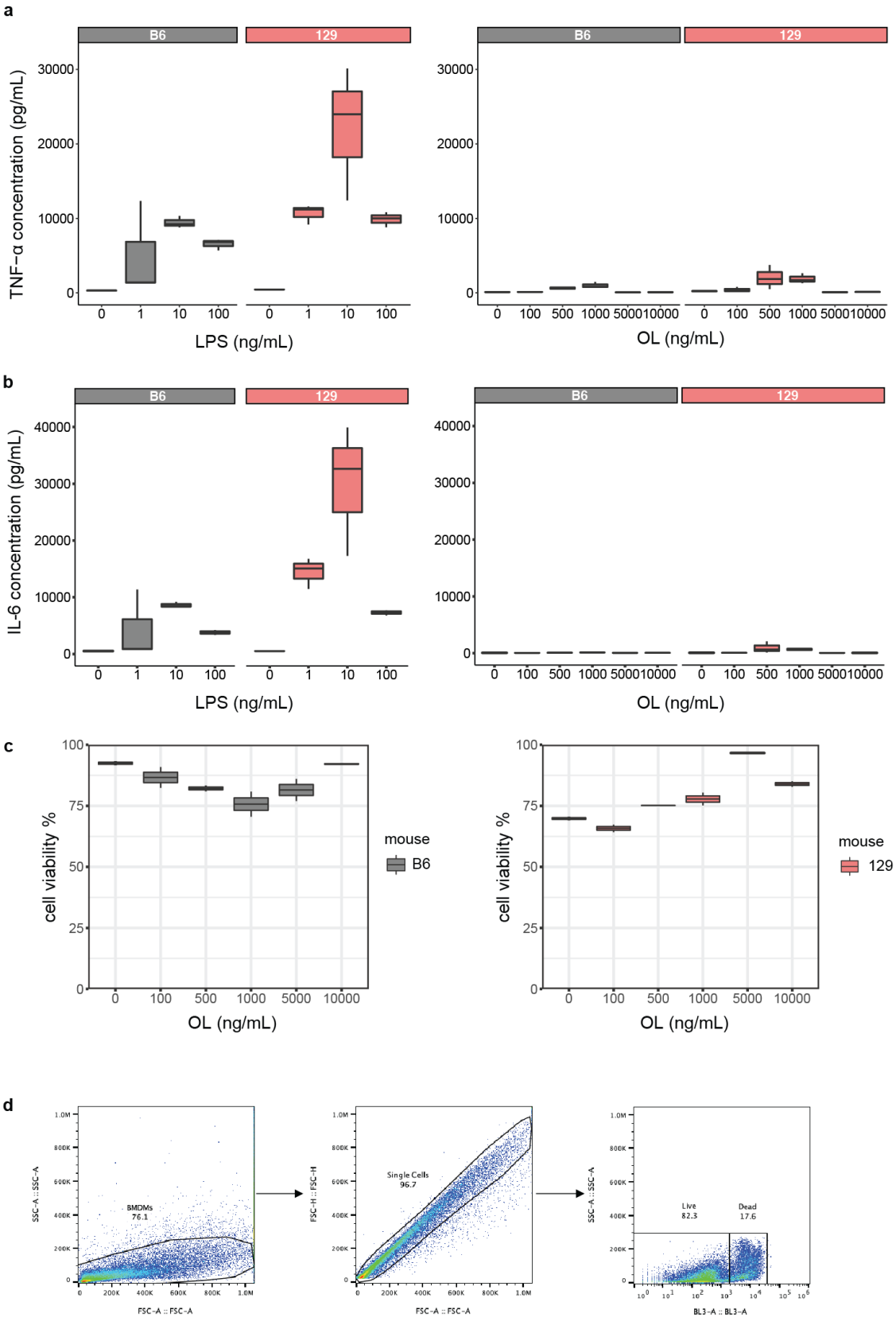
Supplementary Figure 2.4 DO gut microbiome QTL hotspot and SNP associations. **a**, Presence/absence of KOs that mapped to Chr15 hotspot across all MAGs. Sporulation functions were not detected in Bacteroidetes. **b**, Estimated founder allele effects for Bacteroidetes and Firmicutes, and Bacteroidetes/Firmicutes ratio (left panel). Observed abundance of Bacteroidetes Firmicutes and Bacteroidetes/Firmicutes ratio in founder strains as determined by Kemis et al. (right panel, n = 9-12 animals/founder strain). **c**, SNPs significantly associated with these traits in Chr15 hotspot include two intron SNPs in *Gsdmc* and *Gsdmc2* genes. Box and whisker plots denote the interquartile range, median and spread of points within 1.5 times the interquartile range, data beyond the end of the whiskers are plotted individually.


a
Lipid category
Correlation R
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
cluster1
cluster2
c3
cluster4
cluster5
Lipid category
Fatty.Acyl
Glycerolipid
Sphingolipid
Phospholipid
UNK
MAGs phylum
Actinobacteria
Bacteroidetes
Firmicutes
Tenericutes
Verrucomicrobia
unclassified
c1
cluster2
cluster3
cluster4
cluster5
cluster6
MAGs phylum
b
Phospholipid: PE
Sphingolipid: Cer [BS]
UNK: UNK
Fatty.Acyl: FA
Sphingolipid: Cer [NS]
Sphingolipid: Cer [NP]
Sphingolipid: Cer [AP]
Sphingolipid: SP
Sphingolipid: Cer [AS]
Sphingolipid: SHexCer
Glycerolipid: DG
Phospholipid: Plasmenyl-PC
Phospholipid: PC
Phospholipid: Plasmanyl-PE
Phospholipid: LysoPE
Phospholipid: PI
Phospholipid: PE
Fatty.Acyl: AC
Phospholipid: CL
Sphingolipid: Cer [AS]
Phospholipid: LysoPC
Sphingolipid: HexCer [AP]
Phospholipid: LysoPI
Phospholipid: Plasmenyl-PC
Glycerolipid: TG
Glycerolipid: MGDG
Phospholipid: PG
Phospholipid: PC
Phospholipid: Plasmanyl-PC
Phospholipid: Lyso-PE
cluster1
cluster2
cluster3
cluster4
cluster5
cluster6
Lipid category
Fatty.Acyl
Glycerolipid
Phospholipid
Sphingolipid
UNK
1
2
3
4
5
10
20
80
-log10(1/P adj-BH)

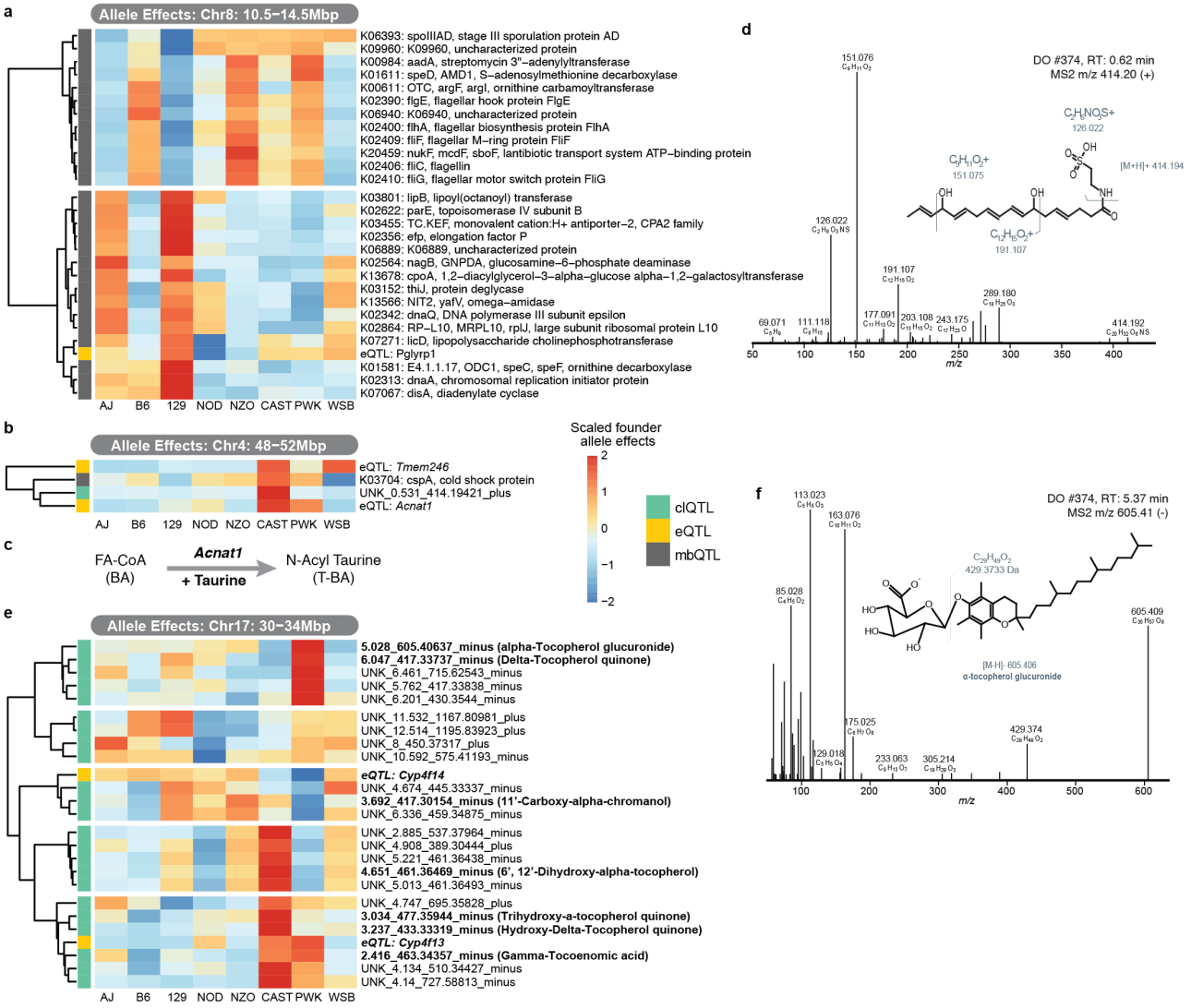
Supplementary Figure 2.5 Correlation between gut bacterial MAGs and caecal lipids.. **a**, Presence/absence of KOs that mapped to Chr15 hotspot across all MAGs. Sporulation functions were not detected in Bacteroidetes. **b**, Estimated founder allele effects for Bacteroidetes and Firmicutes, and Bacteroidetes/Firmicutes ratio (left panel). Observed abundance of Bacteroidetes Firmicutes and Bacteroidetes/Firmicutes ratio in founder strains as determined by Kemis et al. (right panel, n = 9-12 animals/founder strain). **c**, SNPs significantly associated with these traits in Chr15 hotspot include two intron SNPs in *Gsdmc* and *Gsdmc2* genes. Box and whisker plots denote the interquartile range, median and spread of points within 1.5 times the interquartile range, data beyond the end of the whiskers are plotted individually.



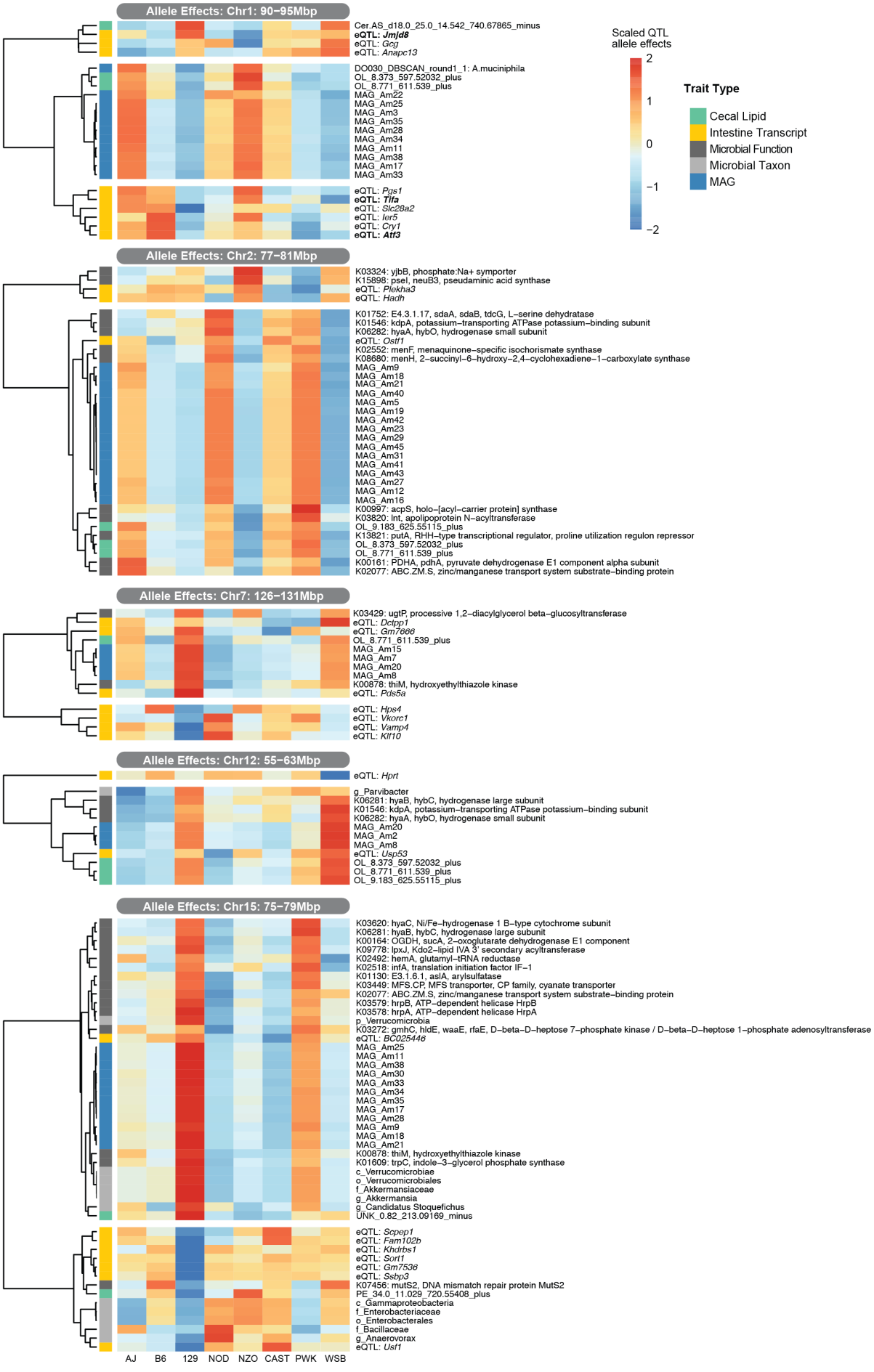
Supplementary Figure 2.6 Detection of ornithine lipids (OL) in *Akkermansia muciniphila*. **a**, Heatmap showing relative abundance of all OL species detected in cell pellets from *A. muciniphila* grown in vitro in defined media supplemented with different levels of phosphate: 20 µM, 200 µM and 2000µM. **b**, Relative abundance of lipid features detected in cell pellets from *A. muciniphila* grown in defined media with different levels of phosphate. Top 200 most abundant lipids features are shown. **c**, Relative abundance of OL features detected extracellular vesicles (AmEVs) purified from *A. muciniphila* grown in defined medium with the comparison to *A. muciniphila* cells.


a
B6
129
TNF−α concentration (pg/mL)
LPS (ng/mL)
OL (ng/mL)
b
IL-6 concentration (pg/mL)
LPS (ng/mL)
OL (ng/mL)
c
cell viability %
OL (ng/mL)
mouse
B6
129
d
BMDMs
76.1
Single Cells
96.7
Live
82.3
Dead
17.6
FSC-A :: FSC-A
SSC-A :: SSC-A
FSC-H :: FSC-H
BL3-A :: BL3-A

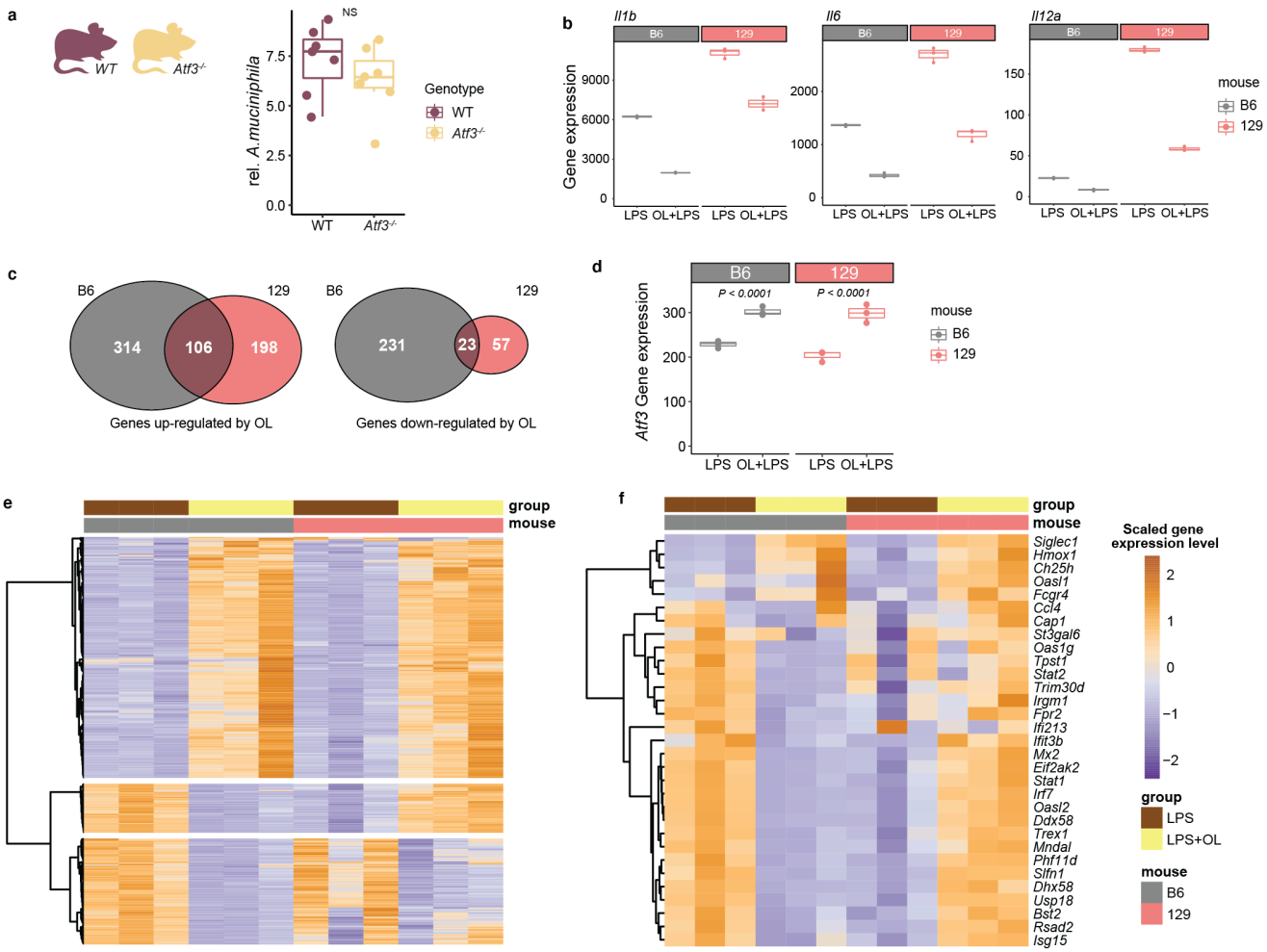
Supplementary Figure 2.7 Cytokine production by BMDM. **a**,**b**, (**a**) TNF-α and (**b**) IL-6 levels detected in supernatants from BMDM cells in B6 and 129 mice treated for six hours with different concentrations of LPS or OL. **c**, Cell viability of BMDM cells in B6 and 129 mice treated for six hours with 10 ng/mL LPS and different concentrations of OL. **d**, Flow cytometry gating strategy for BMDM cell viability assays. N = 3 biological replicates/treatment group. Box and whisker plots denote the interquartile range, median and spread of points within 1.5 times the interquartile range, data beyond the end of the whiskers are plotted individually.



Supplementary Figure 2.8 Examples of co-mapping QTL. **a**, At Chr8: 10.5-14.5 Mbp, co-mapping of gut bacterial lipopolysaccharide cholinephosphotransferase function with *Pglyrp1* eQTL was observed. **b**, At Chr4: 50 Mbp, co-mapping of an unidentified caecal feature and a local *Acnat1* eQTL was observed. **c**, The knowledge of *Acnat1* conjugating taurine to fatty acids guided the identification of the feature as an N-acyl taurine. **d**, Fragmentation pattern of identified N-acyl taurine. **e**, At Chr17: 30-34 Mbp, several unidentified features co-mapped which subsequently could be identified as tocopherols and exemplarily shown for the most significant feature alpha-tocopherol glucuronide. **f**, Fragmentation pattern of identified alpha-tocopherol glucuronide.


Allele Effects: Chr1: 90−95Mbp
Cer.AS_d18.0_25.0_14.542_740.67865_minus
eQTL: Jmjd8
eQTL: Gcg
eQTL: Anapc13
DO030_DBSCAN_round1_1: A.muciniphila
OL_8.373_597.52032_plus
OL_8.771_611.539_plus
MAG_Am22
MAG_Am25
MAG_Am3
MAG_Am35
MAG_Am28
MAG_Am34
MAG_Am11
MAG_Am38
MAG_Am17
MAG_Am33
eQTL: Pgs1
eQTL: Tlfa
eQTL: Slc28a2
eQTL: Ier5
eQTL: Cry1
eQTL: Atf3
Scaled QTL allele effects
2
1
0
−1
−2
Trait Type
Cecal Lipid
Intestine Transcript
Microbial Function
Microbial Taxon
MAG
Allele Effects: Chr2: 77−81Mbp
K03324: yjbB, phosphate:Na+ symporter
K15898: pseI, neuB3, pseudaminic acid synthase
eQTL: Plekha3
eQTL: Hadh
K01752: E4.3.1.17, sdaA, sdaB, tdcG, L−serine dehydratase
K01546: kdpA, potassium−transporting ATPase potassium−binding subunit
K06282: hyaA, hybO, hydrogenase small subunit
eQTL: Ostf1
K02552: menF, menaquinone−specific isochorismate synthase
K08680: menH, 2−succinyl−6−hydroxy−2,4−cyclohexadiene−1−carboxylate synthase
MAG_Am9
MAG_Am18
MAG_Am21
MAG_Am40
MAG_Am5
MAG_Am19
MAG_Am42
MAG_Am23
MAG_Am29
MAG_Am45
MAG_Am31
MAG_Am41
MAG_Am43
MAG_Am27
MAG_Am12
MAG_Am16
K00997: acpS, holo−[acyl−carrier protein] synthase
K03820: lnt, apolipoprotein N−acyltransferase
OL_9.183_625.55115_plus
K13821: putA, RHH−type transcriptional regulator, proline utilization regulon repressor
OL_8.373_597.52032_plus
OL_8.771_611.539_plus
K00161: PDHA, pdhA, pyruvate dehydrogenase E1 component alpha subunit
K02077: ABC.ZM.S, zinc/manganese transport system substrate−binding protein
Allele Effects: Chr7: 126−131Mbp
K03429: ugtP, processive 1,2−diacylglycerol beta−glucosyltransferase
eQTL: Dctpp1
eQTL: Gm7666
OL_8.771_611.539_plus
MAG_Am15
MAG_Am7
MAG_Am20
MAG_Am8
K00878: thiM, hydroxyethylthiazole kinase
eQTL: Pds5a
eQTL: Hps4
eQTL: Vkorc1
eQTL: Vamp4
eQTL: Klf10
Allele Effects: Chr12: 55−63Mbp
eQTL: Hprt
g_Parvibacter
K06281: hyaB, hybC, hydrogenase large subunit
K01546: kdpA, potassium−transporting ATPase potassium−binding subunit
K06282: hyaA, hybO, hydrogenase small subunit
MAG_Am20
MAG_Am2
MAG_Am8
eQTL: Usp53
OL_8.373_597.52032_plus
OL_8.771_611.539_plus
OL_9.183_625.55115_plus
Allele Effects: Chr15: 75−79Mbp
K03620: hyaC, Ni/Fe−hydrogenase 1 B−type cytochrome subunit
K06281: hyaB, hybC, hydrogenase large subunit
K00164: OGDH, sucA, 2−oxoglutarate dehydrogenase E1 component
K09778: lpxJ, Kdo2−lipid IVA 3' secondary acyltransferase
K02492: hemA, glutamyl−tRNA reductase
K02518: infA, translation initiation factor IF−1
K01130: E3.1.6.1, aslA, arylsulfatase
K03449: MFS.CP, MFS transporter, CP family, cyanate transporter
K02077: ABC.ZM.S, zinc/manganese transport system substrate−binding protein
K03579: hrpB, ATP−dependent helicase HrpB
K03578: hrpA, ATP−dependent helicase HrpA
p_Verrucomicrobia
K03272: gmhC, hldE, waaE, rfaE, D−beta−D−heptose 7−phosphate kinase / D−beta−D−heptose 1−phosphate adenosyltransferase
eQTL: BC025446
MAG_Am25
MAG_Am11
MAG_Am38
MAG_Am30
MAG_Am33
MAG_Am34
MAG_Am35
MAG_Am17
MAG_Am28
MAG_Am9
MAG_Am18
MAG_Am21
K00878: thiM, hydroxyethylthiazole kinase
K01609: trpC, indole−3−glycerol phosphate synthase
c_Verrucomicrobiae
o_Verrucomicrobiales
f_Akkermansiaceae
g_Akkermansia
g_Candidatus Stoquefichus
UNK_0.82_213.09169_minus
eQTL: Scpep1
eQTL: Fam102b
eQTL: Khdrbs1
eQTL: Sort1
eQTL: Gm7536
eQTL: Ssbp3
K07456: mutS2, DNA mismatch repair protein MutS2
PE_34.0_11.029_720.55408_plus
c_Gammaproteobacteria
f_Enterobacteriaceae
o_Enterobacterales
f_Bacillaceae
g_Anaerovorax
eQTL: Usf1
AJ
B6
129
NOD
NZO
CAST
PWK
WSB

Supplementary Figure 2.9 Founder allele effects on co-mapping traits associated with *A. muciniphila* levels. *A. muciniphila*, caecal OL and eQTL genes co-mapping at Chr1: 90-95 Mbp, Chr2: 77-81 Mbp, Chr7: 126-131 Mbp, Chr12: 55-63 Mbp and Chr15: 75-79 Mbp.



Supplementary Figure 2.10 Expression of genes in BMDM treated with OL/LPS. **a**, Abundance of *A. muciniphila* in faecal pellets from *Atf3*$^{-/-}$ mice and WT mice (n = 7 mice/genotype; four females, three males for both genotypes). **b**, Gene expression level of *Il1b*, *Il6* and *Il12a* from BMDM cells derived from B6 and 129 mice treated for six hours with LPS (10 ng/ml) or with LPS (10 ng/mL) and OL (1 µg/mL). N = 3 biological replicates/treatment group. **c**, Number of differentially expressed genes in BMDM derived from B6 and 129 mice. **d**, Gene expression levels of *Atf3* in BMDM from B6 and 129 mice treated for six hours with LPS (10 ng/mL) or LPS (10 ng/mL) and OL (1 µg/mL). N = 3 biological replicates/genotype/treatment group. **e**, Differentially expressed genes in BMDM from B6 and 129 mice. **f**, Previously reported ATF3 regulated genes in BMDM. Impact of OL on these genes in B6 and 129 mice. Box and whisker plots denote the interquartile range, median and spread of points within 1.5 times the

interquartile range; data beyond the end of the whiskers are plotted individually. Statistical difference between treatment groups was tested by two-sided Welch's *t*- test.

CHAPTER 3: Host genetics effects on enterotype-associated gut microbiome in mice

Zhang Q[*], *Hutchison ER, Pan C, Lusis AJ, Rey FE*[+]

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

3.1 Abstract

Host genetic variation has been associated with gut microbiome composition differences in humans and mice, including several large scale human cohorts. However, we lack a molecular understanding of how host genetics influences microbial community and metabolic niches. We leveraged 90 inbred hyperlipidemic mouse strains from the Hybrid Mouse Diversity Panel (HMDP). Metagenomic analysis of fecal DNA followed by genome-wide association analysis identified genomic loci that were associated with microbial enterotypes in the gut. We discovered genetic variants of *Amy1* gene that were associated with abundance of Firmicutes (*Lachnospiraceae* family) and Bacteroidetes (*Muribaculaceae* family) taxa that encode distinct starch and sugar metabolism functions. Host amylase activity impacts availability of carbohydrates to the host and potentially to gut bacteria. The genetic variants described above were associated with distinct gut microbial communities (enterotypes) with different metabolic capacity. Mendelian randomization revealed that host physiology phenotypes, including liver fibrosis and plasma HDL-cholesterol levels were causally associated with gut microbiome enterotypes. This work provides evidence for a causal relationship between host genetic variation, gut microbial enterotypes and host physiology phenotypes in mice.

3.2 Introduction

The microbial communities that inhabit mammals have profound effects on host biology and health. Alterations in the intestinal microbiome have been associated with myriad conditions including cardiovascular disease and metabolic disorders (Lynch & Pedersen, 2016; Wang et al., 2011). Emerging evidence showed the influences of gut bacteria on host metabolism and immunity (Belkaid & Hand, 2014; Tremaroli & Bäckhed, 2012). A range of intrinsic factors, including host genetics, and environmental factors, including diet, modulate the gut microbiome and subsequently affect its interactions with the host (Gacesa et al., 2022; Kurilshikov et al., 2021). Over the years several groups have identified distinct gut microbial communities that reflect stratification in a population and define the metabolic niche and functional characteristics of the gut microbiome. These well-defined and frequently recurring microbial communities were termed enterotypes (Costea et al., 2017; MetaHIT Consortium (additional members) et al., 2011). However, it remains largely unknown how host genetics modulate microbial enterotypes as well as enterotype associated functions and metabolic pathways in the gut.

Carbohydrates are important energy source for both human and microbial cells. Dietary compounds that cannot be digested by human enzymes, including plant polysaccharides and other carbohydrates such as resistant starches, reach the distal gut where they are broken down and fermented by resident bacteria and serve as energy source for these gut microbes (Oliphant & Allen-Vercoe, 2019). However, it is also known that digestible carbohydrates such as starch and sugars can influence the gut microbiota and its associations with the host (Hutchison et al., 2023; Murga-Garrido et al., 2021). Thus, digestible carbohydrates can also be accessed by gut microbes as the ability of different hosts to access these carbohydrates is variable. Human amylase facilitate starch digestion (Perry et al., 2007). Recent work revealed that humans with lower amylase gene copy number harbor gut microbiomes with a greater capacity for breakdown of complex carbohydrates (Poole et al., 2019). This work suggested that genetic variation in host amylase gene may potentially impact gut microbiome and its subsequential effects on the host.

In addition to the study mentioned above, several mouse and human studies have examined the role of host genetics in shaping the composition of gut microbiota. These efforts mostly applied 16S rRNA gene sequencing (Kemis et al., 2019; Org et al., 2015; Sanna, 2022). Shotgun metagenomics allows comprehensive profiling of the functions and metabolic pathways present in gut communities. However, there is lack of metagenomic characterization of gut microbiome in large-scale genetically diverse phenotypically characterized cohorts. A valuable resource for such genetically diverse cohort is the Hybrid Mouse Diversity Panel (HMDP), which consists of over 100 common inbred strains. Recently, each strain from this panel was made hyperlipidemic by transgenic expression of human apolipoprotein E-Leiden (APOE-Leiden) and human cholesteryl ester transfer protein (CETP). This was done by breeding each of the HMDP strains to a common strain (C57BL/6J) that donated these two dyslipidemia-inducing transgenes. Genetic differences among these animals arise only from sequence variations present in the individual recipient strains. This set of F1 animals termed Ath-HMDP mice (Bennett et al., 2015). Here, we used metagenomics to characterize the gut microbiome from 90 Ath-HMDP strains fed high-fat supplemented with 1% cholesterol. Our analysis identified three microbial enterotypes, dominated by Firmicutes, Bacteroidetes and Verrucomicrobiota, in this mouse cohort and identified two host genomic loci that are significantly associated with the enterotypes. Our results suggest that genetic variance, potentially in host amylase genes, could be a causal determinant of gut microbial enterotypes by selecting specific carbohydrate metabolizing bacterial species. Furthermore, our results also suggest that the functional niches of the selected enterotypes subsequently influence host biomarkers including plasma cholesterol and triglycerides and disease phenotypes including liver fibrosis.

3.3 Results

Characterization of the gut metagenomes from 90 Ath-HMDP mouse strains

We generated metagenomic datasets from cecal DNA collected from 355 F1 Ath-HMDP mice encompassing 90 strains (2-6 mice/sex/strain, 188 male and 161 female) fed high-fat with 1% cholesterol containing diet for 16 weeks (37.8 million paired-end reads/sample). Generation of this mouse cohort has been previously described (Bennett et al., 2015). These animals developed atherosclerosis and liver fibrosis which varies widely as a function of strain (Bennett et al., 2015; Hui et al., 2018), additional clinical data is also available for these animals (**Supplementary Table 1**).

Phylogenetic and functional analyses identified 461 bacterial taxa (7 phyla, 43 classes, 49 orders, 63 families, 131 genera and 166 species), 2,127 microbial functions (KEGG orthology database) and 294 metabolic pathways (MetaCyc database) across all mice. Gut microbiota composition was highly variable across the 90 strains; for example, the relative abundance of Firmicutes (Bacillota) phylum ranged from 10% to 55% whereas the relative abundance of Bacteroidetes (Bacteroidota) ranged from 5% to 77% (**Supplementary Figure 3.1a**). The most abundant species (> 0.5%) that were present in at least 90% of mice are *Muribaculaceae* family (*Paramuribaculum intestinale*, *Muribaculum intestinale*, *Muribaculum gordoncarteri, Duncaniella muris, Duncaniella freteri*), *Alistipes finegoldii*, *Parabacteroides goldsteinii*, *Faecalibaculum rodentium*, *Lachnospiraceae bacterium*, *Bilophila sp.* and *Parasutterella sp* (**Supplementary Figure 3.1b**). For the purpose of our analyses, we define these taxa as the “core microbiome species” of Ath-HMDP mice.

Enterotypes and associated bacterial species

Principal coordinate analysis (PCoA) of the microbial communities described above using species abundance data resulted in three clusters each dominated by a different phylum: (i) Firmicutes, (ii) Bacteroidetes and (iii) Verrucomicrobiota respectively (**Supplementary Figure**

3.2a). Using microbial function abundance resulted in similar clustering each dominated by one these three phyla (**Supplementary Figure 3.2b**). The first principal component (PC1) of the microbial functions, which explains the most variance for the functional profiles, was highly correlated with several taxa, including the *Lachnospiraceae* family (Spearman's $\rho = 0.86$, $P = 5.7 \times 10^{-99}$) and the *Desulfovibrionaceae* family (Spearman's $\rho = 0.72$, $P = 1.9 \times 10^{-53}$) (**Supplementary Table 2**). Because we observed gradient stratification in taxonomic composition and functions in the gut microbiome data, we next detected three enterotypes using partitioning around medoid (PAM) clustering of Bray-Curtis distance of species abundance. Each enterotype is identifiable by the levels of Firmicutes, Bacteroidetes and Verrucomicrobiota respectively (**Figure 3.1a**). Previous studies reported distinct enterotypes dominated by Bacteroides, Prevotella and Ruminococcaceae in human gut. We observed clusters dominated by different taxa as compared to humans, which is likely due to the distinct overall gut microbiota composition detected in these two mammals.

To identify microbial species that shape the enterotypes, we summarized the top abundant species (relative abundance > 0.1% and present at least > 20% mice) into co-abundance groups (CAGs) by co-occurrence scores (**Figure 3.1b**). The Bacteroidetes-dominated CAG contained mostly Bacteroidetes taxa including *Muribaculaceae*, the most abundant family in mouse gut (Lkhagva et al., 2021), (*Duncaniella muris, Duncaniella freteri*, *Duncaniella dubosii*, *Muribaculaceae CAG-495 sp., Muribaculaceae CAG-873 sp.*), *Bacteroides* genus (*Bacteroides acidifaciens*, *Bacteroides sp.*) and *Alistipes sp*. The Firmicutes-dominated CAG consisted mostly of Firmicutes taxa including *Lachnospiraceae* family (*Roseburia sp., Dorea sp., Acetatifactor sp., Sporofaciens sp., Kineothrix sp.* and *Schaedlerella arabinosiphila*), *Oscillospiraceae* family (*Lawsonibacter sp., Enterenecus sp.* and *Pelethomonas sp.*) and Proteobacteria including *Bilophila sp.*

Associations between gut microbiome features and host genetics

To identify associations between mouse genomic variation and gut bacterial features, we performed genome-wide association analysis and mapped the abundance of bacterial functions, taxa and metabolic pathways to the mouse genome (**Figure 3.2a**). We observed 646 functions, 138 taxa and 109 pathways that were associated with at least one significant locus by genome-wide significance cutoff ($P < 4 \times 10^{-6}$). We further examined the density of these significantly associated gut microbial traits over the whole mouse genome and identified a GWAS hotspot on chromosome 3 at 113-115 Mbp; this genomic locus was associated with 347 different functions. Pathway enrichment analysis using Fisher's exact test revealed that genes encoding for bacterial flagellar assembly, bacterial chemotaxis and bacterial motility proteins were significantly overrepresented among the functional traits mapping to this GWAS hotspot (**Figure 3.2b**).

The genomic regions identified by GWAS most likely contain candidate gene/genes that are in strong linkage disequilibrium (LD) with the peak SNP. Nine protein coding genes are in the LD region (determined by correlation r^2 with peak SNP > 0.8) at chromosome 3 GWAS hotspot, including mouse amylase cluster genes (*Amy1*, *Amy2a1*, *Amy2a2*, *Amy2a3*, *Amy2a4* and *Amy2a5*), *Rnpc3*, *Col11a1* and *Olfm3* (**Figure 3.2c**). Gene expression data were available from a subset (n = 4-8 mice/ strain; 40 strains) of the same Ath-HMDP mice (Bennett et al., 2015). We searched for candidate genes by correlation between functions' PC1 and these candidate genes and found a significantly correlation for *Amy1* gene (**Supplementary Figure 3.3a,b**). In humans, the amylase gene copy number is associated with nearby SNPs, structural haplotypes of the amylase locus (Usher et al., 2015), and gut microbiome composition (Poole et al., 2019). In addition, a previous study showed that individuals with higher amylase gene copy number had higher levels of Firmicutes, the primary phylum containing flagellated bacteria in the human gut (Poole et al., 2019). Our data suggested that genomic variants of the amylase gene locus in mice might be associated with abundance of bacterial functions.

Because of multiple functional traits mapped to this hotspot locus, we also found that the PC1 of function abundance mapped to the same locus as expected (**Figure 3.3a**). PC1 explained the most variance of microbial functions, which are also clustered by the three enterotypes (**Supplementary Figure 3.1b**). The lead SNP rs31001780 has two alleles, A and G, which are associated with the enterotypes (**Figure 3.3b**). More specifically, mice with the allele A at SNP rs31001780 had higher Firmicutes and lower Bacteroidetes levels in the gut (**Figure 3.3c,d**). The other genetic locus associated with the enterotypes was at Chr1 (**Figure 3.3e**). This locus has the lead SNP rs31965376 with two alleles, A and T, which are associated with *Akkermansia* levels in the gut (**Figure 3.3f-h**).

Co-abundance groups (CAGs) are associated with bacterial carbohydrate metabolism

To explore potential bacterial functions and pathways that were modulated by enterotype associated SNPs. We examine the metabolic niche differences between the two identified CAGs, we first compared correlations between functions from starch and sugar metabolism pathway with the species abundance from two CAGs (**Supplementary Figure 3.4**). Glycoside hydrolases are enzymes that catalyze the hydrolysis of the glycosidic linkage in sugars. Glucosidases that breakdown oligo- and disaccharides, such as amylosucrase (K05341, predicted to degrade sucrose), β−glucosidase (K05350), β−fructofuranosidase (K01193) and oligo−1,6−glucosidase (K01182), were associated with Firmicutes species. α−amylase (K07405 and K01176, predicted to degrade starch), endoglucanase (K01179, predicted to degrade cellulose), dextranase (K05988, predicted to degrade dextran) and pullulanase (K01200, predicted to degrade pullulan), were associated with Bacteroidetes species. In addition, bacteria species from Firmicutes CAG were positively correlated with sugar transport system and phosphotransferase system (PTS).

Higher mouse amylase gene expression was associated with higher gut Firmicutes levels (*Lachnospiraceae* family) and lower Bacteroidetes levels (*Muribaculaceae* family) representing

two enterotypes. We reasoned that the variant of host amylase genes could lead to different carbohydrates available in the gut for microbes. This in turn would lead to distinct gut microbial community (enterotype) that has different metabolic capacity. To test this hypothesis, we characterized gut microbiome by shotgun metagenomic sequencing from three mouse strains with a variety of amylase gene copy number (CN): C57BL/6J (B6) has high amylase CN (*Amy1*, *Amy2a1*, *Amy2a2*, *Amy2a3*, *Amy2a4*, *Amy2a5*, *Amy2b*, *Amy2-ps1*); NZO/HLtJ (NZO) has medium amylase CN (*Amy1*, *Amy2a1*, *Amy2a2*, *Amy2b*, *Amy2-ps1*); CAST/EiJ (CAST) has low amylase CN (*Amy1*, *Amy2a2*, *Amy2-ps1*). These mice were fed a high carbohydrate diets for 16 weeks. We found the abundance of *Muribaculaceae* family is significantly higher in CAST mouse compared to those in B6 and NZO (**Figure 3.4a**) and the bacteria alpha-amylase enzyme (K07405) are higher in CAST compared to those in B6 and NZO (**Figure 3.4b**). Furthermore, recent studies showed that *Muribaculaceae*, who have the starch utilization genes, were dramatically more abundant in acarbose-treated mice (Smith et al., 2019, 2021), Acarbose inhibits α-amylase activity in the small intestine, which resulted in increased starch availability in the lower digestive tract. These results suggest that lower copy number of amylase genes from mouse genome resulted in increased *Muribaculaceae* abundance in the gut. These results also support the notion that genetic variation is amylase gene region is causal for enterotype associated variants at Chromosome 3 in mice.

Enterotype species are associated with host physiology traits

We next explored the associations between CAG species with previously reported host cardiometabolic phenotypes for these mice. Species from the two CAGs discussed above showed distinct associations (**Figure 3.5a**). Atherosclerotic lesion size was positively associated with *Bacteroides sp.* (Spearman's $\rho = 0.35$, $P = 4.1 \times 10^{-6}$) and *Bilophila sp.* (Spearman's $\rho = 0.17$, $P = 4.4 \times 10^{-2}$) and negatively correlated with *Roseburia sp.* (Spearman's $\rho = -0.31$, $P = 5.6 \times 10^{-5}$)

as reported in a previous study (Kasahara et al., 2018). We further identified positive associations for *Turicimonas muris* (Spearman's $\rho = 0.18$, $P = 0.03$), *Atopobiaceae bacterium sp.* (Spearman's $\rho = 0.2$, $P = 0.015$) and negative associations for *Dorea sp.* (Spearman's $\rho = -0.29$, $P = 2 \times 10^{-4}$) and *Enterenecus sp.* (Spearman's $\rho = -0.16$, $P = 0.05$). Plasma levels of HDL and LDL/VLDL were positively associated with bacteria from Firmicutes CAG and negatively associated with bacteria from Bacteroidetes CAG. Liver fibrosis was negatively associated with Bacteroidetes CAG species such as *Bacteroides sp.* and *Muribaculaceae bacterium* and positively associated with Firmicutes CAG species such as *Sporofaciens sp*., *Enterenecus sp*. And *Kineothrix sp*. These results indicated the host physiology phenotypes are highly associated with gut microbiome enterotypes (Firmicutes and Bacteroidetes CAG) in the mice.

Gut bacterial flagella is causally associated with host phenotypes

We further examined whether other gut microbial function abundances associated with enterotypes with a focus on bacterial flagellar genes as they were overrepresented among the functions that mapped at Chr3 locus and the noteworthy characteristic that distinguish Firmicutes and Bacteroidetes CAGs (**Supplementary Figure 3.5**). We observed a high correlation between *fliC* abundance (K02406; encoding flagellar filament structural protein) with LDL+VLDL cholesterol (Spearman's $\rho = 0.31$, $P = 5.8 \times 10^{-10}$), HDL cholesterol (Spearman's $\rho = 0.17$, $P = 1.8 \times 10^{-3}$), liver fibrosis area (Spearman's $\rho = 0.17$, $P = 2.5 \times 10^{-3}$), liver cholesterol levels (Spearman's $\rho = 0.24$, $P = 4.1 \times 10^{-5}$) and liver triglycerides (Spearman's $\rho = -0.27$, $P = 4.3 \times 10^{-6}$) (**Supplementary Figure 3.6a**). We also detected correlation between gut bacterial *fliC* abundance with aortic lesion area size in male mice (Spearman's $\rho = 0.2$, $P = 0.052$). Structural variations in flagellin genes determine the recognition by host Toll-like receptor 5 (TLR5) and antibody responses (Bourgonje et al., 2023; Clasen et al., 2023). We examined the most abundant (relative abundance > 0.01%) flagellin genes, from Lachnospiraceae (*Roseburia*, *Dorea*)

and Desulfovibrionaceae (*Desulfovibrio*) family, and their individual associations with the host phenotypes (**Supplementary Figure 3.6b**). We found the conserved N-terminal and C-terminal motifs from flagellin genes which are important for TLR5 binding and activation, but the amino acid variations in these motifs may determine the different associations with the host phenotypes.

We next performed bi-directional Mendelian randomization (MR) to assess whether gut bacterial flagellar causally contributes to host traits. We focused on gut bacterial flagellin encoding gene *fliC* abundance, as this trait was associated with four independent genetic variants with $P < 1 \times 10^{-4}$, which were used as instrument variables in MR. We carried out inverse variance weighted (IVW) test using *fliC* abundance as exposure and nine clinical traits as outcome. We observed significant causal relationships between gut bacterial *fliC* with liver fibrosed area ($P = 7.3 \times 10^{-3}$) and HDL cholesterol ($P = 4.6 \times 10^{-4}$) (**Figure 3.5b**). When we tested MR considering clinical traits as exposure and *fliC* abundance as outcome, we didn't observe significant causal effects (**Figure 3.5c**).

3.4 Discussion

Genome-wide association studies have identified multiple host genomic loci associated with gut microbiome in humans and mice, however most of these efforts have many of which focused on microbial organismal composition and there is lack of evidence linking specific functions and pathways with host genetic variation. Additionally, gut microbiome enterotypes were described in human cohorts where a small number of bacterial taxa determines the stratification of whole community, but no such study conducted in large genetically diverse mouse cohort so far. Our work comprehensively characterized gut microbiome composition, functions, and metabolic pathways in the Ath-HMDP mouse cohort. We identified three enterotypes dominated by different phyla including Firmicutes, Bacteroidetes and Verrucomicrobiota. We further found that the enterotypes were associated with bacterial taxa (Firmicutes and Bacteroidetes CAGs) and microbial functions (starch and sugar metabolism and flagellar assembly). The Bacteroidetes CAG included many bacteria from *Muribaculaceae*, the most abundant family in mouse gut (Lkhagva et al., 2021).

Using genetic mapping, we identified host genomic loci associated with bacterial taxa, functions and pathways, where two of these loci were associated with enterotypes. The genetic variant rs31001780 (A/G) at Chr3 locus was significantly associated with Firmicutes and Bacteroidetes enterotypes and genetic variant rs31965376 (A/T) at Chr1 locus was significantly associated with the Verrucomicrobiota enterotype. Importantly, expression level of *Amy1* gene, which spans in LD region of Chr3 locus, was correlated with Firmicutes, Bacteroidetes and Verrucomicrobiota. Given the same carbohydrate rich diets, mouse genetic variants of *Amy1* gene may induce different carbohydrate accessibility for gut microbiota and shape the different enterotypes. In humans, the amylase gene copy number determines the abundance of gut and oral microbiome; individuals with higher amylase gene copy number show higher levels of many genera within the Firmicutes in the gut (Poole et al., 2019), which aligns with our results in in mice; i.e., *Amy1* gene expression is positively associated with Firmicutes abundance. Bacteria in this

enterotype harbors different metabolic capacity. We reason that the differences in sugar and starch metabolism between Firmicutes and Bacteroidetes CAG species, may explain how genetic variants of *Amy1* gene shapes the enterotypes in mouse gut. Previous work showed that acarbose-treated mice show increased abundance of the *Muribaculaceae* family (Smith et al., 2019, 2021). Analyses of genome content revealed the starch utilization genes from *Muribaculaceae* MAGs (Smith et al., 2021). We found *Muribaculaceae* were more abundant in low amylase gene copy number mouse CAST. Although loss of function mouse experiment (*Amy1*) should be conducted to eliminate the effects of other genetic variants to gut microbiome in the future.

We also found the enterotype associated bacteria species correlated with host cardiometabolic phenotypes, especially the bacterial flagellin causally associated with increased liver fibrosed area and HDL cholesterol levels. Previous studies showed that gut microbiome partially explained the variations of plasma triglyceride and HDL cholesterol levels in human (Fu et al., 2015). Another study showed that the high-fat diet increased flagellated bacteria in the gut, which increased apolipoprotein A1 (ApoA1) production and HDL cholesterol levels in mice (Yiu et al., 2020). Interestingly, we also observed gut Firmicutes levels and bacterial *fliC* abundance were significantly positively correlated with ApoA1 gene expression levels in liver. Mendelian randomization (MR) seeks to find causal effects between phenotypes. Successful applications of MR in humans revealed the causal relationships between gut microbiome and other molecular traits, including blood metabolites (Liu et al., 2020), short-chain fatty acids (Sanna et al., 2019), and host metabolic traits (Qin et al., 2020; Rühlemann et al., 2021). To our best knowledges, our study is the first MR application of gut microbiome in genetically diverse mouse cohort. Our MR results confirmed the casual relationship between gut flagellated bacteria and plasma HDL cholesterol levels. We further reasoned that not only high-fat diet can increase the flagellated bacteria in the gut, but the amylase gene copy number can also affect flagellated bacteria abundance. A recent study showed bacteria flagellin gene variants from Lachnospiraceae family

were associated with TLR5 activation. We also found that nD1 TLR5 epitope motif in fliC gene were associated differently with host physiology phenotypes, including atherosclerotic lesion. This indicated the importance of bacterial genetic variations in gut microbiome association studies. A recent study found bacterial SNPs in human gut microbiome were associated with host BMI (Zahavi et al., 2023). Future investigation of bacterial SNPs in the mouse gut microbiome, especially in genetically diverse cohort such as HMDP, is required for a better understanding of their associations with the host.

Together, our work highlights how host genetics can shape different microbial enterotypes in mouse gut and identifies a potential candidate host gene involved. The microbial enterotype associated functions and pathways are the consequence of host genetic variants and are also the causes of host cardiometabolic phenotype variations, making them pivotal during host-gut microbiome interactions.

3.5 Methods

Mouse cohort

Male and female mice from the F1 Ath-HMDP panel are maintained in a temperature and humidity controlled environment under a 12 h light/dark cycle (lights on at 6:00 and off at 18:00). Mice were housed by strain. All mice were fed a high fat diet (33 kcal % fat from cocoa butter) supplemented with 1% cholesterol (Research Diets D10042101) for 16 weeks. Cecum contents were collected after animals were euthanized. Clinical phenotypes for this cohort was described in a previous study (Bennett et al., 2015).

Metagenomic DNA sequencing

Cecal DNA was extracted from individual mice using the PowerSoil DNA Isolation Kit. DNA concentration was verified using the Qubit® dsDNA HS Assay Kit (Life Technologies, Grand Island, NY). Samples were prepared using Illumina NexteraXT library preparation kit. Quality and quantity of the finished libraries were assessed using an Agilent bioanalyzer and Qubit® dsDNA HS Assay Kit, respectively. Libraries were standardized to 2nM. Paired end, 150 bp sequencing was performed using the Illumina NovaSeq6000. Images were analyzed using the standard Illumina Pipeline, version 1.8.2.

Profiling microbiome taxon

Gut microbial taxon was profiled by MetaPhlAn4 pipeline (ver 4.0.2) using the MetaPhlAn database (mpa_vOct22) and the ChocoPhlAn pan-genome database (mpa_vOct22_CHOCOPhlAnSGB_202212) that contain a collection of around 1 million prokaryotic metagenome-assembled genomes (Blanco-Míguez et al., 2023). The taxonomy clades with average relative abundance > 0.01% and present in at least >20% samples are kept as microbial taxon for downstream analyses. The unclassified SBG taxa were further annotated

to Genome Taxonomy Database (GTDB) using mpa_vOct22_CHOCOPhlAnSGB_202212_SGB2GTDB.tsv data from MetaPhlan4 pipeline.

Profiling microbiome function

Raw reads quality control was performed using Trimmomatic (ver. 0.39) with default parameters. To identify and eliminate host sequences, reads were aligned against the mouse genome (mm10/GRCm38) using Bowtie2 (ver. 2.3.4) with default settings and microbial DNA reads that did not align to the mouse genome were identified using samtools (ver. 1.3; samtools view -b -f 4 -f 8). Samples with total read depth <10 million were excluded for downstream analyses. Quantification of microbial genes was done by aligning clean paired end reads to a previous published mouse gut microbiome non-redundant gene catalog (Zhang et al., 2023) using Bowtie2 (ver. 2.3.4) and default parameters. RSEM (ver. 1.3.1) was used to estimate microbial gene abundance. Relative abundance of microbial gene counts per million (CPM) were calculated using microbial gene expected counts divided by gene effective length then normalized by the total sum. To obtain abundance information for microbial functions, CPM of genes with the same KEGG orthologous (KO) annotation were summed together. In case there were multiple KO annotations for a single gene, we used all KO annotations.

Microbiome pathway profiling

Gut microbial pathways were profiled using the HUMAnN3 pipeline (ver 3.0.0), the MetaPhlAn database (mpa_v20_m200), the ChocoPhlAn pan-genome database (v296_v201901b) and the UniRef90 protein database (ver 0.1.1) (Beghini et al., 2021). Pathways with average relative abundance > 0.01% detected in at least >20% samples were used for downstream analyses.

Enterotype clustering

Gut microbiome data was clustered using partitioning around medoid (PAM) clustering via pam() function from R package cluster (ver. 2.1.2). The Bray-Curtis distance of species abundance was used for PAM clustering. The dominant phylum in each cluster was determined by highest abundant taxon comparing to other clusters.

Microbial co-abundance groups

Similarity scores between species were calculated using CCREPE (compositionality corrected by renormalization and permutation) package (ver. 1.1.3) (Gevers et al., 2014). The species network was visualized using ggnet2() function from R package ggnet (ver. 0.1.0).

Genome-wide association of gut microbiome

The Mouse Diversity Genotyping Array was used for genotyping and gave approximately 450,000 SNPs. SNPs used for each trait were filtered by the following criteria: the minor allele frequency (MAF) > 5% and missing genotype frequency < 10%. GWAS analyses was performed using FaST-LMM (Lippert et al., 2011) (Python ver. 3.7.4). When testing SNPs on chromosome N, all SNPs from other chromosome besides N were used for kinship matrix construction, that is leave out one chromosome (LOOC) approach. Sex was used as covariance in the regression model. GWAS significant thresholds were determined by permutation tests. The significant of a genome-wild association has threshold of $P < 4 \times 10^{-6}$. We defined a study-wide significance threshold of $P < 4 \times 10^{-6} / (2127+108+300) = 1.58 \times 10^{-9}$.

Heritability estimation

Broad sense heritability for each trait was estimated using “repeatability()” function from “heritability” (ver. 1.3) R package. Narrow sense heritability for each trait was calculated using all filtered SNPs to estimate the proportion to explain total variations for each trait.

Data and statistical analysis

All data integration and statistical analysis were performed in R (v3.6.3). Differences between groups were evaluated using unpaired two-tailed Welch's t-test. Enrichment analysis was performed with Fisher's exact test using a custom R function. Correlation analysis was performed with two-sided Spearman's correlation using the R function 'cor.test()'. For multiple testing, Benjamini-Hochberg false discovery rate (FDR) procedure was used to adjust P values. Data integration was performed using R packages dplyr (v1.0.6), tidyr (v1.1.3), reshape2 (v1.4.4) and data.table (v1.14.0). Heat maps were plotted using the R package pheatmap (v1.0.12). Other plots were created using the R packages ggplot2 (v3.3.3).

3.6 Contributions

F.E.R. and A.J.L. conceived the study. Q.Z designed the experiments. Q.Z. and E.H contributed to sample processing for DNA sequencing. Q.Z. performed metagenomic, network and GWAS analyses. C.P. assisted with GWAS analyses. Q.Z. and F.E.R. wrote the manuscript. All authors approved the final manuscript.

3.7 Acknowledgements

We thank the University of Wisconsin Biotechnology Center DNA Sequencing Facility for providing sequencing and support services; 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 by National Institutes of Health (NIH) grant HL144651 (F.E.R. and A.J.L.), HL148577 (F.E.R. and A.J.L.).

3.8 References

Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A. M., Valles-Colomer, M., Weingart, G., Zhang, Y., Zolfo, M., Huttenhower, C., Franzosa, E. A., & Segata, N. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife*, *10*, e65088. https://doi.org/10.7554/eLife.65088

Belkaid, Y., & Hand, T. W. (2014). Role of the Microbiota in Immunity and Inflammation. *Cell*, *157*(1), 121–141. https://doi.org/10.1016/j.cell.2014.03.011

Bennett, B. J., Davis, R. C., Civelek, M., Orozco, L., Wu, J., Qi, H., Pan, C., Packard, R. R. S., Eskin, E., Yan, M., Kirchgessner, T., Wang, Z., Li, X., Gregory, J. C., Hazen, S. L., Gargalovic, P. S., & Lusis, A. J. (2015). Genetic Architecture of Atherosclerosis in Mice: A Systems Genetics Analysis of Common Inbred Strains. *PLOS Genetics*, *11*(12), e1005711. https://doi.org/10.1371/journal.pgen.1005711

Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L. J., Thompson, K. N., Zolfo, M., Manghi, P., Dubois, L., Huang, K. D., Thomas, A. M., Nickols, W. A., Piccinno, G., Piperni, E., Punčochář, M., Valles-Colomer, M., Tett, A., Giordano, F., Davies, R., Wolf, J., … Segata, N. (2023). Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nature Biotechnology*. https://doi.org/10.1038/s41587-023-01688-w

Bourgonje, A. R., Andreu-Sánchez, S., Vogl, T., Hu, S., Vich Vila, A., Gacesa, R., Leviatan, S., Kurilshikov, A., Klompus, S., Kalka, I. N., Van Dullemen, H. M., Weinberger, A., Visschedijk, M. C., Festen, E. A. M., Faber, K. N., Wijmenga, C., Dijkstra, G., Segal, E., Fu, J., … Weersma, R. K. (2023). Phage-display immunoprecipitation sequencing of the antibody epitope repertoire in inflammatory bowel disease reveals distinct antibody signatures. *Immunity*, *56*(6), 1393-1409.e6. https://doi.org/10.1016/j.immuni.2023.04.017

Clasen, S. J., Bell, M. E. W., Borbón, A., Lee, D.-H., & Henseler, Z. M. (2023). Silent recognition of flagellins from human gut commensal bacteria by Toll-like receptor 5. *SCIENCE IMMUNOLOGY*.

Costea, P. I., Hildebrand, F., Arumugam, M., Bäckhed, F., Blaser, M. J., Bushman, F. D., de Vos, W. M., Ehrlich, S. D., Fraser, C. M., Hattori, M., Huttenhower, C., Jeffery, I. B., Knights, D., Lewis, J. D., Ley, R. E., Ochman, H., O'Toole, P. W., Quince, C., Relman, D. A., … Bork, P. (2017). Enterotypes in the landscape of gut microbial community composition. *Nature Microbiology*, *3*(1), 8–16. https://doi.org/10.1038/s41564-017-0072-8

Fu, J., Bonder, M. J., Cenit, M. C., Tigchelaar, E. F., Maatman, A., Dekens, J. A. M., Brandsma, E., Marczynska, J., Imhann, F., Weersma, R. K., Franke, L., Poon, T. W., Xavier, R. J., Gevers, D., Hofker, M. H., Wijmenga, C., & Zhernakova, A. (2015). The Gut Microbiome Contributes to a Substantial Proportion of the Variation in Blood Lipids. *Circulation Research*, *117*(9), 817–824. https://doi.org/10.1161/CIRCRESAHA.115.306807

Gacesa, R., Kurilshikov, A., Vich Vila, A., Sinha, T., Klaassen, M. A. Y., Bolte, L. A., Andreu-Sánchez, S., Chen, L., Collij, V., Hu, S., Dekens, J. A. M., Lenters, V. C., Björk, J. R., Swarte, J. C., Swertz, M. A., Jansen, B. H., Gelderloos-Arends, J., Jankipersadsing, S., Hofker, M., … Weersma, R. K. (2022). Environmental factors shaping the gut microbiome in a Dutch population. *Nature*, *604*(7907), 732–739. https://doi.org/10.1038/s41586-022-04567-7

Gevers, D., Kugathasan, S., Denson, L. A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S. J., Yassour, M., Morgan, X. C., Kostic, A. D., Luo, C., González, A., McDonald, D., Haberman, Y., Walters, T., Baker, S., Rosh, J., … Xavier, R. J. (2014). The Treatment-Naive Microbiome in New-Onset Crohn's Disease. *Cell Host & Microbe*, *15*(3), 382–392. https://doi.org/10.1016/j.chom.2014.02.005

Hui, S. T., Kurt, Z., Tuominen, I., Norheim, F., C.Davis, R., Pan, C., Dirks, D. L., Magyar, C. E., French, S. W., Chella Krishnan, K., Sabir, S., Campos-Pérez, F., Méndez-Sánchez, N., Macías-Kauffer, L., León-Mimila, P., Canizales-Quinteros, S., Yang, X., Beaven, S. W., Huertas-Vazquez, A., & Lusis, A. J. (2018). The Genetic Architecture of Diet-Induced Hepatic Fibrosis in Mice. *Hepatology*, *68*(6), 2182–2196. https://doi.org/10.1002/hep.30113

Hutchison, E. R., Kasahara, K., Zhang, Q., Vivas, E. I., Cross, T.-W. L., & Rey, F. E. (2023). Dissecting the impact of dietary fiber type on atherosclerosis in mice colonized with different gut microbial communities. *Npj Biofilms and Microbiomes*, *9*(1), 31. https://doi.org/10.1038/s41522-023-00402-7

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., & Rey, F. E. (2018). Interactions between Roseburia intestinalis and diet modulate atherogenesis in a murine model. *Nature Microbiology*, *3*(12), 1461–1471. https://doi.org/10.1038/s41564-018-0272-x

Kemis, J. H., Linke, V., Barrett, K. L., Boehm, F. J., Traeger, L. L., Keller, M. P., Rabaglia, M. E., Schueler, K. L., Stapleton, D. S., Gatti, D. M., Churchill, G. A., Amador-Noguez, D., Russell, J. D., Yandell, B. S., Broman, K. W., Coon, J. J., Attie, A. D., & Rey, F. E. (2019). Genetic determinants of gut microbiota composition and bile acid profiles in mice. *PLOS Genetics*, *15*(8), e1008073. https://doi.org/10.1371/journal.pgen.1008073

Kurilshikov, A., Medina-Gomez, C., Bacigalupe, R., Radjabzadeh, D., Wang, J., Demirkan, A., Le Roy, C. I., Raygoza Garay, J. A., Finnicum, C. T., Liu, X., Zhernakova, D. V., Bonder, M. J., Hansen, T. H., Frost, F., Rühlemann, M. C., Turpin, W., Moon, J.-Y., Kim, H.-N., Lüll, K., … Zhernakova, A. (2021). Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nature Genetics*, *53*(2), 156–165. https://doi.org/10.1038/s41588-020-00763-1

Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I., & Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. *Nature Methods*, *8*(10), 833–835. https://doi.org/10.1038/nmeth.1681

Liu, X., Tong, X., Zou, Y., Lin, X., Zhao, H., Tian, L., Jie, Z., Wang, Q., Zhang, Z., Lu, H., Xiao, L., Qiu, X., Zi, J., Wang, R., Xu, X., Yang, H., Wang, J., Zong, Y., Liu, W., … Zhang, T. (2020). Mendelian randomization analyses support causal relationships between blood metabolites and the gut microbiome. *Nature Genetics*. https://doi.org/10.1101/2020.06.30.181438

Lkhagva, E., Chung, H.-J., Hong, J., Tang, W. H. W., Lee, S.-I., Hong, S.-T., & Lee, S. (2021). The regional diversity of gut microbiome along the GI tract of male C57BL/6 mice. *BMC Microbiology*, *21*(1), 44. https://doi.org/10.1186/s12866-021-02099-0

Lynch, S. V., & Pedersen, O. (2016). The Human Intestinal Microbiome in Health and Disease. *New England Journal of Medicine*, *375*(24), 2369–2379. https://doi.org/10.1056/NEJMra1600266

MetaHIT Consortium (additional members), Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J.-M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., … Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature*, *473*(7346), 174–180. https://doi.org/10.1038/nature09944

Murga-Garrido, S. M., Hong, Q., Cross, T.-W. L., Hutchison, E. R., Han, J., Thomas, S. P., Vivas, E. I., Denu, J., Ceschin, D. G., Tang, Z.-Z., & Rey, F. E. (2021). Gut microbiome variation modulates the effects of dietary fiber on host metabolism. *Microbiome*, *9*(1), 117. https://doi.org/10.1186/s40168-021-01061-6

Oliphant, K., & Allen-Vercoe, E. (2019). Macronutrient metabolism by the human gut microbiome: Major fermentation by-products and their impact on host health. *Microbiome*, *7*(1), 91. https://doi.org/10.1186/s40168-019-0704-8

Org, E., Parks, B. W., Joo, J. W. J., Emert, B., Schwartzman, W., Kang, E. Y., Mehrabian, M., Pan, C., Knight, R., Gunsalus, R., Drake, T. A., Eskin, E., & Lusis, A. J. (2015). Genetic and environmental control of host-gut microbiota interactions. *Genome Research*, *25*(10), 1558–1569. https://doi.org/10.1101/gr.194118.115

Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R., Werner, J., Villanea, F. A., Mountain, J. L., Misra, R., Carter, N. P., Lee, C., & Stone, A. C. (2007). Diet and the evolution of human amylase gene copy number variation. *Nature Genetics*, *39*(10), 1256–1260. https://doi.org/10.1038/ng2123

Poole, A. C., Goodrich, J. K., Youngblut, N. D., Luque, G. G., Ruaud, A., Sutter, J. L., Waters, J. L., Shi, Q., El-Hadidi, M., Johnson, L. M., Bar, H. Y., Huson, D. H., Booth, J. G., & Ley, R. E. (2019). Human Salivary Amylase Gene Copy Number Impacts Oral and Gut Microbiomes. *Cell Host & Microbe*, *25*(4), 553-564.e7. https://doi.org/10.1016/j.chom.2019.03.001

Qin, Y., Havulinna, A. S., Liu, Y., Jousilahti, P., Ritchie, S. C., Tokolyi, A., Sanders, J. G., Valsta, L., Brożyńska, M., Zhu, Q., Tripathi, A., Vazquez-Baeza, Y., Loomba, R., Cheng, S., Jain, M., Niiranen, T., Lahti, L., Knight, R., Salomaa, V., … Méric, G. (2020). Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. *Nature Genetics*. https://doi.org/10.1101/2020.09.12.20193045

Rühlemann, M. C., Hermes, B. M., Bang, C., Doms, S., Moitinho-Silva, L., Thingholm, L. B., Frost, F., Degenhardt, F., Wittig, M., Kässens, J., Weiss, F. U., Peters, A., Neuhaus, K., Völker, U., Völzke, H., Homuth, G., Weiss, S., Grallert, H., Laudes, M., … Franke, A. (2021). Genome-wide association study in 8,956 German individuals identifies influence of ABO histo-blood groups on gut microbiome. *Nature Genetics*, *53*(2), 147–155. https://doi.org/10.1038/s41588-020-00747-1

Sanna, S. (2022). Challenges and future directions for studying effects of host genetics on the gut microbiome. *Nature Genetics*, 7.

Sanna, S., van Zuydam, N. R., Mahajan, A., Kurilshikov, A., Vich Vila, A., Võsa, U., Mujagic, Z., Masclee, A. A. M., Jonkers, D. M. A. E., Oosting, M., Joosten, L. A. B., Netea, M. G., Franke, L., Zhernakova, A., Fu, J., Wijmenga, C., & McCarthy, M. I. (2019). Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nature Genetics*, *51*(4), 600–605. https://doi.org/10.1038/s41588-019-0350-x

Smith, B. J., Miller, R. A., Ericsson, A. C., Harrison, D. C., Strong, R., & Schmidt, T. M. (2019). Changes in the gut microbiome and fermentation products concurrent with enhanced longevity in acarbose-treated mice. *BMC Microbiology*, *19*(1), 130. https://doi.org/10.1186/s12866-019-1494-7

Smith, B. J., Miller, R. A., & Schmidt, T. M. (2021). Muribaculaceae Genomes Assembled from Metagenomes Suggest Genetic Drivers of Differential Response to Acarbose Treatment in Mice. *mSphere*, *6*(6), e00851-21.

Tremaroli, V., & Bäckhed, F. (2012). Functional interactions between the gut microbiota and host metabolism. *Nature*, *489*(7415), 242–249. https://doi.org/10.1038/nature11552

Usher, C. L., Handsaker, R. E., Esko, T., Tuke, M. A., Weedon, M. N., Hastie, A. R., Cao, H., Moon, J. E., Kashin, S., Fuchsberger, C., Metspalu, A., Pato, C. N., Pato, M. T., McCarthy, M. I., Boehnke, M., Altshuler, D. M., Frayling, T. M., Hirschhorn, J. N., & McCarroll, S. A. (2015). Structural forms of the human amylase locus and their relationships to SNPs, haplotypes and obesity. *Nature Genetics*, *47*(8), 921–925. https://doi.org/10.1038/ng.3340

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*(7341), 57–63. https://doi.org/10.1038/nature09922

Yiu, J. H. C., Chan, K.-S., Cheung, J., Li, J., Liu, Y., Wang, Y., Fung, W. W. L., Cai, J., Cheung, S. W. M., Dorweiler, B., Wan, E. Y. F., Tso, P., Xu, A., & Woo, C. W. (2020). Gut Microbiota-Associated Activation of TLR5 Induces Apolipoprotein A1 Production in the Liver. *Circulation Research*, *127*(10), 1236–1252. https://doi.org/10.1161/CIRCRESAHA.120.317362

Zahavi, L., Lavon, A., Reicher, L., Shoer, S., Godneva, A., Leviatan, S., Rein, M., Weissbrod, O., Weinberger, A., & Segal, E. (2023). Bacterial SNPs in the human gut microbiome associate with host BMI. *Nature Medicine*. https://doi.org/10.1038/s41591-023-02599-8

Zhang, Q., Linke, V., Overmyer, K. A., Traeger, L. L., Kasahara, K., Miller, I. J., Manson, D. E., Polaske, T. J., Kerby, R. L., Kemis, J. H., Trujillo, E. A., Reddy, T. R., Russell, J. D., Schueler, K. L., Stapleton, D. S., Rabaglia, M. E., Seldin, M., Gatti, D. M., Keele, G. R., … Rey, F. E. (2023). Genetic mapping of microbial and host traits reveals production of immunomodulatory lipids by Akkermansia muciniphila in the murine gut. *Nature Microbiology*, *8*(3), 424–440. https://doi.org/10.1038/s41564-023-01326-w

3.9 Figures

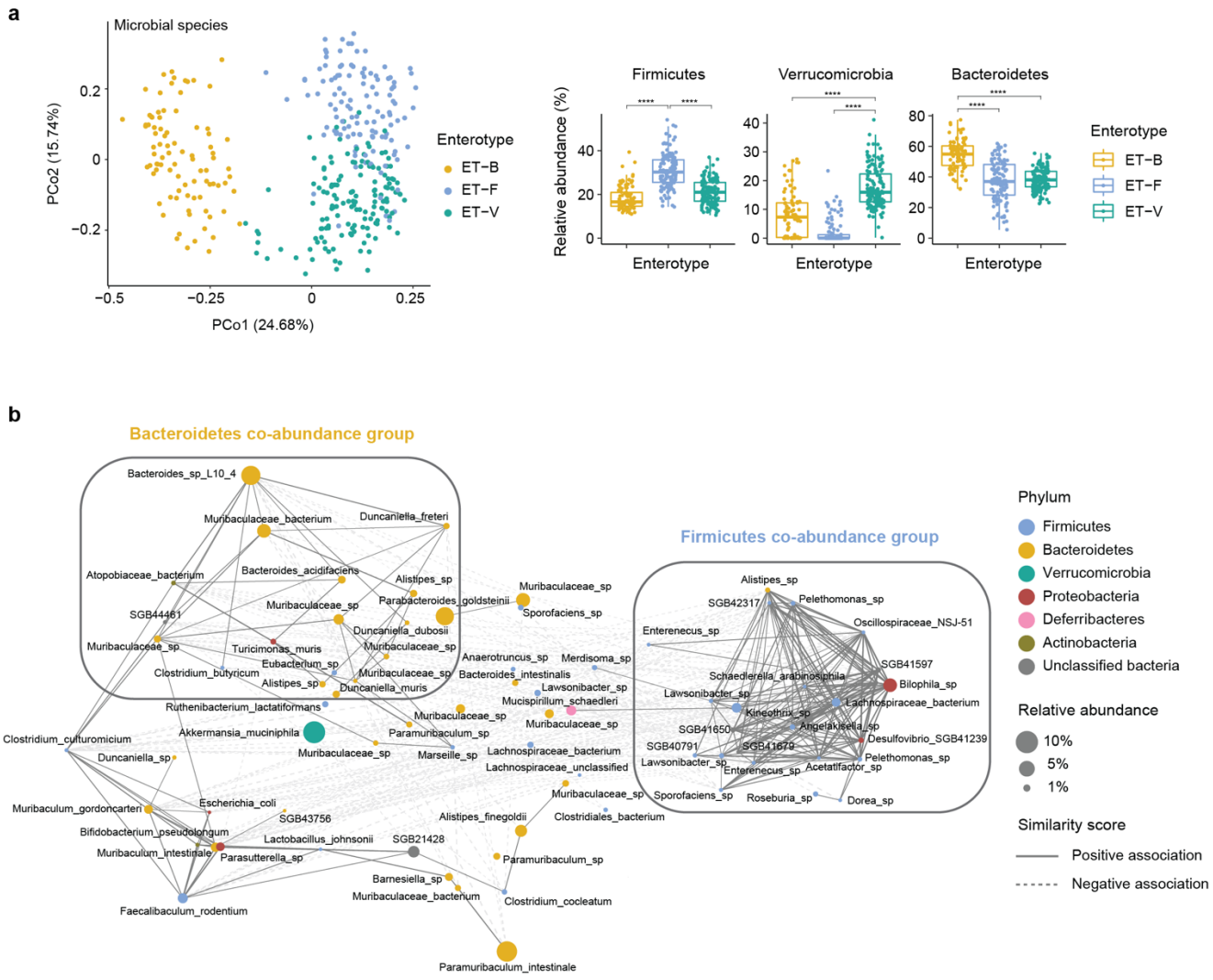
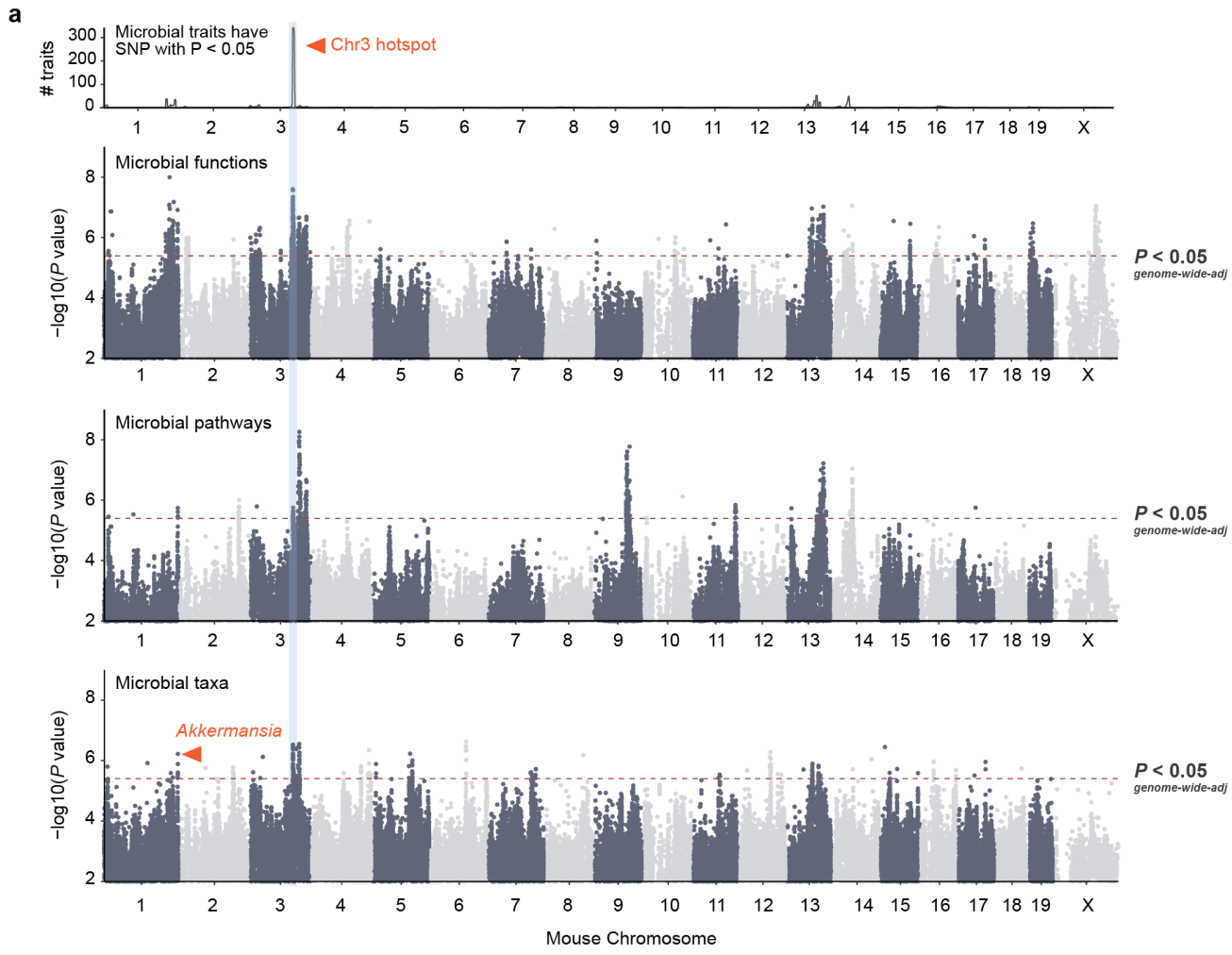
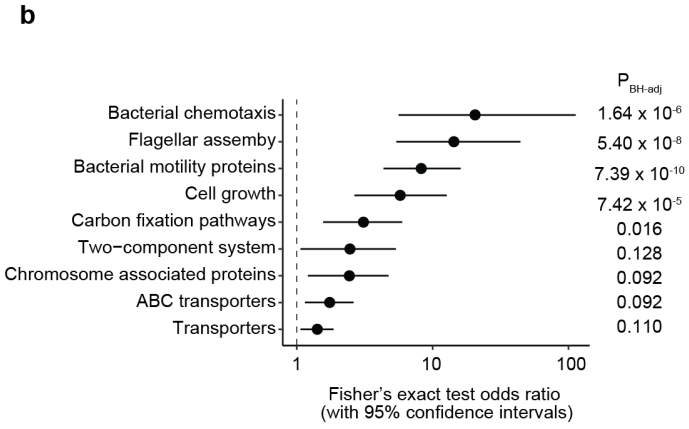


Figure 3.1 Enterotypes and bacterial species co-abundance groups in the gut. a. Clustering identified three enterotypes in F1 Ath-HMDP mouse cohort. Significance was calculated by unpaired two-tailed Welch's t-test and is designated as follows: ** p value < 0.01; *** p value < 0.001; **** p value < 0.0001. ET-B, Bacteroidetes enterotype; ET-F, Firmicutes enterotype; ET-V, Verrucomicrobia enterotype. **b**. The correlation network of bacteria species (average relative abundance > 0.1 % and present in at least 20% of samples) using CCREPE with a checkerboard score, indicating a strong co-occurrence between species of the same behavior and a co-

exclusion between taxa of a different behavior. Nodes represent the species and lines represent similarity score. Solid lines represent co-occurrence species and dashed lines represent co-exclusion species.


a
Microbial traits have SNP with P < 0.05
Chr3 hotspot
traits
Microbial functions
Microbial pathways
Microbial taxa
Akkermansia
−log10(P value)
P < 0.05
genome-wide-adj
Mouse Chromosome


b
Bacterial chemotaxis
Flagellar assemby
Bacterial motility proteins
Cell growth
Carbon fixation pathways
Two−component system
Chromosome associated proteins
ABC transporters
Transporters
P BH-adj
1.64 x 10-6
5.40 x 10-8
7.39 x 10-10
7.42 x 10-5
0.016
0.128
0.092
0.092
0.110
Fisher's exact test odds ratio
(with 95% confidence intervals)

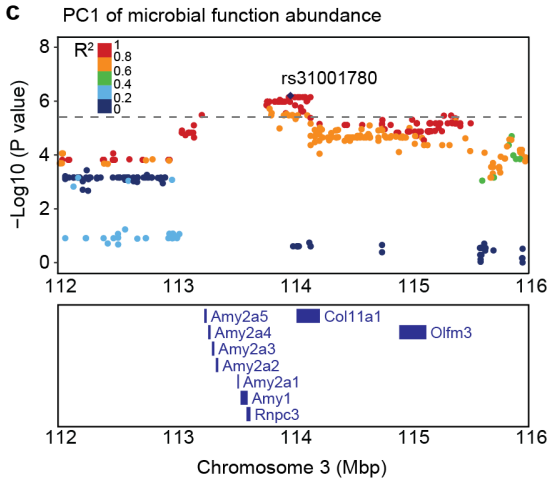

c
PC1 of microbial function abundance
R2
rs31001780
−Log10 (P value)
Amy2a5
Amy2a4
Amy2a3
Amy2a2
Amy2a1
Amy1
Rnpc3
Col11a1
Olfm3
Chromosome 3 (Mbp)

Figure 3.2 Genetic association of gut microbiomes in Ath-HMDP. a. Genetic associations of gut microbial functions, taxa and metabolic pathways. The density of associated gut bacterial function features at mouse genome is showed on the top. Dashed lines represent significance thresholds determined by permutation tests ($P < 4 \times 10^{-6}$). **b.** Enrichment analysis using Fisher's exact test for gut bacterial functions that are mapped at the hotspot locus on Chr3: 112–116 Mbp. **c.** SNP associations for PC1 of microbial functions at chromosome 3. Protein coding genes are displayed for Chr3: 112–116 Mbp region.

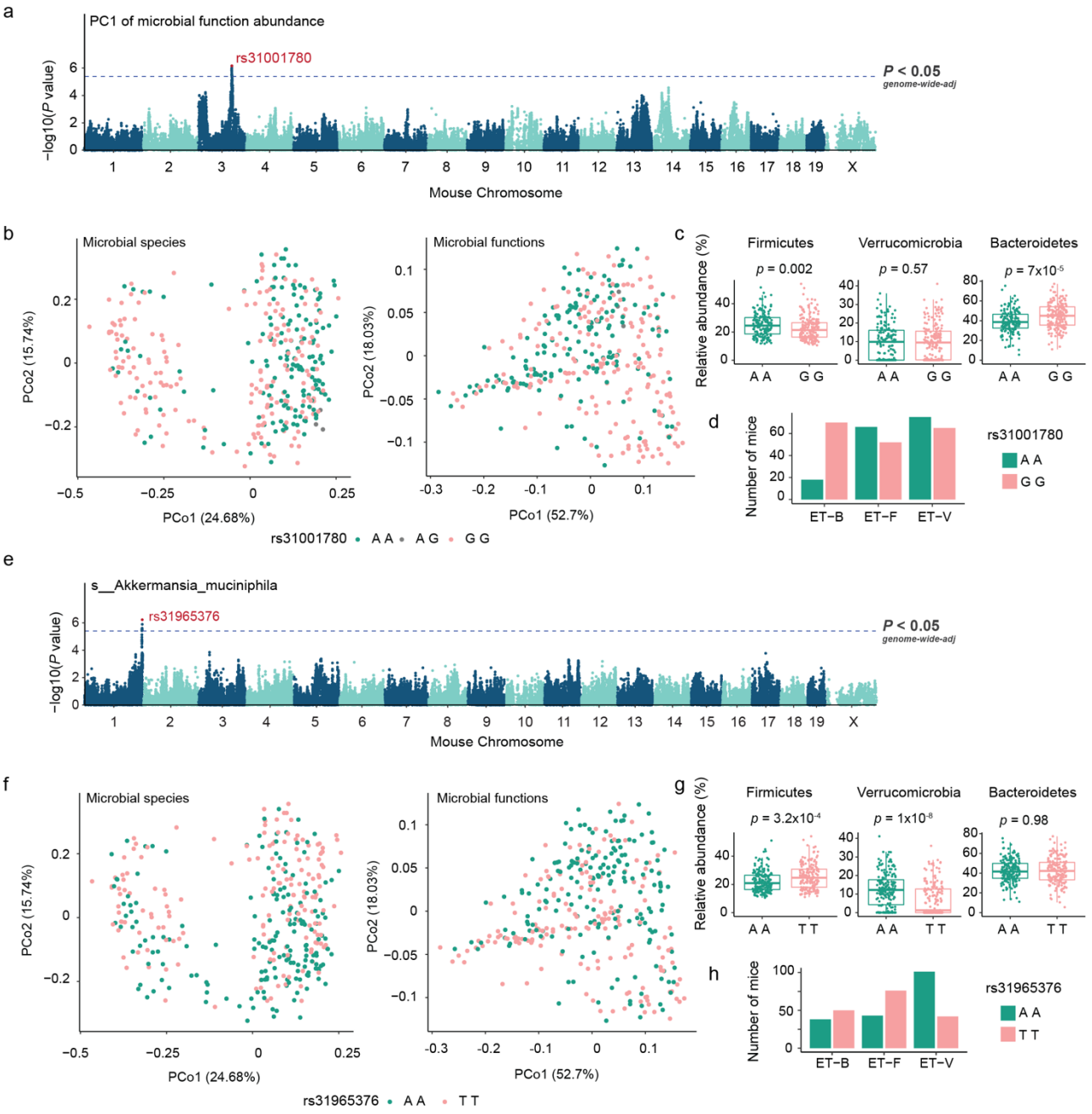


Figure 3.3 Genomic loci are associated microbial enterotypes in gut. a. Genomic locus at Chr3 113-114 Mbp is associated with PC1 of microbial functions. The lead SNP is an intron SNP rs31001780. Dashed lines represent significance thresholds determined by permutation tests ($P < 4 \times 10^{-6}$). **b**. Mouse with allele AA, AG or GG at SNP rs31001780 visualized in PCoA of species beta-diversity (left) and functions beta-diversity (right). **b**. Relative abundance of Firmicutes,

Bacteroidetes and Verrucomicrobia from mouse with AA or GG at SNP rs31001780. **d**. Number of mice with AA or GG at SNP rs31001780 in each of three enterotypes. **e**. Genomic locus at Chr1 193-194 Mbp are associated with *Akkermansia muciniphila* abundance. The lead SNP is rs31965376. Dashed lines represent significance thresholds determined by permutation tests ($P < 4 \times 10^{-6}$). **f**. Mouse with allele AA or TT at SNP rs31965376 visualized in PCoA of species beta-diversity (left) and functions beta-diversity (right). **g**. Relative abundance of Firmicutes, Bacteroidetes and Verrucomicrobia from mouse with AA or TT at SNP rs31965376. **h**. Number of mice with AA or TT at SNP rs31965376 in each of three enterotypes.

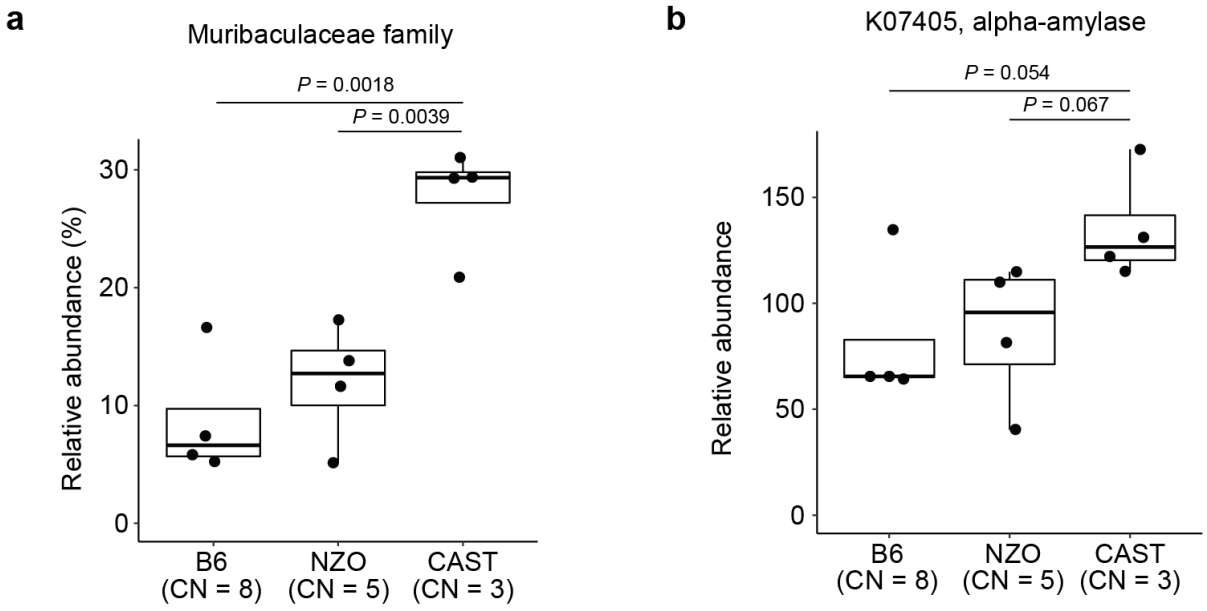


Figure 3.4 Gut microbiome of high/medium/low amylase copy number mouse in high carbohydrate diet. a. Relative abundance of *Muribaculaceae* family in B6, NZO and CAST mouse fed on 16 weeks of high carbohydrate diet. **b**. Relative abundance of bacteria alpha-amylase enzyme (K07405) in B6, NZO and CAST mouse fed on 16 weeks of high carbohydrate diet. Statistical difference between treatment groups was tested by unpaired two-sided Welch's t-test.

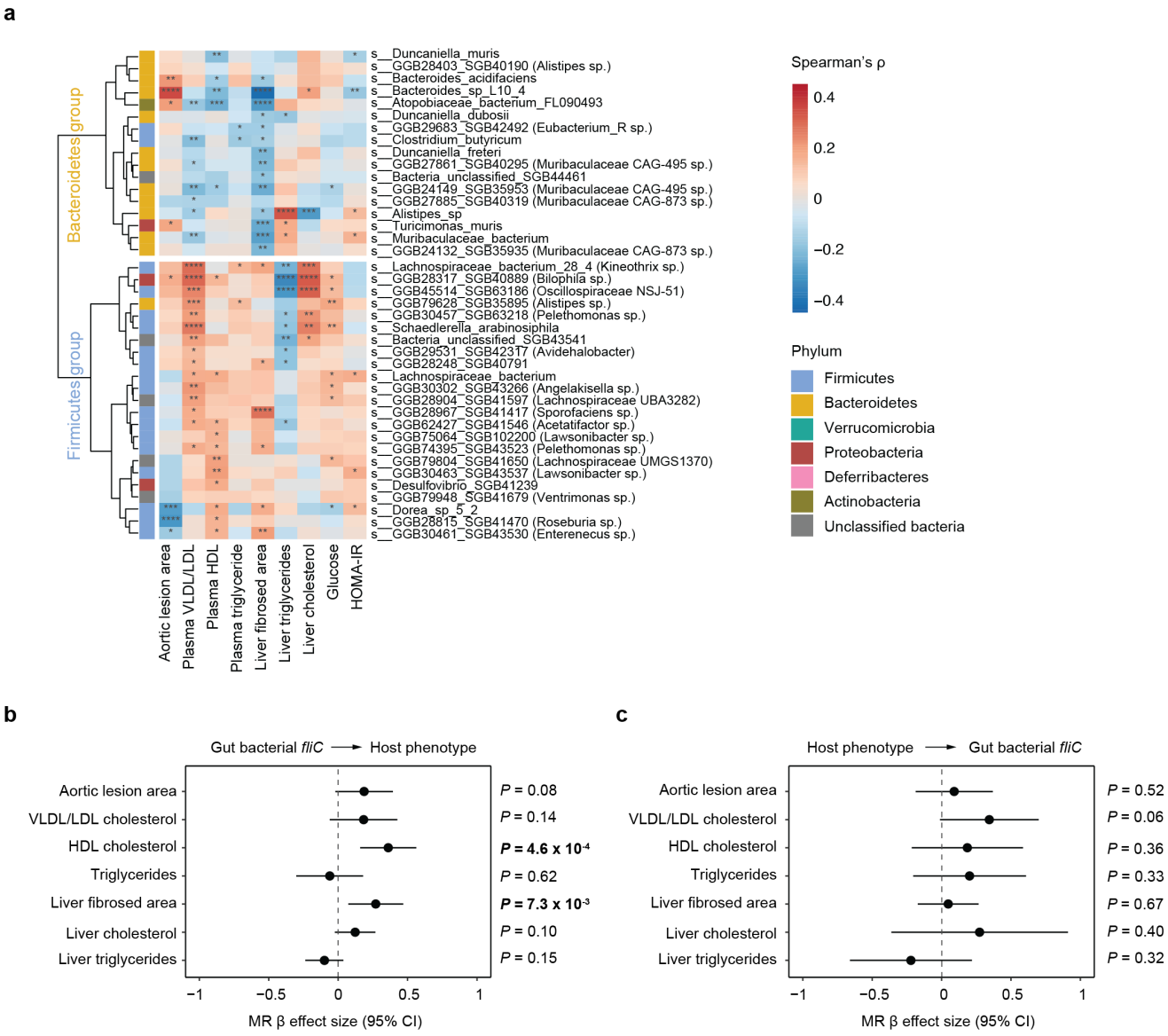
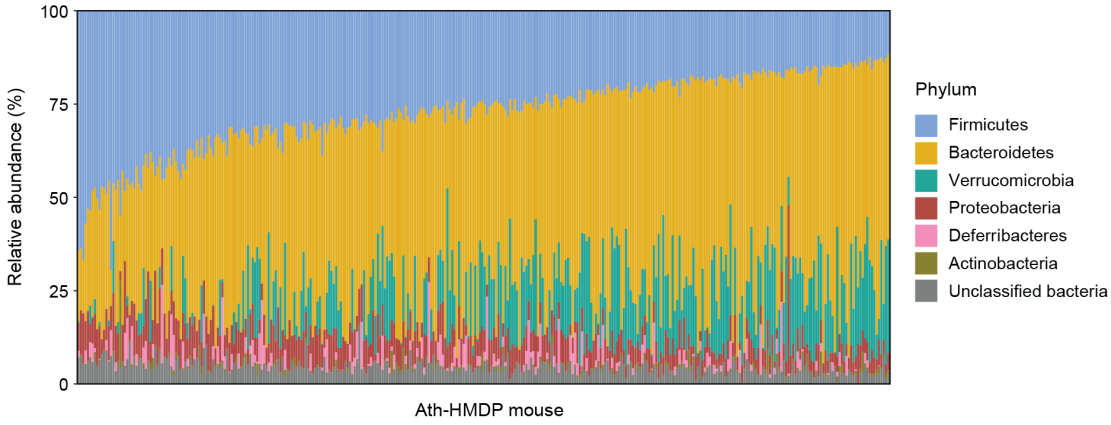
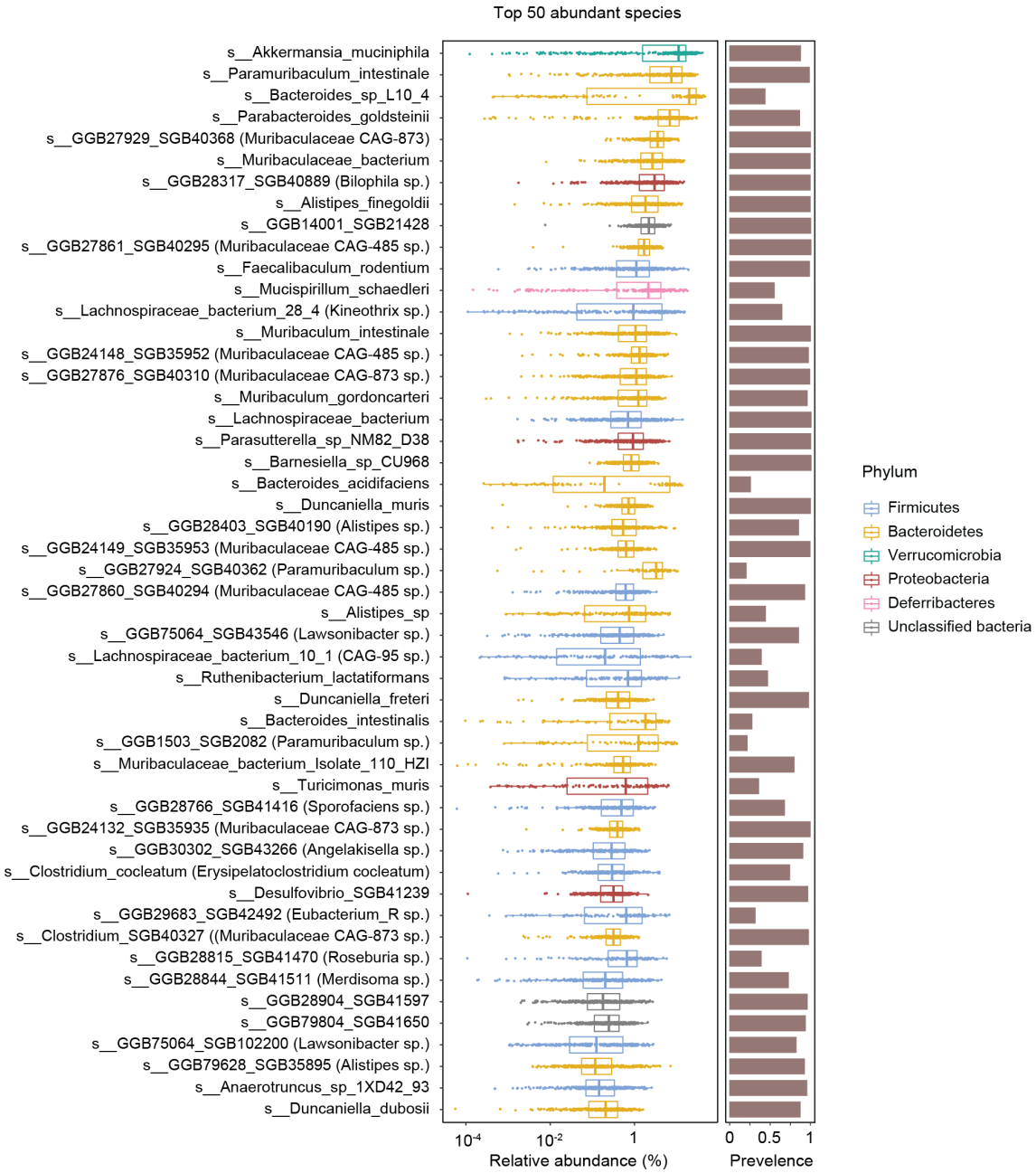


Figure 3.5 Host phenotypes are associated with CAGs and gut bacteria flagellar. a. Spearman's correlation between bacteria species and host physiology phenotypes. Bacteria species from Firmicutes CAG and Bacteroidetes CAG showed different associations. **b**. Bidirectional MR analysis between host clinical traits and gut bacterial flagellin gene abundance. MR results using gut bacterial *fliC* abundance as exposure and clinical traits as outcome. **c**. MR results using clinical traits as exposure and gut bacterial *fliC* abundance as outcome.

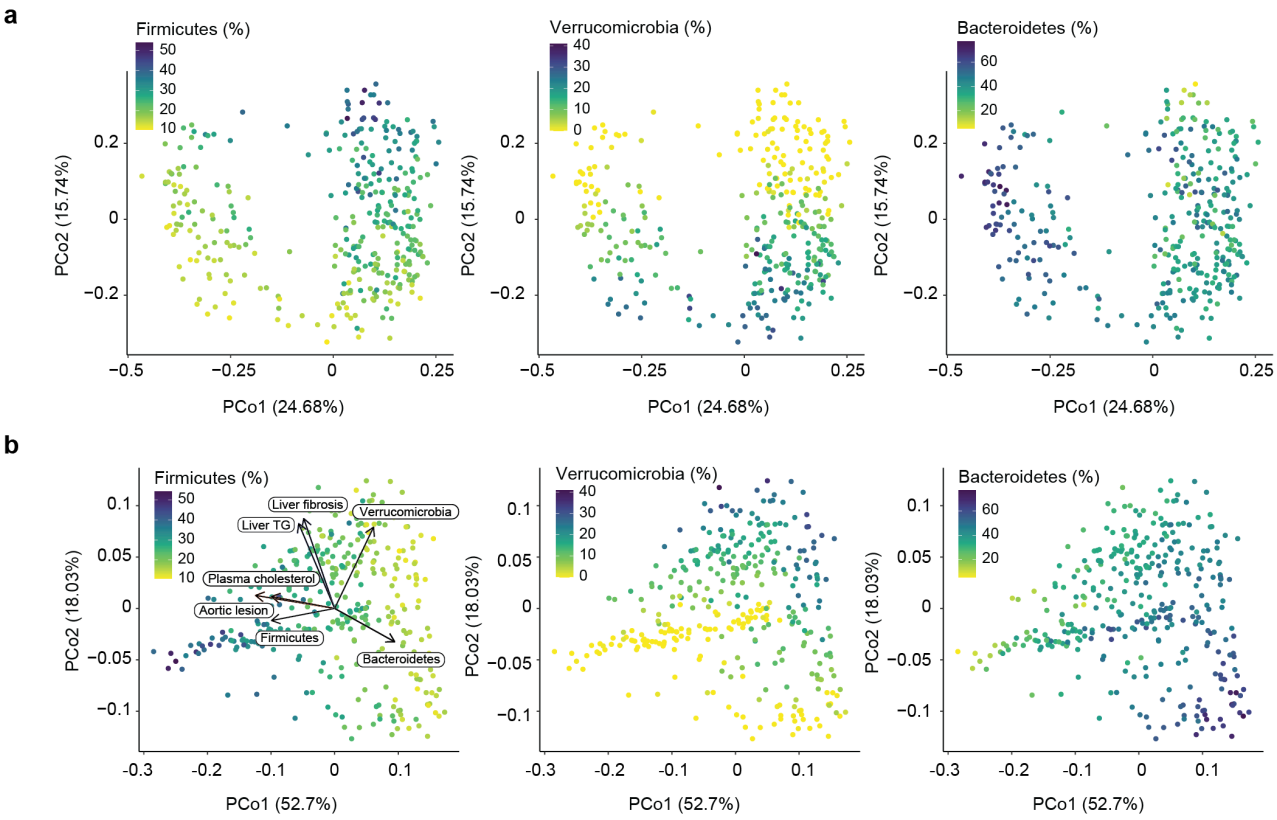
a


100
75
50
25
0
Relative abundance (%)
Ath-HMDP mouse
Phylum
Firmicutes
Bacteroidetes
Verrucomicrobia
Proteobacteria
Deferribacteres
Actinobacteria
Unclassified bacteria

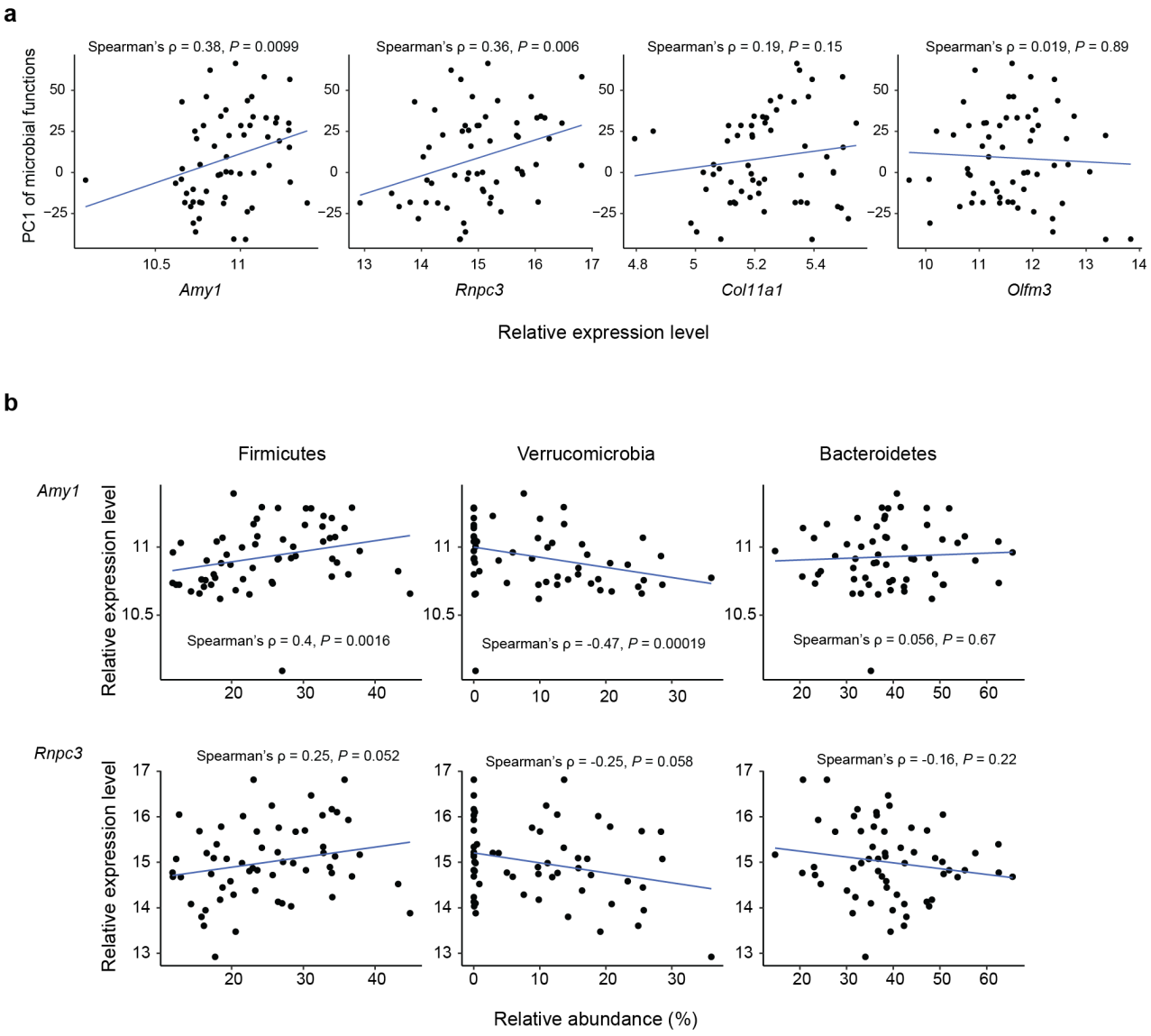
b


Top 50 abundant species
s__Akkermansia_muciniphila
s__Paramuribaculum_intestinale
s__Bacteroides_sp_L10_4
s__Parabacteroides_goldsteinii
s__GGB27929_SGB40368 (Muribaculaceae CAG-873)
s__Muribaculaceae_bacterium
s__GGB28317_SGB40889 (Bilophila sp.)
s__Alistipes_finegoldii
s__GGB14001_SGB21428
s__GGB27861_SGB40295 (Muribaculaceae CAG-485 sp.)
s__Faecalibaculum_rodentium
s__Mucispirillum_schaedleri
s__Lachnospiraceae_bacterium_28_4 (Kineothrix sp.)
s__Muribaculum_intestinale
s__GGB24148_SGB35952 (Muribaculaceae CAG-485 sp.)
s__GGB27876_SGB40310 (Muribaculaceae CAG-873 sp.)
s__Muribaculum_gordoncarteri
s__Lachnospiraceae_bacterium
s__Parasutterella_sp_NM82_D38
s__Barnesiella_sp_CU968
s__Bacteroides_acidifaciens
s__Duncaniella_muris
s__GGB28403_SGB40190 (Alistipes sp.)
s__GGB24149_SGB35953 (Muribaculaceae CAG-485 sp.)
s__GGB27924_SGB40362 (Paramuribaculum sp.)
s__GGB27860_SGB40294 (Muribaculaceae CAG-485 sp.)
s__Alistipes_sp
s__GGB75064_SGB43546 (Lawsonibacter sp.)
s__Lachnospiraceae_bacterium_10_1 (CAG-95 sp.)
s__Ruthenibacterium_lactatiformans
s__Duncaniella_freteri
s__Bacteroides_intestinalis
s__GGB1503_SGB2082 (Paramuribaculum sp.)
s__Muribaculaceae_bacterium_Isolate_110_HZI
s__Turicimonas_muris
s__GGB28766_SGB41416 (Sporofaciens sp.)
s__GGB24132_SGB35935 (Muribaculaceae CAG-873 sp.)
s__GGB30302_SGB43266 (Angelakisella sp.)
s__Clostridium_cocleatum (Erysipelatoclostridium cocleatum)
s__Desulfovibrio_SGB41239
s__GGB29683_SGB42492 (Eubacterium_R sp.)
s__Clostridium_SGB40327 ((Muribaculaceae CAG-873 sp.)
s__GGB28815_SGB41470 (Roseburia sp.)
s__GGB28844_SGB41511 (Merdisoma sp.)
s__GGB28904_SGB41597
s__GGB79804_SGB41650
s__GGB75064_SGB102200 (Lawsonibacter sp.)
s__GGB79628_SGB35895 (Alistipes sp.)
s__Anaerotruncus_sp_1XD42_93
s__Duncaniella_dubosii
10^{-4}
10^{-2}
1
Relative abundance (%)
0
0.5
1
Prevelence
Phylum
Firmicutes
Bacteroidetes
Verrucomicrobia
Proteobacteria
Deferribacteres
Unclassified bacteria

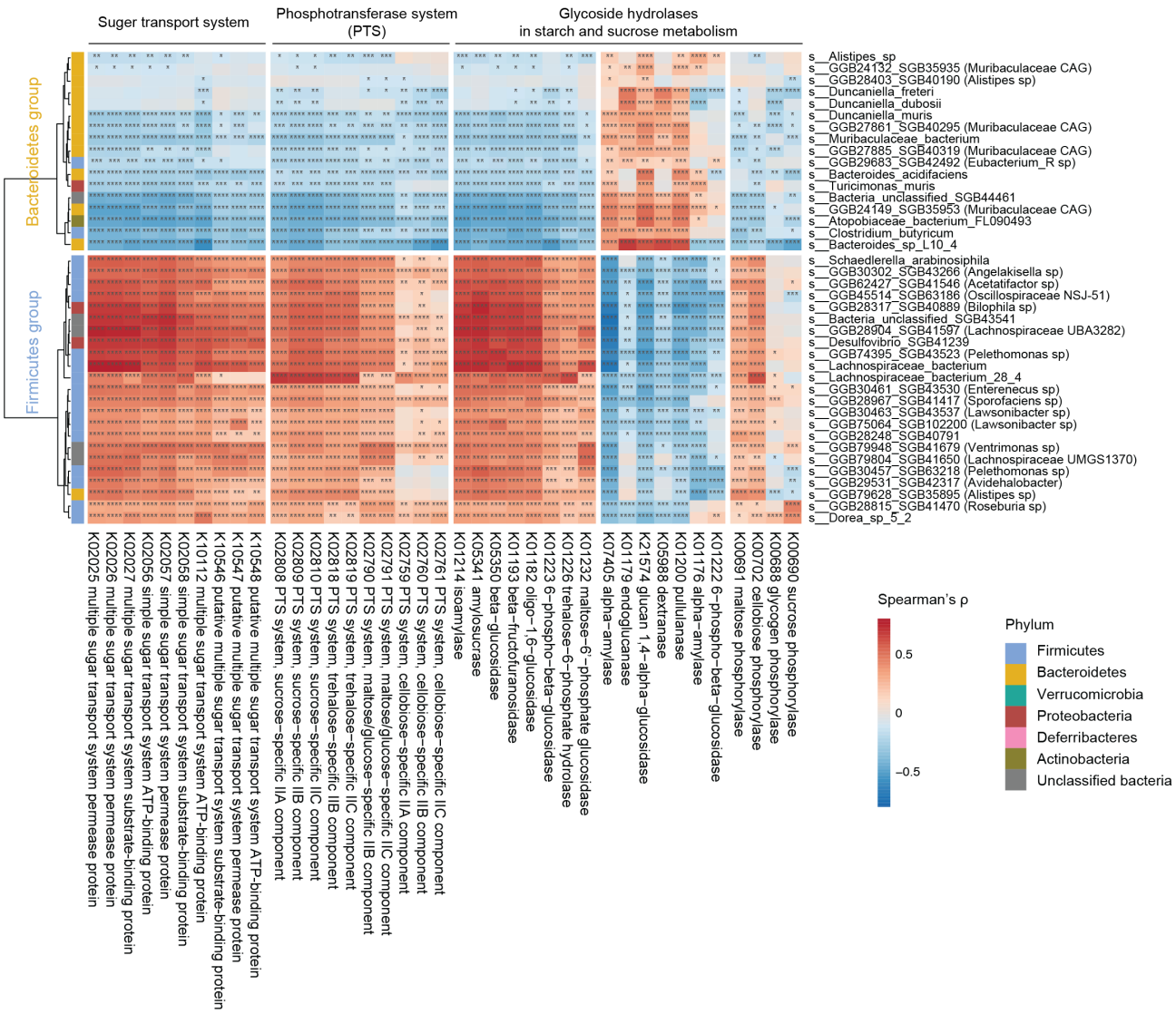
Supplementary Figure 3.1 Gut microbiota composition summary. a. The relative abundance of phyla levels taxa in Ath-HMDP mouse. **b**. The relative abundance of top 50 bacteria species and their prevalence in Ath-HMDP mouse cohort.



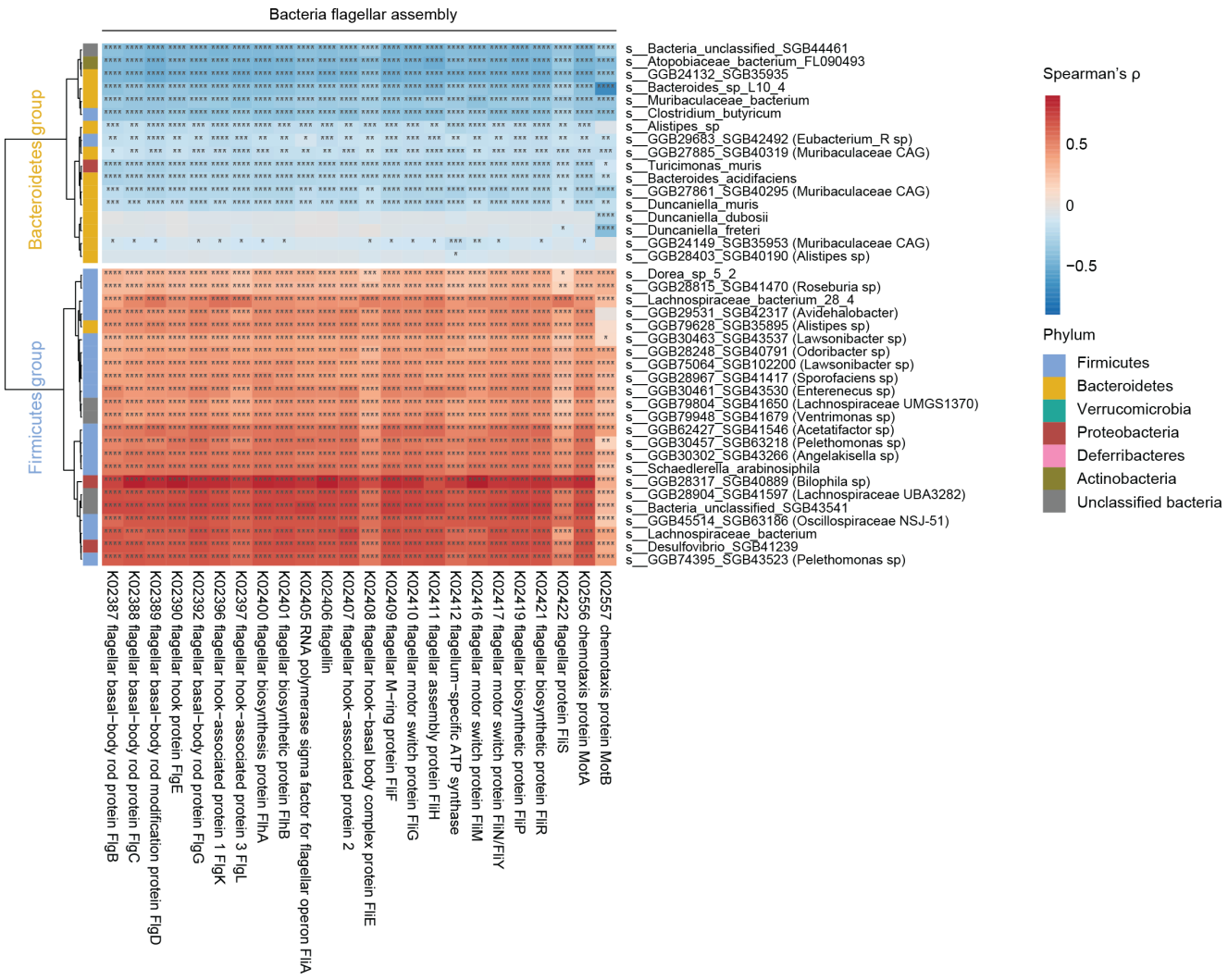
Supplementary Figure 3.2 Gut microbial species and functions in the F1 Ath-HMDP mice. **a**. PCoA showing community-level difference among mice in the cohort. Three major phyla show large inter-individual variance and stratify the mouse samples. **b.** PCoA visualizing the beta-diversity of functions in the cohort. Three major phyla show large inter-individual variance and stratify the mouse samples.



Supplementary Figure 3.3 Amy1 gene expression level correlated with microbial enterotypes. a. Spearman's correlation between candidate genes (*Amy1*, *Rnpc3*, *Col11a1* and *Olfm3*) in Chr3 locus with PC1 of microbial functions. **a**. Spearman's correlation between candidate genes (*Amy1* and *Rnpc3*) in Chr3 locus with abundance of Firmicutes, Verrucomicrobia and Bacteroidetes.

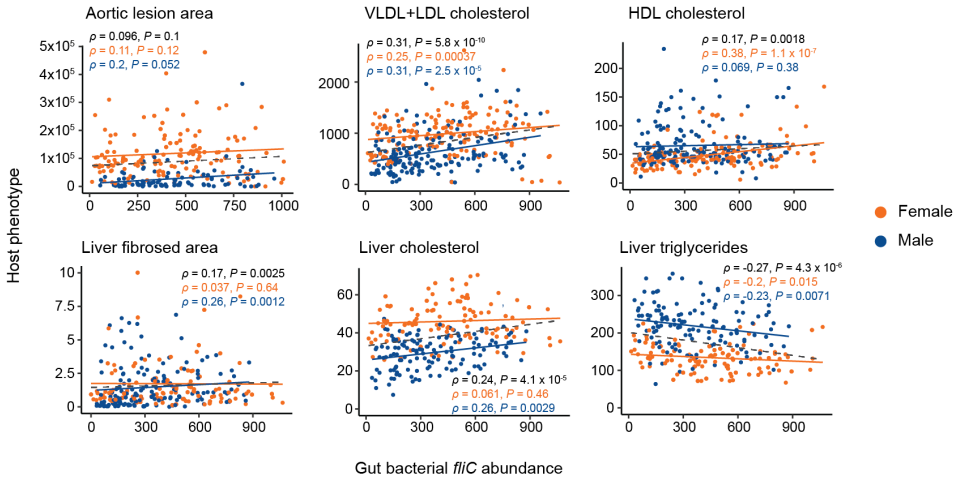


Supplementary Figure 3.4 Bacteria starch and sugar metabolism is associated with microbial enterotypes species. a. Spearman's correlation between bacteria species and bacterial functions involved in starch and sugar metabolism. Bacteria species from Firmicutes CAG and Bacteroidetes CAG showed different associations.

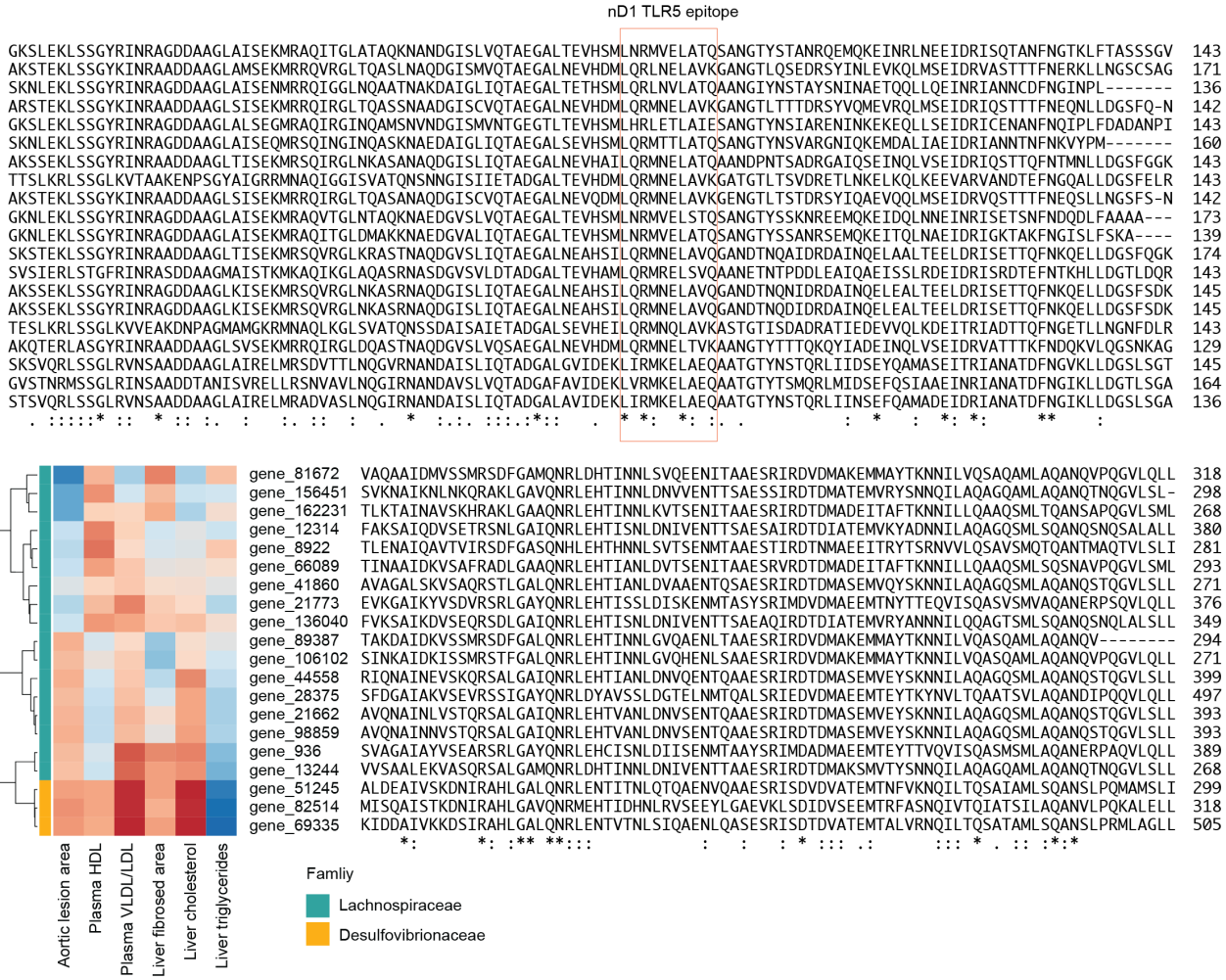


Supplementary Figure 3.5 Bacteria flagellar assembly genes are associated with microbial enterotypes species. a. Spearman's correlation between bacteria species and bacterial functions involved in flagellar assembly. Bacteria species from Firmicutes CAG and Bacteroidetes CAG showed different associations.

a


Aortic lesion area
VLDL+LDL cholesterol
HDL cholesterol
Liver fibrosed area
Liver cholesterol
Liver triglycerides
Host phenotype
Gut bacterial fliC abundance
Female
Male

b


nD1 TLR5 epitope
ρ
Aortic lesion area
Plasma HDL
Plasma VLDL/LDL
Liver fibrosed area
Liver cholesterol
Liver triglycerides
Famliy
Lachnospiraceae
Desulfovibrionaceae

Supplementary Figure 3.6 Correlation between gut bacterial *fliC* gene with host phenotypes. a. Spearman's correlation between host clinical traits and gut bacterial flagellin gene abundance. **b**. Spearman's correlation between host clinical traits and most abundant individual gut bacterial flagellin gene abundance from Lachnospiraceae and Desulfovibrionaceae family. The conserved N-terminal and C-terminal motifs from flagellin genes were aligned.

CHAPTER 4: Conclusion and Future Work

Comprehensive characterization of the gut microbiome in a population is a demanding task. This thesis leverages shotgun metagenomic methods to profile gut microbial composition, function and metabolic pathway in two genetically diverse mouse cohorts, the Diversity Outbred (DO) and the Hybrid Mouse Diverse Panel (HMDP). These datasets provide valuable references for future mouse gut microbiome research. The interindividual variation in gut microbiome observed in these cohorts is associated with host cardiovascular and metabolic phenotypes. This thesis work aims to further understand the effects of host genetics on gut microbiome and how gut microbiome-host interactions contribute to the host health outcomes. Applying of systems genetics methods in DO and HMDP cohorts, I further identified novel connections between gut microbiome and the host. This work provides the foundation for the future mechanistic studies and exploration of potential therapeutic avenues.

Shotgun metagenomic sequencing allows us to characterize the microbial functions and metabolic pathways in addition to the microbial compositions. I performed *de novo* assembly and used reference-based methods to quantify microbial phenotypes using the metagenomic reads. A limitation of current metagenomic studies is the lack of robust references to annotate microbial genes and genomes accurately. This leads to a considerable proportion of genes from analyzed results are taxonomically or functionally uncharacterized. Bioinformatic techniques to analyze shotgun metagenomic data have been developed rapidly in the past five years. New tools and algorithms are enabling us to profile the gut microbiome more efficiently and accurately (Blanco-Míguez et al., 2023, p. 4). Together with the exponentially increase in the number of the reference and metagenome-assembled genomes available over time, re-analyzing the metagenomic data in the future using update databases and tools would allow to detect rare bacteria or uncharacterized microbial functions present in these mice.

Previous efforts for identification of host genetic associations to gut microbiome are limited by the environmental cofounding variables in humans (Sanna, 2022), or only focused on microbial composition in mice (Kemis et al., 2019; Org et al., 2015). In **Chapter 2**, we leveraged the DO

mouse cohort, a genetically diverse population derived from eight founder strains including five common laboratory mouse strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HLtJ) and three wild derived mouse strains (CAST/EiJ, PWK/PhJ, WSB/EiJ). These eight strains were used for high-resolution genetic mapping of gut microbiome. The DO mice were housed under strict and identical environment conditions to eliminate environmental cofounding variables. The DO mice were fed with high fat and high sucrose diets based on the previous study showed that this diet elicits a wide range of metabolic responses in the eight founder strains that are associated with microbiome changes (Kemis et al., 2019). In addition, the diversity of gut microbiome is higher in HF/HS diet mice compared with low fat diet mice, thus mice fed by HF/HS offer more gut microbiome traits to investigate (Kemis et al., 2019).

We found overlapping QTL for the abundance of *Akkermansia muciniphila* and cacal levels of ornithine lipids in DO mice. Interestingly, we identified multiple genomic loci (Chromosome 1, Chromosome 2, Chromosome 7, Chromosome 12, Chromosome 15) that were associated with these two traits. This indicates the abundance of *A. muciniphila* and cacal levels of ornithine lipids (OL) are polygenetic. Indeed, I compared the co-mapping between eQTL genes in small intestine with *A. muciniphila* QTL to select one of the candidate host gene *Atf3* for validation experiments. However, we didn't see significant difference of gut *A. muciniphila* abundance, but only the trend, in *Atf3* knockout and wild type mice. These suggest that there exist more than one QTL, dependently or independently, associated with the abundance of *A. muciniphila* and cacal levels of OL. The current genetic mapping methods focus on single QTL model, that is considering no interactions between genomic loci to a trait. Applying multiple-QTL model in the future would allow to better identify the genetic associations of polygenetic traits such as the abundance of *A. muciniphila*.

We found the *A. muciniphila* derived OL modulate lipopolysaccharide-induced cytokine responses in mouse macrophage cells, suggesting the immunomodulatory role of OL to the host. This inhibitory effects of OL on LPS-induced cytokines aligned with the observation from previous

studies (Kawai et al., 1991; Kawai & Akagawa, 1989). Other studies showed that both natural and synthetic molecules that can inhibit TLR4-mediated LPS signaling—compounds that have anti-inflammatory and anti-neuropathic pain activities *in vivo*. Thus, further investigation using mouse model need to be done to directly test the OL effects on host *in vivo*. In addition, we hypothesized the anti-inflammatory effects of OL to the host may through the TLR4 pathway. Additional experiments using TLR4 knock out or knock down macrophage cells will provide a better understanding of the molecular mechanisms of OL-host interactions in the future.

In **Chapter 3**, we comprehensively characterized gut microbiome composition, functions, and metabolic pathways in 90 inbred hyperlipidemic mouse strains from the HMDP. We identified the genomic loci that were associated with microbial enterotypes in the gut. The genetic variant rs31001780 (A/G) at Chr3 locus was significantly associated with Firmicutes and Bacteroidetes enterotypes. We found the expression level of *Amy1* gene, which spans in LD region of Chr3 locus, was positively correlated with Firmicutes (*Lachnospiraceae* family) and negatively correlated with Bacteroidetes (*Muribaculaceae* family) abundance in the gut. We reasoned that amylase activity differences, that is caused by genetic variantion including single nucleotide polymorphisms and gene copy variants in *Amy1* gene region, lead to different sugar and starch availability in the gut and subsequently modulate the abundance of *Muribaculaceae* bacteria that contain starch utilization genes in their genomes. This hypothesis is supported by the previous evidence that humans with higher amylase gene copy number have more Firmicutes in their gut (Poole et al., 2019). We further examined this idea in B6 and CAST mice, which vary in *Amy* copy number. We found *Muribaculaceae* and bacterial alpha-amylase genes were more abundant in the low amylase gene copy number CAST strain and less abundant in the high amylase gene copy number mouse B6 strain. Further investigation of amylase activity effects on gut microbiome using mouse models need to be done to validate this hypothesis.

We found enterotype associated bacterial species correlated with host cardiometabolic phenotypes, especially the bacterial flagellin, which was causally associated with increased liver

fibrosed area and HDL cholesterol levels. Previous studies showed that gut microbiome partially explained the variations of plasma triglyceride and HDL cholesterol levels in human (Fu et al., 2015). Another study showed that the high-fat diet increased flagellated bacteria in the gut, which increased apolipoprotein A1 (ApoA1) production and HDL cholesterol levels in mice (Yiu et al., 2020). Interestingly, we also observed gut Firmicutes levels and bacterial *fliC* abundance were significantly positively correlated with ApoA1 gene expression levels in liver. Mendelian randomization (MR) seeks to find causal effects between phenotypes. Successful applications of MR in humans revealed the causal relationships between gut microbiome and other molecular traits, including blood metabolites (Liu et al., 2020), short-chain fatty acids (Sanna et al., 2019), and host metabolic traits (Qin et al., 2020; Rühlemann et al., 2021). To our best knowledges, our study is the first MR application of gut microbiome in genetically diverse mouse cohort. Our MR results confirmed the casual relationship between gut flagellated bacteria and plasma HDL cholesterol levels. We further reasoned that not only high-fat diet can increase the flagellated bacteria in the gut, but the amylase gene copy number can also affect flagellated bacteria abundance. A recent study showed bacteria flagellin gene variants from Lachnospiraceae family were associated with TLR5 activation. We also found that nD1 TLR5 epitope motif in *fliC* gene were associated differently with host physiology phenotypes, including atherosclerotic lesion. This indicated the importance of bacterial genetic variations in gut microbiome association studies. A recent study found bacterial SNPs in human gut microbiome were associated with host BMI (Zahavi et al., 2023). Future investigation of bacterial SNPs in the mouse gut microbiome, especially in genetically diverse cohort such as HMDP, is required for a better understanding of their associations with the host.

This thesis presents a comprehensive analysis of host genetic associations to gut microbiome in mouse. Because of the fact that environment dominates over host genetics in shaping gut microbiota (Rothschild et al., 2018), the genetic associations identified in this thesis can be dependent on environmental variables, especially diets. In both Chapter 2 and Chapter 3,

studies were conducted using the same diet among all individuals from the same cohorts. I expect some of the identified genetic associations are diet- dependent. In the future, genetic associations need to be done in different diets to learn the gene by environment (GxE) effects on gut microbiome. Further efforts need to be considered to enhance the gut microbiome-genetic associations research in mouse, using a visualization platform such as that for lipidome study (Linke et al., 2020), or the platform integrating association results across studies and cohorts.

Shotgun metagenomic sequencing of microbial DNA allows us to profile microbial genes. However, genes express differently among the same bacterial genome. Studies showed that the microbiome pathways are transcribed by a limited subset of microorganisms encoding them metagenomically (Abu-Ali et al., 2018) and gene expression varies significantly among subjects with metagenomic concordance (Franzosa et al., 2014). Therefore, RNA-sequencing of gut microbial community (i.e., metatranscriptome) need to be considered for future genetic association study.

The gut microbiome is a high dimensional trait, modulated by environmental and genetic factors. The interplay between gut microbiome and host is complex. The work done in thesis enhances our understanding of effects of host genetics on gut microbiome and provides the framework for mechanisms investigation of gut microbiome-host interactions in the future.

References

Abu-Ali, G. S., Mehta, R. S., Lloyd-Price, J., Mallick, H., Branck, T., Ivey, K. L., Drew, D. A., DuLong, C., Rimm, E., Izard, J., Chan, A. T., & Huttenhower, C. (2018). Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nature Microbiology*, *3*(3), 356–366. https://doi.org/10.1038/s41564-017-0084-4

Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L. J., Thompson, K. N., Zolfo, M., Manghi, P., Dubois, L., Huang, K. D., Thomas, A. M., Nickols, W. A., Piccinno, G., Piperni, E., Punčochář, M., Valles-Colomer, M., Tett, A., Giordano, F., Davies, R., Wolf, J., … Segata, N. (2023). Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nature Biotechnology*. https://doi.org/10.1038/s41587-023-01688-w

Franzosa, E. A., Morgan, X. C., Segata, N., Waldron, L., Reyes, J., Earl, A. M., Giannoukos, G., Boylan, M. R., Ciulla, D., Gevers, D., Izard, J., Garrett, W. S., Chan, A. T., & Huttenhower, C. (2014). Relating the metatranscriptome and metagenome of the human gut. *Proceedings of the National Academy of Sciences*, *111*(22), E2329–E2338. https://doi.org/10.1073/pnas.1319284111

Fu, J., Bonder, M. J., Cenit, M. C., Tigchelaar, E. F., Maatman, A., Dekens, J. A. M., Brandsma, E., Marczynska, J., Imhann, F., Weersma, R. K., Franke, L., Poon, T. W., Xavier, R. J., Gevers, D., Hofker, M. H., Wijmenga, C., & Zhernakova, A. (2015). The Gut Microbiome Contributes to a Substantial Proportion of the Variation in Blood Lipids. *Circulation Research*, *117*(9), 817–824. https://doi.org/10.1161/CIRCRESAHA.115.306807

Kawai, Y., & Akagawa, K. (1989). Macrophage Activation by an Ornithine-Containing Lipid or a Serine-Containing Lipid. *INFECT. IMMUN.*, *57*(7), 2086–2091.

Kawai, Y., Kaneda, K., Morisawa, Y., & Akagawa, K. (1991). Protection of Mice from Lethal Endotoxemia by Use of an Ornithine-Containing Lipid or a Serine-Containing Lipid. *INFECT. IMMUN.*, *59*(8), 2560–2566.

Kemis, J. H., Linke, V., Barrett, K. L., Boehm, F. J., Traeger, L. L., Keller, M. P., Rabaglia, M. E., Schueler, K. L., Stapleton, D. S., Gatti, D. M., Churchill, G. A., Amador-Noguez, D., Russell, J. D., Yandell, B. S., Broman, K. W., Coon, J. J., Attie, A. D., & Rey, F. E. (2019). Genetic determinants of gut microbiota composition and bile acid profiles in mice. *PLOS Genetics*, *15*(8), e1008073. https://doi.org/10.1371/journal.pgen.1008073

Linke, V., Overmyer, K. A., Miller, I. J., Brademan, D. R., Hutchins, P. D., Trujillo, E. A., Reddy, T. R., Russell, J. D., Cushing, E. M., Schueler, K. L., Stapleton, D. S., Rabaglia, M. E., Keller, M. P., Gatti, D. M., Keele, G. R., Pham, D., Broman, K. W., Churchill, G. A., Attie, A. D., & Coon, J. J. (2020). A large-scale genome–lipid association map guides lipid identification. *Nature Metabolism*, *2*(10), 1149–1162. https://doi.org/10.1038/s42255-020-00278-3

Liu, X., Tong, X., Zou, Y., Lin, X., Zhao, H., Tian, L., Jie, Z., Wang, Q., Zhang, Z., Lu, H., Xiao, L., Qiu, X., Zi, J., Wang, R., Xu, X., Yang, H., Wang, J., Zong, Y., Liu, W., … Zhang, T. (2020). Mendelian randomization analyses support causal relationships between blood metabolites and the gut microbiome. *Nature Genetics*. https://doi.org/10.1101/2020.06.30.181438

Org, E., Parks, B. W., Joo, J. W. J., Emert, B., Schwartzman, W., Kang, E. Y., Mehrabian, M., Pan, C., Knight, R., Gunsalus, R., Drake, T. A., Eskin, E., & Lusis, A. J. (2015). Genetic and environmental control of host-gut microbiota interactions. *Genome Research*, *25*(10), 1558–1569. https://doi.org/10.1101/gr.194118.115

Poole, A. C., Goodrich, J. K., Youngblut, N. D., Luque, G. G., Ruaud, A., Sutter, J. L., Waters, J. L., Shi, Q., El-Hadidi, M., Johnson, L. M., Bar, H. Y., Huson, D. H., Booth, J. G., & Ley, R. E. (2019). Human Salivary Amylase Gene Copy Number Impacts Oral and Gut Microbiomes. *Cell Host & Microbe*, *25*(4), 553-564.e7. https://doi.org/10.1016/j.chom.2019.03.001

Qin, Y., Havulinna, A. S., Liu, Y., Jousilahti, P., Ritchie, S. C., Tokolyi, A., Sanders, J. G., Valsta, L., Brożyńska, M., Zhu, Q., Tripathi, A., Vazquez-Baeza, Y., Loomba, R., Cheng, S., Jain, M., Niiranen, T., Lahti, L., Knight, R., Salomaa, V., … Méric, G. (2020). Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. *Nature Genetics*. https://doi.org/10.1101/2020.09.12.20193045

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P. I., Godneva, A., Kalka, I. N., Bar, N., Shilo, S., Lador, D., Vila, A. V., Zmora, N., Pevsner-Fischer, M., Israeli, D., Kosower, N., Malka, G., Wolf, B. C., … Segal, E. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature*, *555*(7695), 210–215. https://doi.org/10.1038/nature25973

Rühlemann, M. C., Hermes, B. M., Bang, C., Doms, S., Moitinho-Silva, L., Thingholm, L. B., Frost, F., Degenhardt, F., Wittig, M., Kässens, J., Weiss, F. U., Peters, A., Neuhaus, K., Völker, U., Völzke, H., Homuth, G., Weiss, S., Grallert, H., Laudes, M., … Franke, A. (2021). Genome-wide association study in 8,956 German individuals identifies influence of ABO histo-blood groups on gut microbiome. *Nature Genetics*, *53*(2), 147–155. https://doi.org/10.1038/s41588-020-00747-1

Sanna, S. (2022). Challenges and future directions for studying effects of host genetics on the gut microbiome. *Nature Genetics*, 7.

Sanna, S., van Zuydam, N. R., Mahajan, A., Kurilshikov, A., Vich Vila, A., Võsa, U., Mujagic, Z., Masclee, A. A. M., Jonkers, D. M. A. E., Oosting, M., Joosten, L. A. B., Netea, M. G., Franke, L., Zhernakova, A., Fu, J., Wijmenga, C., & McCarthy, M. I. (2019). Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. *Nature Genetics*, *51*(4), 600–605. https://doi.org/10.1038/s41588-019-0350-x

Yiu, J. H. C., Chan, K.-S., Cheung, J., Li, J., Liu, Y., Wang, Y., Fung, W. W. L., Cai, J., Cheung, S. W. M., Dorweiler, B., Wan, E. Y. F., Tso, P., Xu, A., & Woo, C. W. (2020). Gut

Microbiota-Associated Activation of TLR5 Induces Apolipoprotein A1 Production in the Liver. *Circulation Research*, *127*(10), 1236–1252. https://doi.org/10.1161/CIRCRESAHA.120.317362

Zahavi, L., Lavon, A., Reicher, L., Shoer, S., Godneva, A., Leviatan, S., Rein, M., Weissbrod, O., Weinberger, A., & Segal, E. (2023). Bacterial SNPs in the human gut microbiome associate with host BMI. *Nature Medicine*. https://doi.org/10.1038/s41591-023-02599-8

APPENDIX A: Gut bacterial metabolism contributes to host global purine homeostasis

The work presented in this appendix has been published:

*Ksahara K[*], Kerby RL[*], **Zhang Q**, Pradhan M, Mehrabian M, Lusis AJ, Bergström G, Bäckhed F, Rey FE[+]*

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

Data and supplementary information available online

A.1 Abstract

The microbes and microbial pathways that influence host inflammatory disease progression remain largely undefined. Here, we show that variation in atherosclerosis burden is partially driven by gut microbiota and is associated with circulating levels of uric acid (UA) in mice and humans. We identify gut bacterial taxa spanning multiple phyla, including Bacillota, Fusobacteriota, and Pseudomonadota, that use multiple purines, including UA as carbon and energy sources anaerobically. We identify a gene cluster that encodes key steps of anaerobic purine degradation and that is widely distributed among gut-dwelling bacteria. Furthermore, we show that colonization of gnotobiotic mice with purine-degrading bacteria modulates levels of UA and other purines in the gut and systemically. Thus, gut microbes are important drivers of host global purine homeostasis and serum UA levels, and gut bacterial catabolism of purines may represent a mechanism by which gut bacteria influence health.

A.2 Introduction

Metabolic disorders including obesity, type 2 diabetes, and atherosclerosis have historically been viewed as lipid conditions primarily driven by overindulgence of calorie-dense foods. However, it is now widely appreciated that chronic inflammation plays a central role in the development and progression of these disorders (Wolf & Ley, 2019). Atherosclerosis, the leading cause of cardiovascular disease (CVD), is characterized by vascular inflammation and is influenced by multiple genetic and environmental factors (Lusis, 2000; Stylianou et al., 2012; Wolf & Ley, 2019). Large-scale genome-wide analyses in human populations have identified over 100 loci significantly associated with atherosclerosis itself (Kessler & Schunkert, 2021) and hundreds of additional loci for traits associated with atherosclerosis such as plasma lipids, obesity, and diabetes (MAGIC et al., 2010; Selvaraj et al., 2022; Xue et al., 2018, p. 143). Nevertheless, while genetics significantly influences atherosclerosis, the environment, especially diet, also plays a major role in its progression. Furthermore, several recent studies have provided evidence suggesting that dietary contributions to disease progression are often mediated by the gut microbiome (Kasahara et al., 2018; Koeth et al., 2013; Wang et al., 2011).

The gut microbiome exerts profound influence on metabolism and inflammatory diseases (Brandsma et al., 2019; Tilg et al., 2020). Diet and host-derived factors modulate the composition of the gut microbiome, which in turn transforms dietary components consumed by the host, generating bioactive molecules that interact with the immune system and virtually every host organ, including the vascular system. Changes over the last century in food production, dietary habits, antibiotic usage, and lifestyle have caused major changes in the microbiome and have affected human health in discordant ways (Bolte et al., 2021; De Filippo et al., 2010; Modi et al., 2014): the prevalence of acute infectious diseases has decreased, while it increased for chronic inflammatory diseases. Furthermore, several diet-derived gut-bacterially produced metabolites have been uncovered as potential drivers of metabolic and cardiovascular ailments. These metabolites constitute a direct link between environmental exposures and host cellular function

and encompass several uremic toxins (i.e., waste products that cannot be eliminated properly by subjects with impaired kidney function), including p-cresyl sulfate, indoxyl sulfate, trimethylamine-N-oxide (TMAO), and phenylacetylglutamine (Beker et al., 2022). For example, dietary choline, betaine, and carnitine serve as substrates for TMAO production, which is generated in the liver from gut-bacterially produced trimethylamine (TMA) (Wang et al., 2011). TMAO enhances inflammation and aortic thrombosis in mice and it is associated with CVD risks in humans (Tang et al., 2013; Wang et al., 2011; W. Zhu et al., 2016). More recently it was found that phenylacetylglutamine, synthesized by the microbiota from dietary protein, enhances platelet activation and thrombosis via host G-protein-coupled receptors (Nemet et al., 2020). Together, this evidence supports the notion that gut bacteria metabolism contributes to CVD-related traits by modulating abundance of uremic toxins in circulation.

Several purines, including xanthine, hypoxanthine, and uric acid (UA) are also considered uremic toxins (Falconi et al., 2021) and contribute to several symptoms observed in subjects with chronic kidney disease (Beker et al., 2022; Falconi et al., 2021). UA—the end product of the metabolic breakdown of purines in humans—is mostly studied for the complications it causes when its concentration reaches saturation levels, forming pro-inflammatory crystals that deposit in joints (e.g., gout). However, a recent study showed that concentrations of UA within the solubility range can promote atherosclerosis via induction of AMP-activated protein kinase (AMPK)-mediated inflammation (Kimura et al., 2020). UA exacerbates inflammation, endothelial dysfunction, increases the renin-angiotensin-aldosterone system activity (Falconi et al., 2021), and it is increased in patients with hypertension and heart failure (Falconi et al., 2021). Several studies have suggested that pharmacological interventions effective at reducing UA production or increasing its excretion in hyperuricemic patients improves cardio-renal outcomes (Weisman et al., 2019), although these benefits are not consistently observed. While the kidneys play a major role in regulating levels of UA in circulation, a significant fraction of this metabolite is secreted into the intestine (Méndez-Salazar & Martínez-Nava, 2022), and a recent metagenomic

study identified bacterial pathways associated with blood levels of UA (Chu et al., 2021). However, no causal relationships have been established between abundance of this uremic toxin and specific gut bacteria, and although the bacterial mechanisms of anaerobic purine metabolism have been studied biochemically, their genetic underpinnings remains undefined (Hartwich et al., 2012).

We sought to examine the role of the gut microbiome on atherosclerosis and identify potential microbial pathways that contribute to disease burden. First, we transplanted microbial communities derived from mouse strains with disparate atherosclerosis phenotypes into germ-free (GF) *Apolipoprotein E* knockout (*ApoE* KO) mice. We found that microbial-driven variation on atherosclerosis progression was associated with abundance of purine metabolites including UA. We also observed that this pro-inflammatory metabolite was associated with atherosclerosis burden and gut microbial features in a human cohort. We identified bacterial taxa able to degrade purines anaerobically, uncovered a gene cluster encoding key components needed for anaerobic purine degradation, demonstrated environmental factors affecting its activity, and showed that colonization with taxa containing this locus lowered multiple purines in the gut and UA systemically in mice. Altogether, this work strengthens the connection between gut microbes and atherosclerosis and provides insights into how bacterial metabolism influences host biology.

A.3 Results

Gut microbes modulate atherosclerosis progression and plasma metabolites associated with disease burden in mice

Previous work revealed a large degree of variation in atherosclerosis burden among 100 inbred strains of mice from the Hybrid Mouse Diversity Panel (HMDP), which harbor distinct microbial communities (Bennett et al., 2015; Org et al., 2015). We hypothesized that the gut microbiome contributed to the variation observed in disease progression among HMDP strains. We transplanted cecal samples from four HMDP strains into GF *ApoE* KO recipient mice. We selected two strains that exhibited large atherosclerotic lesions (AXB10/PgnJ and BXD5/TyJ) and two strains that showed little signs of disease (BTBR T+tf/J and BXA8/PgnJ), hereinafter referred as “AXB10,” “BXD5,” “BTBR,” and “BXA8,” respectively. Transplanted mice were maintained on a chow diet supplemented with 0.2% cholesterol for 8 weeks. After this period, atherosclerotic lesions, gut microbiome composition, and disease biomarkers were evaluated (Figure A.1a and Supplemental Figure A.1). We found that mice colonized with cecal communities from HMDP donors prone to atherosclerosis development (i.e., AXB10 and BXD5) exhibited larger lesions compared with recipient mice colonized with samples from donors that showed little signs of atherosclerosis (i.e., BTBR and BXA8). These results support the notion that the gut microbiome contributes to the development of atherosclerosis and possibly to the variation in disease burden observed among the HMDP strains (Figure A.1b–g). Neither traditional CVD risk factors such as body weight and cholesterol, nor previously identified gut microbiota-derived metabolites including lipopolysaccharides (LPSs), TMAO, and short-chain fatty acids explained the differences in atherosclerosis burden observed among the transplanted mice (Supplementary Figure A.1).

Shotgun metagenomic analyses of cecal contents from the transplanted mice identified 1,649 functional features and 52 bacterial taxonomic features. Principal component analysis of functional features shows distinct clustering by donor strain (p value <0.001, PERMANOVA),

suggesting unique gut bacterial function profiles for each of the four HMDP strains used (Figure A.1). Pearson correlation analysis identified several bacterial functions associated with atherosclerosis lesion size among recipient mice (Supplementary Figure A.2a,b). These included several functions related to production and conversion of purines. We also observed that bacterial pathways including energy metabolism and amino acid metabolism were enriched among functions positively correlated with atherosclerosis lesion size, although these did not survive multiple hypothesis correction.

To further investigate whether microbiota transplants impacted circulating metabolites associated with disease, we performed metabolome analysis of plasma samples using ultra-high performance liquid chromatography (uHPLC)-tandem mass spectrometry (MS/MS). A total of 682 metabolites were measured (Table S1). Pearson correlation analyses identified purine metabolites including xanthine, xanthosine, inosine, and UA, positively associated with atherosclerotic lesions size (Figure A.1i,j). Altogether, these results may suggest that gut microbes influence atherosclerosis progression and abundance of purines in the blood of transplanted mice.

Serum UA is correlated with gut microbial features and subclinical atherosclerosis in a human cohort

UA is the end product of purine metabolism in humans, and it has been shown to cause inflammation (Braga et al., 2017; Kimura et al., 2020; Martinon et al., 2006), induce endothelial dysfunction (Khosla et al., 2005), and stimulate smooth muscle cell proliferation (Rao et al., 1991). We explored associations between atherosclerosis, gut bacteria, and UA in a human cohort previously characterized for gut microbiome and glucose homeostasis (n = 998) (Wu et al., 2020). Coronary artery calcium (CAC) score measurements were assessed for disease burden. Calcification of arteries is an accepted proxy for estimating overall plaque burden of atherosclerosis. We first classified individuals based on their CAC score status: CAC score = 0

(i.e., no detectable vascular calcification, n = 492) vs. CAC > 0 (n = 497). Logistic regression analysis revealed that distribution of UA concentrations was significantly different between individuals from these two groups (Figure A.2a) with individuals with CAC > 0 showing higher mean and median UA levels. Furthermore, Spearman correlation analysis for individuals with CAC score >0 showed a significant positive association between UA and CAC score (rho coefficient = 0.14, p value < 0.001, Figure A.2b). We then applied extreme gradient boosting (XGBoost) regression to identify gut bacterial taxa correlated with UA levels. XGBoost is a decision-tree-based ensemble machine learning algorithm that uses the gradient boosting method. The top 10 features associated with UA levels are shown in Figure A.2c and Table S2 after adjusted for covariates (CAC score, Body Mass Index, gender, triglycerides, and HemoglobinA1c) in a mixed linear regression model. Interestingly, we found multiple taxa within the *Clostridia* class that were negatively associated with levels of UA. Altogether, these results suggest that the gut microbiome, particularly taxa within the *Clostridia*, may influence UA levels. These data are also consistent with gnotobiotic mouse work reported above and previous work connecting UA with CVD in humans (Agarwal et al., 2013; Drivelegka et al., 2020; Rahimi-Sakak et al., 2019; Sun et al., 2015).

Gut microbiome modulates purines in cecum and circulation

We next investigated whether the gut microbiome modulated abundance of purines in the intestine and circulation. We quantified purine-related metabolites, including nucleotides, nucleosides, and nucleobases by liquid chromatography-tandem mass spectrometry (LC-MS/MS) in cecal contents and plasma from GF mice and conventionally raised (Conv) animals (Tables S3A and S3B). We found that most purines were decreased in the cecal contents from GF mice compared with Conv mice, with a few exceptions being increased in GF mice, especially UA and allantoin, both terminal purine metabolites in mice (Figures A.3a,b). Given that cecal/fecal purines

arise from diet type, bacterial and host turnover, and metabolism, these analyses cannot establish the sources and fates of these compounds in the gut.

Partial least squares-discriminant analysis (PLS-DA) of cecal purines showed separation of the two groups by principal components 1 and 2. While the separation for plasma samples was less evident (Supplementary Figures A.3a,b), we found that GF mice had significantly increased UA levels in plasma compared with Conv mice (Supplementary Figures A.3a,c). This result was confirmed using an enzymatic assay to quantify UA (Figure A.3c). These results again suggested that the gut microbiome modulates abundance of purines both in the gut and systemically and was the impetus for attempts to isolate anaerobic purine-degrading bacteria (PDB).

Human gut bacteria degrade purines anaerobically

The intestine is a key organ for purine homeostasis. Dietary purines are absorbed in the gut, resident microbes produce and recycle purines needed for their anabolism and ~30% of the UA generated by the body is secreted into the intestine (Sorensen & Levinson, 1975; Yun et al., 2017). We hypothesized that gut bacteria influence purine levels by metabolizing them to non-purine products. While this notion has been previously discussed, isolates from the human or mouse gut able to grow on these metabolites anaerobically have not been identified. We attempted anaerobic enrichments using fecal slurries (human samples) on media supplemented with UA as the primary source of carbon and energy. Culture medium included non-fermentable acetate (often beneficial for butyrate-producing Firmicutes) plus 0.1% yeast extract as the sole complex nutrient. Enrichments were plated and colonies isolated on bilayer agar plates bearing a top agar layer supplied with saturating amounts of UA or other purines as described in the Methods section. We obtained several isolates from the Bacillota (Firmicutes) and Pseudomonadota (Proteobacteria) phyla including species identified as *Enterocloster bolteae* and *Escherichia coli* by 16S rRNA gene sequencing. Additional confirmatory assays were used to verify the identity of the *E. coli* isolate (designated “I-11,” see Methods).

Using the same medium, we verified growth and UA utilization by *E. bolteae* ATCC BAA-613—the species type strain, which has been previously sequenced. We observed that 10-ml cultures of this strain supplemented with 12-mg UA degraded 49.6 µmol of substrate in a 24-h period, with accumulation of 106.9 µmol acetate, a known fermentation product of this organism (Song et al., 2003). HPLC and headspace gas chromatography (GC) analysis did not detect significant ethanol, formate, propionate, lactate, and butyrate accumulation under this condition.

We then expanded our search for bacteria with this capability and by using methods described above screened a culture collection of 34 isolates encompassing gut-dwelling bacteria from six phyla (Supplementary Figure A.4; a representative subset of these species is shown in Figure A.4). The screen utilized monolayer plates containing no nitrogen (other than 0.1% yeast extract) or fermentable carbon or energy source, the same medium supplemented with soluble substrates (glucose or allantoin), or bilayer plates (briefly described above, see also Methods) containing insoluble purines (UA, adenine, or hypoxanthine). These assays demonstrated the anaerobic allantoin- and purine-dependent growth characteristics of these strains as evident by growth of the applied bacterial patch and a zone of disappearance of the insoluble purine substrates. Our screen showed evidence of purine utilization among the Bacillota (Firmicutes), Fusobacteriota, and Pseudomonadota (Proteobacteria) phyla (Figures A.4 and Supplementary Figure A.4), although this property was not universal among strains belonging to these phyla. Of particular note, these assays showed different purine utilization capacities among UA-degrading strains: (1) *E. coli* MS 200-1 and lab isolate *E. coli* I-11 showed greater UA utilization compared with the commonly used *E. coli* K12 strain (Figure A.4); (2) allantoin supported growth of *E. bolteae* and *E. coli* but not medically-relevant *Clostridioides difficile* or *Edwardsiella tarda* (Figure A.4); (3) adenine supported the growth of several strains of *Proteobacteria* (*E. coli* MS200-1, *E. coli* I-11, *E. tarda*) and *Firmicutes* (*E. bolteae*, *E. asparagifome*), but not *C. difficile*, *C. sporogenes* or *E. coli* K12) (Figures A.4 and Supplementary Figure A.4); (4) consistent with previous work, growth of *Proteobacteria* strains on UA was enhanced by the addition of formate (Supplementary

Figure A.4) (Iwadate & Kato, 2019), while the addition of formate had more modest effects, if any, for strains from other phyla tested, and did not appear to enhance the utilization of the more reduced purines (adenine, hypoxanthine, Supplementary Figure A.5a) in any of the strains tested. In multiple organisms, the ability to use UA was diminished or eliminated in media containing glucose or fructose, consistent with regulation by catabolite repression (Supplementary Figure A.4 and Supplementary Figure A.5b). It is also important to note that none of the six *Bacteroides* strains tested showed any growth on UA, allantoin, or adenine (Figures A.4 and Supplementary Figure A.4).

Extensive biochemical analyses of environmental clostridial isolates has previously demonstrated molybdenum and selenium dependences for purine metabolism (Dürre & Andreesen, 1982, 1983; Schiefer-Ullrich et al., 1984). While no attempt was made to limit trace minerals—i.e., the medium contained a high amount of phosphate buffer plus low levels of yeast extract and cysteine (a possible source of Se contamination), and the inocula were prepared in rich medium—a requirement for micromolar additions of Mo and Se was evident for both species of Proteobacteria tested on UA (Supplementary Figure A.5b). However, these effects were not observed with the two Firmicutes tested, *C. difficile* and *E. bolteae* (Supplementary Figure A.5b). Additional assays did not show an effect of Fe, Co, Mn, Ni, W, or Zn supplementation on UA utilization by any of the strains tested (*E. bolteae*, *C. difficile*, *E. coli* MS 200-1, *E. tarda*, again the media were not rigorously depleted of these metals). Altogether, the results presented above suggested that common gut bacteria can use purines for carbon and energy, and that availability of other carbon sources and metals could modulate this process.

PDB modulate abundance of purines in cecum and circulation

To test the impact of PDB identified above in vivo, we created synthetic bacterial communities that varied in their capacity to degrade UA and used them to colonize GF mice. We colonized GF mice with a core community that included seven species spanning major phyla from

the human gut and that did not degrade purines in vitro as determined by the assays described above (and as described below they lack genes encoding functions necessary for anaerobic purine metabolism). These included *Bacteroides caccae*, *Bifidobacterium dentium*, *Blautia hansenii*, *Bacteroides thetaiotaomicron*, *Coprococcus comes*, *Mitsuokella multacida*, *and Ruminococcus torques*. Half of the animals were also colonized with three PDB, including *E. bolteae*, *Hungatella hathewayi*, and *E. coli* isolate I-11 (Figure A.5a). In vitro tests of UA utilization for each of the strains used in these communities are shown in Figure A.4 (highlighted in red), and their combined activities measured using fecal innocula verified that colonization with PDB was required for UA degradation (Figure A.5a). A third group of animals remained GF throughout the experiment. The engrafted bacterial communities were analyzed by COPRO-seq (community profiling by sequencing) analysis (Faith et al., 2011). All taxa included in these communities successfully colonized the gut of GF mice. *E. bolteae* was the most abundant among the three PDB (relative abundance: *E. bolteae* 19.5%, *H. hathewayi* 5.6%, and *E. coli* 0.9%; Figure A.5b). We next performed targeted quantification of purines, pyrimidines, and related metabolites in cecal contents and plasma (Tables S4A and Supplementary Figure A.4b). Global analysis of the data showed distinct patterns for cecal purine-related metabolites between GF mice and mice with the “core” or the “core plus PDB,” where nucleosides were increased, but nucleotides were decreased in the cecum of GF mice (Figure A.5c). Surprisingly, mice colonized with the core plus PDB community showed significantly higher levels of UA in the cecum compared with mice colonized with the core community, while cecal levels of other purines/nucleosides including hypoxanthine and allantoin were significantly reduced in mice co-colonized with the three PDB (Figure A.5d). It is important to note that these metabolites were detected at significantly higher concentrations in the gut of mice relative to UA (Figure A.5d; Table S4A). PLS-DA analysis separated plasma samples from the core plus PDB from the ones from the other two groups (Supplementary Figure A.6a,b). Colonization with core plus PDB resulted in consistently lower levels of several purine metabolites in plasma, including UA (Supplementary Figure A.6a,c),

recapitulating previous data (Figure A.3c). Plasma UA results were again confirmed by an enzymatic assay (Figure A.5e). Altogether these results suggested that PDB impacted levels of several purines in the gut and specifically UA both locally and in circulation.

Transcriptional analysis identifies bacterial genes required for anaerobic growth on multiple purines

Having established the role of PDB on lowering levels of UA systemically and having identified gut isolates capable of anaerobic purine metabolism, we sought to identify genes encoding these functions. Cultures of *E. bolteae* were cultivated in medium supplemented with UA or xylose plus NH4Cl (henceforth, “xylose”). We selected xylose for comparison as the growth rate of *E. bolteae* on this substrate was similar to that of UA, with doubling times of 2.6 h for xylose and 4.6 h for UA. For both substrates, log-phase cells were harvested, and libraries subjected to sequencing. We obtained ~3.6 × 107 reads/sample, of which 99.2% mapped to the *E. bolteae* genome. Figure A.6a shows reads per million (RPM) normalized to gene size plotted against the relative expression level for growth on the two substrates, limited to the 3,217 (of 5,993) differentially-expressed genes (FDR < 0.01, Table S5). As expected, genes encoding 30S and 50S RNA polymerase (RNAP) subunits show a slight bias (1.6-fold, Supplementary Figure A.7) toward the xylose substrate side, consistent with the faster growth rate observed on this substrate and the rate-limiting nature of RNAP subunit expression (Gaal et al., 1997). Growth on UA promoted higher expression of 51 genes relative to xylose (cutoff > 25-fold; Figure A.6a), including several predicted to encode micronutrient transport functions, one glycine-cleavage system, and a probable electron bifurcating hydrogenase (Supplementary Figure A.7).

Two adjacent and divergently oriented putative operons, each encoding 6 genes, amounted to 12% of all RNA-seq reads in *E. bolteae* grown on UA. These highly upregulated genes are indicated by the filled blue and red circles with their corresponding operons and putative gene products diagrammed (Figures A.6a,b). Notably, the encoded proteins are predicted to

catalyze C-N cleavage and Se-dependent hydrolytic reactions. Also indicated are a purine permease, likely specific for UA uptake based on the conservation of residues S99 and S314 (T100 and S317 in E. coli UacT) (Papakostas et al., 2013) and a knotted carbamoyl transferase/carbamate kinase, presumably required for ATP synthesis (Y. Li et al., 2011).

Alignments of conserved chromosomal regions of purine-fermenting organisms illustrated conservation of five genes (*E. coli* nomenclature: *dpaL*, *hydA*, *ssnA*, *ygeY*, and *xdhD*) among all taxa, although not with a conserved organization nor exclusively present in a contiguous genomic region across phyla (Figure A.6c). A variant of *E. coli* MS 200-1 bearing a deletion of the *ygeW-dpaL-ygeY-hydA-arcC* operon grew as the wild-type strain in medium supplied with glucose or allantoin but was unable to grow anaerobically using UA, adenine, or hypoxanthine as the carbon and energy source. Conversely, a variant of *E. coli* MS 200-1 bearing a deletion of *allB*, encoding the enzyme catalyzing the first step of anaerobic allantoin metabolism (Cusa et al., 1999), was unable to utilize allantoin but retained the ability to catabolize UA, adenine, and hypoxanthine, indicating distinct mechanisms of allantoin and purine metabolism in this organism (Figure A.6d), and in good agreement with low expression level of all genes in UA-grown *E. bolteae* cells. Finally, a variant of *E. coli* MS 200-1 bearing a deletion of the *ygeV* gene (Figure A.6c) encoding a single-component sigma 54-type transcription factor grew normally on glucose and allantoin but failed to grow on all tested purines (Figure A.6d), consistent with previous data indicating transcription regulation of the adjacent *ygeW* operon by YgeV (Iwadate & Kato, 2019).

To assess the role of this bacterial gene cluster involved in purine metabolism in vivo, we used the same approach as described for Figure A.5, where GF mice were colonized with the core community which lacks PDB, the “core community plus *E. coli* MS 200-1 wild-type,” or the “core community plus the deletion variant FER041” (Δ(*ygeW-arcC*)::*tetA-sacB*) (Figure A.6e). Both the wild-type and the variant showed comparable levels of colonization in the gut under the conditions tested (Figure A.6f). Colonization with the wild-type strain resulted in lower levels of plasma UA compared with mice colonized with the core or the deletion variant FER041 (Figure

A.6g). Collectively, these results suggest that the bacterial functions encoded by these genes contribute to UA homeostasis in vivo.

Detection of genes encoding purine degradation functions in bacterial genomes and transplanted mice

Having identified genes required for anaerobic purine metabolism and UA homeostasis in mice we then sought to identify bacterial taxa containing these genes. We performed BLASTP of the NCBI RefSeq Genome Database (refseq_genomes) using parameters described in Methods. We detected 230 non-redundant bacterial taxa that had the five genes reliably detected among all experimentally confirmed purine-degrading taxa (*dpaL, hydA, ssnA, ygeY*, and *xdhD*). These potential UA degraders included bacterial taxa belonging to Actinobacteria, Firmicutes, Proteobacteria, Fusobacteria, and Spirochaetes (Table S6).

Lastly, we assessed the abundance of these genes in the cecum of *ApoE* gnotobiotic mice (Figures A.1 and A.7a) and correlated their abundance with levels of purine-related metabolites quantified in their cecum (Table S7). We found that cecal levels of several purine-related metabolites including deoxyxanthosine, xanthosine, and UA were negatively associated with the abundance of genes involved in anaerobic purine degradation (Figure A.7b). Altogether these results highlight the potential of these genes as biomarkers for purine breakdown in the gut. Understanding how to manipulate the representation and function (as opposed to abundance of relevant genes) of purine-consuming species in the intestinal microbiota could potentially lead to means for preventing or treating hyperuricemia and associated conditions, with caveats (see Discussion).

A.4 Discussion

In this study, we sought to identify microbially regulated metabolites involved in atherosclerosis progression. Our initial gnotobiotic mouse transplant studies revealed that levels of several purines including UA were influenced by gut microbes and associated with atherosclerosis burden. These initial results motivated a deeper exploration into the role of the gut microbiome on purine homeostasis and led us to identify bacteria able to break down purines anaerobically, and to uncover a cluster of bacterial genes required for anaerobic catabolism of these substrates. We also demonstrated that taxa encoding these functions lowered circulating UA levels in mice. Altogether this work has implications beyond atherosclerosis: it provides insights into how gut bacterial metabolism may influence UA in the circulation and suggests that microbes able to catabolize purines anaerobically are important drivers of host purine homeostasis both locally (i.e., gut) and systemically. A recent preprint using complementary approaches, including hyperuricemic humans and mutagenesis in Clostridia species, identified the same bacterial gene cluster associated with UA degradation, and resulted in similar conclusions with regards to its role on host UA homeostasis (Liu et al., 2022).

In humans, nearly two-thirds of purines are endogenously produced, whereas the remainder comes from diet, which are primarily absorbed in the duodenum, but absorption could occur in the large intestine as transporters of nucleosides are expressed in epithelial cells (Errasti-Murugarren et al., 2007; Fernández-Calotti et al., 2016; Pastor-Anglada & Pérez-Torras, 2018), and their abundance increases when demand for purines increases, for example, in colon cancer (Naes et al., 2023). Absorbed purines can be used by enterocytes or colonocytes or be degraded to UA. In most mammals, UA can be further metabolized to allantoin, but in hominoids the presence of an uricase gene bearing multiple mutations and premature stop codons results in the accumulation of UA (Kratzer et al., 2014). Since UA is relatively insoluble, humans are susceptible to diseases resulting from precipitation of UA including gout and kidney stones. In fact, the prevalence of hyperuricemia in the US is ~20% among adults and has been steadily increasing

in recent decades (Chen-Xu et al., 2019; G. Singh et al., 2019). This increase may be related to the prevalence of a high-purine diets, fructose beverages (known to increase UA levels), and alcohol consumption (Choi, 2010).

The association between UA and CVD has been demonstrated for several conditions such as hypertension (Agarwal et al., 2013), chronic kidney diseases, metabolic syndrome (Sun et al., 2015), atherosclerosis (Drivelegka et al., 2020), and adverse cardiovascular outcomes (Rahimi-Sakak et al., 2019), even for UA levels in the normal to high range (from 5.2 to 6 mg/dL) (Feig et al., 2008). Several mechanisms have been proposed to explain the role of UA in CVD, including endothelial dysfunction, systemic inflammation, and renin-angiotensin-aldosterone system activation. Our results in gnotobiotic *ApoE* KO mice (Figure A.1) and humans (Figure A.2) support this notion and suggest that levels of UA within the soluble range may contribute to disease progression. This is consistent with a recent mechanistic study suggesting that soluble UA activates the NLRP3 (NOD- LRR- and pyrin-domain-containing protein 3) inflammasome.22 However, whether elevated serum UA levels are an independent risk factor for CVD remains controversial. There is conflicting evidence regarding the benefits of UA-reducing strategies for treating patients with CVD (Ju et al., 2020; Mackenzie et al., 2022; J. A. Singh et al., 2017). Importantly, the beneficial effects of UA-lowering therapies on CVD are not yet established in large-scale randomized trials. Moreover, several Mendelian randomization studies showed mixed results for causal effects of serum UA on CVD outcomes (Keerman et al., 2020; J. Zhu et al., 2022). These inconsistent results are also observed in experimental animals. For example, in one study increasing the levels of UA with a high-purine diet did not change the development of atherosclerosis (Wakuda et al., 2014), while in a second study, lowering UA levels by administration of xanthine oxidase inhibitor or using uricase transgenic mice reduced the atherosclerosis development (Kimura et al., 2020). Interestingly, the latter showed that UA promoted the production of the inflammatory cytokine IL-1b only in the presence of LPS, suggesting that the effects of UA on the cardiovascular system may be context dependent.

Collectively, more work is needed to clarify the contribution of hyperuricemia to atherosclerosis in humans: our study introduces an unappreciated factor, the microbial composition and nutritional milieu of the gut as discussed below.

The kidneys play a critical role in maintaining plasma UA levels through complex transport systems that mediate both reabsorption and secretion of UA. Intestinal secretion is a substantial contributor to extra-renal elimination of UA, accounting for about one-third of total elimination of UA (Sorensen & Levinson, 1975; Yun et al., 2017). Our results showing that PDB lower the abundance of some purines in the intestine (Figure A.5) suggest that these organisms may lower circulating UA levels by decreasing the burden of purines bioavailable to the host. Alternatively, the increased levels of UA detected in the cecum of mice colonized with PDB (Figure A.5) may suggest that in the presence of these organisms, consumption of certain purines may trigger more secretion of UA into the intestine. More studies are needed to clarify how PDB influences gut and systemic levels of purines. Given the differential purine-metabolizing capabilities and environmental modulation of purine consumption among the examined PDB, this may vary widely depending on the specific PDB colonizing an individual.

Anaerobic purine utilization by bacteria was first described over 100 years ago, yet relatively few species have been identified and these isolates were obtained from environmental sources and are obligate purinolytic Firmicutes. The biochemistry of their purine catabolic pathway was delineated prior to the advent of genetic manipulations and only in the past decade have their genomes been sequenced, although the genetic elements encoding the process remain undefined (Hartwich et al., 2012). More nutritionally diverse anaerobic purine-utilizing organisms, including Firmicutes and Proteobacteria isolated from termite intestines, metabolize purines and evidently recycle purine nitrogen in the nitrogen-limited host diet, but further biochemical and genetic analyses of these organisms have not been published (Tydell et al., 2002). Operons encoding proposed pathways of anaerobic purine degradation have been identified through in silico analyses (Barba et al., 2013; Haft & Self, 2008), through overexpression

of individual genes (Y. Li et al., 2011; Uo et al., 2002), or by using cultures and cell suspensions cultivated in complex media containing additional carbon and energy sources (Iwadate & Kato, 2019). While these efforts identified the appropriate genomic regions encoding functions necessary for anaerobic purine catabolism, the appropriate conditions allowing reproducible growth on multiple purines were not formulated, and a full understanding of the encoded metabolism has remained elusive.

We asked to what extent UA could serve as a source of carbon and energy for gut bacteria, and to what extent the gut microbiota composition might affect host systemic purine concentrations. We tested 34 human gut isolates from six phyla cognizant of (1) the demonstrated requirements of trace elements (Mo and Se) for metabolic function (Schiefer-Ullrich et al., 1984) and (2) the well-known phenomenon of catabolite repression reported for the yge operon of E. coli and likely present in other bacteria (Z. Li et al., 2019). We identified representatives for three different phyla, primarily belonging to the *Firmicutes* and *Proteobacteria*, that readily degraded UA. A limited subset of organisms was shown to utilize adenine and hypoxanthine, and, independent of the ability to use UA or adenine, some organisms including strains of *E. coli* grew anaerobically in medium supplied with allantoin (Figures A.4, Supplementary Figure A.4, and Supplementary Figure A.5a), in contrast to reports describing that this compound only serves as a source of nitrogen (Cusa et al., 1999). However, these properties were not consistently present in any taxonomic group, and even strain differences between species were identified. Moreover, the presence of an alternative carbon source (glucose and/or fructose) reduced or eliminated UA metabolism, confirming transcriptional regulation by catabolite repression, and indicating one nutritional parameter that could modulate purine utilization in the gut.

In two Proteobacteria, the ability to utilize UA was influenced by the presence of Se and Mo (Supplementary Figure A.5b). However, the effects of metals were not evident in *C. difficile* and *E. bolteae*, likely reflecting the high expression of multiple metal uptake systems identified in the RNA-seq data for *E. bolteae* (Supplementary Figure A.7) and the requirement for multiple

metal-limited transfers to affect mineral limitation in other purinolytic Firmicutes (Dürre & Andreesen, 1983; Schiefer-Ullrich et al., 1984; Schiefer-Ullrich & Andreesen, 1985). The notion that dietary Se might affect the metabolism of purines by human gut microbiota has been suggested (Dürre & Andreesen, 1982; Haft & Self, 2008), but the in vitro results shown here indicate that any requirement could depend upon microbiota composition: necessary if purine utilizers mainly are members of the Proteobacteria but perhaps of less importance if Firmicutes predominate. Altogether, these results (1) suggest that phylogeny is a poor predictor of microbial purine utilization; (2) indicate that the presence of the identified genes does not correlate with the breadth of purines utilized by an organism; (3) demonstrate effects on purine metabolism of two nutritional parameters, i.e., carbon source and metals availability; and (4) underscore the need for assessments beyond genomics when making predictions about purine metabolism by the gut microbiota.

Lastly, it is important to note that while the notion of using bacteria able to degrade purines might be an appealing strategy to lower pro-inflammatory UA, more work is needed to fully understand the consequences to the host. For example, the intestinal epithelium is the most vigorously self-renewing tissue of adult mammals, which imposes a high demand of nucleotides that are needed for proliferation and energy (Crosnier et al., 2006). Adenine—a purine consumed by several taxa in our study (Figures A.4 and Supplementary Figure A.4)—is a precursor of nucleic acids in intestinal cells unable to synthesize purines de novo (Savaiano & Clifford, 1981). Furthermore, recent work from the Colgan group showed that gut bacteria are a major source of purines that are used for nucleotide generation by the intestinal mucosa. Importantly, supplementation of purines directly through bacterial colonization improved intestinal epithelial cell wound healing and barrier restitution capabilities and suggested that purines play essential roles for colonic epithelial proliferation, energy balance, and mucin barrier integrity (Lee et al., 2018, 2020). Adenine also inhibits TNF (tissue necrosis factor)-α signaling in intestinal epithelial cells and reduces mucosal inflammation in a dextran-sodium-sulfate-induced colitis mouse model

(Fukuda et al., 2016). Consistent with these results, a recent study identified intestinal purine starvation associated with irritable bowel syndrome (Mars et al., 2020). Our results showing that PDB can metabolize a variety of purines and lower hypoxanthine levels in the intestine, and that the abundance of genes encoding for key proteins in anaerobic purine degradation is associated with lower purines in the gut, suggest that the influence of these organisms on purine availability to intestinal epithelial cells and barrier function needs to be carefully examined, especially considering that many of the taxa identified as anaerobic purine degraders, including *E. coli, E. bolteae, F. varium*, and *C. difficile*, have been associated with disease (Bartlett et al., 2022; Chandra et al., 2021). Thus, further studies are warranted to examine the contribution of purine degradation to the fitness of these taxa and host health.

In summary, the work presented here shows that anaerobic purine utilization is widespread among gut-dwelling bacteria and suggests that microbial purine degraders are important modulators of host purine homeostasis in the gut and of UA levels in circulation. Studies are needed to dissect the contribution of aerobic vs. anaerobic purine-consuming pathways to the purine economy, gut ecology, and health conditions including atherosclerosis.

A.5 Methods

Bacterial culture conditions

Cultures and plates were prepared in an anaerobic (ca. 75% N2/20% CO2/5% H2) chamber. Media formulations and preparations are detailed in Methods S1. Cultures were routinely grown at 37°C in anaerobic septum-stoppered “Hungate tubes” containing rich, well-buffered (pH 7) media, typically “Mega Medium”95 supplemented with maltose (0.9 g/l), cellobiose (0.86 g/l), fructose (0.46 g/l) and NaHCO3 (1.68 g/l) although most clostridial strains were reliably recovered from 20% glycerol freezer stocks using medium 11E (see Methods S1). Freshly-prepared cultures were combined at roughly equivalent levels (normalized to OD600) for the purpose of colonizing germ free mice. Growth on purines utilized medium 23B, which contains 0.1% yeast extract (0.05% for the RNA-Seq cultures) as the sole undefined component, with additions of carbohydrate and NH4Cl (25 and 10 mM, respectively) or purines (UA, 1 mg/ml; adenine, 1 mg/ml; allantoin, 45 mM) as carbon and nitrogen sources. These levels of UA and adenine did not dissolve fully, and growth in tubes was monitored by observing purine disappearance as well as culture OD600, measured using a Spectronic 20D+ (ca. 1.4 cm sample path length) after the saturating purine had settled (ca. 15 min). Commercially-available strains and isolates have been verified by 16S rRNA gene sequencing and are specified in the key resources table. As described below, additional steps were employed to verify isolate Escherichia coli I-11.

Gnotobiotic husbandry

All GF C57BL/6J and *ApoE* KO mice were maintained in a controlled environment in either plastic flexible film gnotobiotic isolators or individually ventilated cages under a strict 12 h light cycle and received sterilized water and standard chow (LabDiet 5021) ad libitum unless otherwise stated. The age and sex of mice used in different experiments are specified below. Sterility of GF animals was assessed by routine PCR testing (16S rRNA gene) and by incubating freshly

collected fecal samples under aerobic and anaerobic conditions using standard microbiology methods. The animal experiments were conducted according to relevant national and international guidelines and were approved by the UCLA Animal Research Committee, the UCLA IACUC, or the University of Wisconsin-Madison Animal Care and Use Committee, as appropriate.

Human studies

The Impaired Glucose Tolerance (IGT) and Microbiota study is a prospective observational study of subjects aged between 50 and 64 years. In the IGT-Microbiota study, more than 5,000 men and women born in Sweden with a range of glucose tolerance based on their fasting glucose values and an oral glucose tolerance test (OGTT) were screened. Here we included the same sub-population (n=1011, 44% male) as in Wu et al (Wu et al., 2020). where plasma urate levels were measured resulting in a cohort of 998 individuals after exclusion. Exclusion criteria were: known diabetes, other severe disease that may jeopardize interpretation of results, e.g. inflammatory bowel disease, rheumatic diseases, malignancy (unless no relapse during 5 years of follow-up), treatment with steroids or immune-modulating treatment, pharmacological treatment of infection during the last 3 months and major cognitive dysfunction. The ethics committee at Gothenburg University approved the study (Dnr 560-13) which was conducted in accordance with the Declaration of Helsinki. Participants gave written informed consent.

Plate assay of anaerobic purine utilization

These assays employed saturating levels of UA, adenine, or hypoxanthine using bilayer plates similar to those described previously (Barnes & Impey, 1974). Specifically, working in the glove bag a 25 ml base layer consisting of a 1:1 mixture of medium 26B plus 2.4% molten Bacto agar was poured into each 100 × 15 mm petri dish and allowed to solidify. The following day, the base layer was overlaid with a 7 ml top layer medium + agar plus (per 7 ml) 84 mg UA, 96 mg

adenine, or 91 mg hypoxanthine (see Methods S1). As for the liquid cultures, these levels of purines were saturating and formed and opaque overlay. Where indicated, filter-sterilized additions of concentrated stock solutions (e.g., glucose, fructose, NH4Cl, formate) were added to indicated levels in both base and overlay medium layers. Plates were allowed to dry for two days, then spotted with 4 µl of an overnight culture grown in rich medium (CMM or 11E) and incubated anaerobically at 37°C for 2 (UA, soluble substrates), 3 (hypoxanthine) or 7 (adenine) days, unless otherwise specified. Cultures that utilized purines both grew on the medium and formed zones of clearing as the saturating purine was depleted. Attempts to prepare overlay plates containing xanthine or guanine at levels useful to support growth did not show clear zones of purine utilization. Similarly, monolayer plates containing UA resulted in relatively indistinct zones of clearing relative to bilayer plates, often with formation of a dense ammonium urate precipitate especially after storage at 4°C. Five cultures were spotted/plate--*E. bolteae* or *E. coli* MS 200-1 as a positive control plus 4 test strains.

Trace mineral requirements for growth on UA

Tests utilized the standard phosphate-buffered basal medium formulation (“26B” with added 100 nM Mn, Ni, Zn, 50 nM Co, W) plus combinations of the following: 2.5 µM Fe, 5 µM Mo, 0.5 µM Se as indicated (specific mineral compounds are listed in Methods S1). No attempt was made to rigorously exclude the individual metal being tested, and it is likely that other medium components—particularly phosphate buffer, cysteine and yeast extract—supplied trace levels; therefore, evidence of mineral requirements indicates a substantial demand.

Inactivation of *allB*, *ygeV*, and *ygeW-arcC*

The *allB* gene (Locus: NZ_GG773866; HMPREF9553_RS01540), the *ygeV* gene (HMPREF9553_RS03160), and the *ygeW-arcC* operon (HMPREF9553_RS03165 - RS03180) of *E. coli* MS 200-1 were deleted and replaced with the *tetA-sacB* cassette amplified from T-SACK

in accord with standard recombineering methods, followed by elimination of the pSIM5 helper plasmid.86 The resulting constructs were verified by sequencing across the cassette-genome junctions. Primers used in these constructs are listed in the key resources table. Unfortunately *E. coli* MS 200-1 proved to be resistant to phage P1 transduction and the constructs could not be transferred to naive recipients.

Isolation of E. coli strain I-11

This strain was isolated from a de-identified human fecal sample collected in accord with University of Wisconsin Health Science Institutional Review Board. For the isolation, approximately 20 mg of aseptically-sampled material was injected into a 10-ml medium 23B anaerobic culture supplemented with 10 mg of UA, then incubated at 37°C. Upon observation of growth and disappearance of the UA precipitate, the culture was transferred (1:100 dilution) into the same medium, maintained under the same conditions, this was repeated then the enrichment was streaked to UA bilayer plates. Colonies demonstrating UA metabolism were purified and identified by sequencing of the 16S rRNA gene. Strain I-11 was a facultative rod-shaped organism, which fermented glucose, sucrose and lactose but not cellobiose, as expected for *E. coli*. A diagnostic PCR analysis confirmed the identification.

RNA isolation and sequencing

10 ml cultures were grown in triplicate in tubes containing medium 23B (with 0.05% yeast extract) plus a) 25 mM xylose + 10 mM NH_4Cl or b) 12 mg UA and harvested at OD_{600} = 0.25 by plunging the tubes into an ice water slurry, then pelleting cells prior to storage at -80°C. RNA was isolated using the Monarch Total RNA Miniprep kit with yields of 2000 – 3000 ng RNA per sample, as assayed by dye binding (Qubit). Samples in were submitted to the Microbial Genome Sequencing Center (MiGS, Pittsburgh, PA) where Illumina Stranded RNA library preparation with RiboZero Plus rRNA depletion was performed, followed by Illumina sequencing [paired-end reads

(2x50bp)]. The sequence data was processed by MiGS as follows: quality control and adapter trimming was performed with bcl2fastq [ver. 2.20.0.445]. Read mapping was performed with HISAT2 [ver. 2.2.0] (Kim et al., 2019). Read quantification was performed using Subread's featureCounts [ver. 2.0.1] functionality (Liao et al., 2014). Read counts were loaded into R [ver. 4.0.2, R Core Team (2020)] and normalized using edgeR's Trimmed Mean of M values (TMM) algorithm (ver. 1.14.5) (Robinson et al., 2010). Subsequent values were then converted to counts per million (cpm).

Mouse experimental design

We conducted four animal experiments as follows. i) At UCLA six-week-old female AXB10/PgnJ, BTBR T+tf/J, BXD5/TyJ, and BXA8/PgnJ conventionally-raised mice from the HMDP cohort were fed a Western diet (Research Diets D10042101) for 4 weeks, and cecal contents were collected. Frozen cecal samples were shipped to the University of Wisconsin-Madison for microbiota transplant. Ten-week-old GF female *ApoE* KO mice (C57BL/6J background) fed a standard chow (LabDiet 5021) were inoculated by oral gavage with 0.2 mL of resuspended cecal slurry from these HMDP strains. Mice were switched to a standard chow diet supplemented with 0.2% cholesterol (TD.07798, Envigo). Mice were then euthanized at 18 weeks of age after 4h fasting and tissues collected. ii) Cecal and plasma samples were collected from eighteen-week-old GF or conventionally-raised *ApoE* KO mice fed the 0.2% cholesterol-supplemented diet (TD.07798) for 8 weeks. iii) Three groups of adult gnotobiotic C57BL/6J mice on a standard chow were tested for purine metabolism: a) mice bearing a "core" community which included seven species that do not degrade purines in vitro, *Bacteroides caccae, Bifidobacterium dentium, Blautia hansenii, Bacteroides thetaiotaomicron, Coprococcus comes, Mitsuokella multacida,* and *Ruminococcus torques*; b) mice bearing the "core plus purine-degrading bacteria (PDB)" community that added three PDB to the "core" community mixture including *E. bolteae, H. hathewayi*, and *E. coli*; and c) one group remained germ-free. iv) Three groups of adult gnotobiotic

C57BL/6J mice on a standard chow were tested for purine metabolism: a) the same “core” community as above; b) the “core plus *E. coli* MS 200-1 wild-type”; and c) the “core plus the *E. coli* MS 200-1 deletion variant FER041”. In the latter two experiments, oxonic acid, a uricase inhibitor, was supplemented in the drinking water (1.5% w/v) as previously described (Dankers et al., 2013) and samples were collected 4 weeks after the colonization.

Atherosclerotic lesion assessments

Atherosclerotic lesions were assessed as previously described (Kasahara et al., 2017). Briefly, mice were anesthetized and the aorta was perfused with PBS. To assess the atherosclerotic lesion size at the aortic sinus, the samples were cut in the ascending aorta, and the proximal samples containing the aortic sinus were embedded in Tissue-Tek OCT compounds. Five consecutive sections (10 μm thickness) taken at 100 μm intervals (i.e. 50, 150, 250, 350, and 450 μm from a bottom of the aortic sinus) were collected from each mouse and stained with Oil Red O. Plaque area and Oil Red O-positive area were measured using Image J software. The volume of atherosclerosis in the aortic sinus was expressed as mean size of the 5 sections for each mouse. Immunohistochemistry was performed on formalin-fixed cryosections of mouse aortic roots using antibodies to identify macrophages (1:50), followed by detection with biotinylated secondary antibodies (1:400) and streptavidin-horseradish peroxidase (1:500). Smooth muscle cells were identified by immunostaining with fluorescein isothiocyanate (FITC)-conjugated primary antibody against α- smooth muscle actin (1:100), followed by anti-FITC biotin-conjugated secondary antibody (1:400). Negative controls were prepared by substitution with an isotype control antibody. Staining with Masson’s trichrome was used to delineate the fibrous area according to the manufacturer’s instructions. Stained sections were digitally captured, and the percentage of the stained area (the stained area per total atherosclerotic lesion area) was calculated.

Plasma biochemical analysis

Blood samples were collected by cardiac puncture into EDTA-rinsed syringes under anesthesia using isoflurane. Plasma was acquired by centrifugation and stored at -80°C until measurement. The levels of triglycerides, total cholesterol, and high-density lipoprotein cholesterol were measured with commercially available kits from Wako Chemicals. Plasma LPS levels were quantitated with the QCL-1000 Endpoint Chromogenic LAL Assay. Plasma UA levels were determined by the Vistro DT60 II Analyzer at the University of Massachusetts Medical School MMPC (National Mouse Metabolic Phenotyping Center). For the experiment shown in Figure 6G, plasma UA was measured using the HPLC system described below, having first diluted plasma samples 8-fold in PBS and passing the diluted plasma through a 3kDa ultrafiltration device (Amicon Ultra-0.5).

DNA extraction from cecal and fecal samples

DNA was extracted from samples according to published bead-beating procedures (Kasahara et al., 2018; Kreznar et al., 2017; Turnbaugh et al., 2009). In short, fecal or cecal samples were resuspended in a solution containing 500 µl of 2× 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. Cells were mechanically disrupted using a bead beater (BioSpec Products) for 3 min at room temperature. The aqueous layer was collected and DNA precipitated using 600 µl isopropanol and 60 µl 3M Na-acetate. Pellets were rinsed with ethanol, dried, and resuspended in TE buffer. A NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel) was used to remove contaminants. Isolated DNA was stored at -80°C until downstream processing.

qPCR analysis

The level of E. coli in the bacterial communities described in Figure 6F were assessed by quantitative extraction of DNA from fecal samples by bead beating followed by qPCR analysis of 2 ng purified DNA samples using the E. coli-specific primers 401F and 611R and SsoAdvanced SYBR Green Supermix (BioRad) with 40 amplification cycles (Walker et al., 2017). The method generated a linear standard curve (0.02 - 20000 pg) using purified E. coli MS 200-1 DNA and a uniform melt curve for all standards and sample products.

COPRO-Seq analysis

Bacterial communities resulting from inoculation of GF animals were analyzed using Illumina sequencing according to the COPRO-Seq (community profiling by sequencing) method.93 Feces were collected 4 weeks after the colonization. In short, DNA isolated from feces via bead beating was used to prepare libraries for shotgun Illumina sequencing. Five hundred nanograms of DNA from each sample was fragmented by sonication and subjected to enzymatic blunting and adenine tailing. Customized Illumina adapters containing maximally distant 8-bp barcodes were ligated to the poly (A)-tailed DNA. Gel-extracted DNA (size selection ~250 to 300bp) was amplified by PCR using primers and cycling conditions recommended by Illumina. Purified PCR products were submitted to the UW-Madison Biotechnology Center for a single end 50-bp Illumina MiSeq run. Results were processed using the software pipeline detailed by McNulty et al.93

Metagenomic shotgun DNA sequencing

DNA was extracted from cecal contents of individual mice as described above. Following DNA extraction, Illumina paired end libraries were constructed using a previously described protocol (Faith et al., 2011), with a modification of gel selecting DNA fragments at ~450 bp in length. Paired end (PE) reads (2 × 125) were generated using HiSeq 2500 platform.

Metagenomic reads processing

Raw reads were preprocessed using Fastx Toolkit (ver. 0.0.13): (1) for demultiplexing raw samples, fastx_barcode_splitter.pl, with –partial 2, mismatch 2 was used; (2) when more than one forward and reverse read file existed for a single sample (due to being run on more than one lane, more than one platform, or at more than one time), read files were concatenated into one forward and one reverse read file; (3) barcodes were trimmed to form reads (fastx_trimmer -f 9 -Q 33); (4) and reads were trimmed to remove low quality sequences (fastq_quality_trimmer -t 20 -l 30 -Q33). Following trimming, unpaired reads were eliminated from the analysis using custom Python scripts. To identify and eliminate host sequences, reads were aligned against the mouse genome (mm10/GRCm38) using bowtie2 (ver. 2.3.4) (Langmead & Salzberg, 2012, p. 2) with default settings and microbial DNA reads that did not align to the mouse genome were identified using samtools (ver. 1.3; samtools view -b -f 4 -f 8).

Microbiome trait quantification

Quantification of microbial genes was done by aligning clean paired end reads from each sample to a previous published mouse gut microbiome non-redundant gene catalog using Bowtie2 (ver. 2.3.4) and default parameters. RSEM (ver. 1.3.1) was used to estimate microbial gene abundance.89 Relative abundance of microbial gene counts per million (CPM) were calculated using microbial gene expected counts divided by gene effective length then normalized by the total sum. To obtain abundance information for microbial functions, CPM of genes with the same KEGG Orthology (KO) annotation were summed together. In case there were multiple KO annotations for a single gene, we used all KO annotations. To obtain taxonomic abundance, CPM of genes with the same NCBI taxa annotation were summed together at phylum, order, class, family, and genus levels with a minimum of 10 genes in each taxon.

HPLC analysis

Analyses of growth media and ultra-filtered serum samples (Figure A.6g) were performed using a Shimadzu system comprised of a CBM-40 controller, LC-40D pumps, SIL-40C autosampler, CTO-40C column oven, and SPD-M40 diode array detector. Samples were applied to a Phenomenx Luna Omega 5µm Polar C18 LC column maintained at 25°C with a 0.5 ml/min gradient composed of (A) 100 mM KxHxPO4 pH 2.4 and (B) A with 40% acetonitrile as follows (Time in minutes, %A, %B): 0, 99, 1 / 15, 96, 4 / 25, 25, 75 / 35, 25, 75 / 36, 99, 1 / 50, 99,1. Under these conditions formate, lactate and acetate eluted at 7.1, 10.1, and 10.6 min and 1 mM concentrations were readily detected at 205 nm. UA eluted at 17.4 min and 10 µM was detected at 284 nm. As some samples contained saturating substrate levels, for UA analyses all samples were vigorously mixed and immediately diluted 40-fold in phosphate buffered saline to allow full UA dissolution prior to analysis.

Headspace gas chromatography

Analyses of short-chain fatty acids and ethanol in growth media were performed using a Shimadzu headspace GC/FID as previously described (Murga-Garrido et al., 2021). Briefly, samples were added to chilled 20 ml headspace vials containing 2.0 g NaHSO4 and 1.0 ml of 60 µM 2-butanol (internal standard). Vials were crimp sealed immediately after sample addition and vortexed periodically to disperse and mix the contents. Headspace GC analyses were performed using a Shimadzu HS-20 headspace sampler connected to a Shimadzu GC-2010 Plus GC equipped with a 30 m SH-Stabilwax column linked to a FID. Samples were equilibrated with shaking to 80°C for 20 min and pressurized to 80 kPa for 3 min prior to column injection (2 ml injection loop, load time 0.2 min, sample and transfer line temperature 150°C, 1:15 split ratio, N2 column flow 1.2 ml/min), with the following column temperature program: 40°C/2 min, increased to 200°C (20°C/min), held 2 min, decreased to 120°C (20°C/min), decreased to 40°C (40°C/min), and stabilized 1 min prior to the subsequent injection. The GC cycle time was approximately 23

min. Standard mixtures were prepared and analyzed by the same method, and peak areas determined using Shimadzu Lab Solution software (version 5.92).

GC-MS measurement of short-chain fatty acids

Sample preparation was based on a previously described procedure (Kasahara et al., 2018). Cecal contents were weighed into 4 ml polytetrafluoroethylene (PTFE) screw cap vials and 10 µl of a mixture of internal standards (20 mM of acetic acid-D4, propionic acid-D6, and butyric acid-D7) was subsequently added to each vial, followed by 20 µl of 33% HCl and 1 ml diethyl ether. For plasma samples, 50 µl of each sample, 1.25 µl of the internal standard mix, 5 µl of 33% HCl, and 0.75 ml of diethyl ether were mixed. The mixture was vortexed vigorously for 3 min and then centrifuged (4,000 × g, 10 min). The upper organic layer was transferred to another vial and a second diethyl ether extraction was performed. After combining the two ether extracts, a 60 µl aliquot was removed, combined with 2 µl N-tert-butyldimethylsilyl-N-methyltrifluoroacetamide (MTBSTFA) in a GC auto-sampler vial with a 200 µl glass insert, and incubated for 2 h at room temperature. Derivatized samples (1 µl) were injected onto an Agilent 7890B/5977A GC/MSD instrument with an Agilent DB1-ms 0.25 mm x 60 m column with 0.25 µm bonded phase. A discontinuous oven program was used starting at 40°C for 2.25 min, then ramping at 20°C/min to 200°C, then ramping at 100°C/min to 300°C and holding for 7 min. The total run time was 18.25 min. Linear column flow was maintained at 1.26 ml/min. The inlet temperature was set to 250°C with an injection split ratio of 15:1. Quantitation was performed using selected ion monitoring (SIM) acquisition mode and metabolites were compared to relevant labeled internal standards using Agilent Mass Hunter v. Acquisition B.07.02.1938. The m/z of monitored ions are as follows: 117 (acetic acid), 120 (acetic acid-D4), 131 (propionic acid), 136 (propionic acid-D6), 145 (butyric acid), and 151 (butyric acid-D7). Concentrations were normalized to mg of cecal contents.

uHPLC-MS/MS analysis of metabolites

Plasma samples were prepared for analysis by precipitating proteins with 4 volumes of ice-cold methanol spiked with 2.5 μM deuterium-labeled choline and TMAO internal standards. Samples were centrifuged at 18,213 × g at 4°C for 3 min. The recovered supernatants were diluted 1:1 in uHPLC-grade water prior to screening. Identification and quantification of choline and TMAO was performed using a uHPLC (Dionex 3000) coupled to a high-resolution mass spectrometer (Thermo Scientific Q Exactive). Liquid chromatography separation was achieved on a Dikma Bio-Bond C4 column (150 mm by 2.1 mm; 3-μm particle size) using a 7-min isocratic gradient (50:50 methanol-water, 5 mM ammonium formate, and 0.1% formic acid). A heated electrospray ionization interface, working in positive mode, was used to direct column eluent to the mass spectrometer. Quantitation of TMAO and D9-TMAO was performed via targeted MS/MS using the following paired masses of parent ions and fragments: TMAO (76.0762 and 58.0659) and D9-TMAO (85.1318 and 68.1301). Quantitation of choline and d9-choline was performed in full-MS scan mode by monitoring their exact masses: 104.1075 and 113.1631, respectively.

UPLC-MS/MS for untargeted plasma metabolome

Untargeted mass spectrometry data were collected at Metabolon Inc. Plasma samples were prepared using the automated MicroLab STAR system (Hamilton Company). To remove protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix and to recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was divided into five fractions: two for analysis by two separate reverse phase (RP)/UPLC-MS/MS methods with positive ion mode electrospray ionization (ESI), one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, one for analysis by HILIC/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. Samples were placed briefly on a TurboVap (Zymark) to remove the organic solvent. The sample extracts were stored overnight under nitrogen before preparation for analysis.

All methods utilized a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo Scientific Q-Exactive high resolution/accurate mass spectrometer interfaced with a heated electrospray ionization (HESI-II) source and Orbitrap mass analyzer operated at 35,000 mass resolution. The sample extract was dried then reconstituted in solvents compatible to each of the four methods. Each reconstitution solvent contained a series of standards at fixed concentrations to ensure injection and chromatographic consistency. One aliquot was analyzed using acidic positive ion conditions, chromatographically optimized for more hydrophilic compounds. In this method, the extract was gradient eluted from a C18 column (Waters UPLC BEH C18–2.1×100 mm, 1.7 µm) using water and methanol containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% formic acid (FA). Another aliquot was also analyzed using acidic positive ion conditions, however it was chromatographically optimized for more hydrophobic compounds. In this method, the extract was gradient eluted from the aforementioned C18 column using methanol, acetonitrile, water, 0.05% PFPA and 0.01% FA and was operated at an overall higher organic content. Another aliquot was analyzed using basic negative ion optimized conditions using a separate dedicated C18 column. The basic extracts were gradient eluted from the column using methanol and water, amended with 6.5mM ammonium bicarbonate at pH 8. The fourth aliquot was analyzed via negative ionization following elution from a HILIC column (Waters UPLC BEH Amide 2.1 × 150 mm, 1.7 µm) using a gradient consisting of water and acetonitrile with 10mM ammonium formate, pH 10.8. Compounds were identified by comparison to library entries based upon retention time/index, mass to charge ratio (m/z) and chromatographic data, and peaks were quantified using area-under-the curve.

Purine and pyrimidine metabolite quantitation

Targeted purine/pyrimidine mass spectrometry data were collected at The Metabolomics Innovation Centre (Victoria, Canada). An internal standard (IS) solution containing 13C- and/or 15N-labeled AMP, ATP, GMP, GTP, UMP, UTP, xanthine, guanine and adenine, hypothanxine,

guanosine and adenosine was prepared in 80% methanol (for cecal samples) or 10% methanol (for plasma samples). Serially-diluted standard solutions containing standard substances of the targeted nucleotides, nucleosides and nucleobases were prepared in a concentration range of 0.00001 to 50 nmol/ml in 80% methanol (for cecal samples) or 0.00001 to 10 nmol/ml in 10% methanol (for plasma samples). The cecal samples were lyophilized to dryness. The powder of each sample was weighted into an Eppendorf tube and 80% methanol at 20 µl per mg of powder was added. The samples were homogenized on a mill mixer at 30 Hz for 2 min, three times, with the aid of 2 metal beads, followed by sonication in an ice-water bath for 5 min. The tubes were centrifuged at 21,000 × g and 5°C for 20 min. 100 µl of the clear supernatant of each sample or 100 µl of each standard solution was in turn mixed with an equal volume of the IS solution, 100 µl of water and 120 µl of dichloromethane. The mixtures were vortex-mixed at 3,000 rpm for 2 min and subsequently centrifuged at 21,000 × g for 5 min. 120 µl of the clear supernatant was collected and then dried at 30°C under a nitrogen gas flow. For plasma samples, 20 µl was mixed with 200 µl of the IS solution and 780 µl of methanol. After vortex mixing at 3000 rpm for 1 min and sonication in an ice-water bath for 2 min, the sample was centrifuged at 21,000 × g and 5°C for 15 min. 500 µl of the clear supernatant of each sample was transferred to another tube, 100 µl of water and 400 µl of dichloromethane then added. The mixture was vortex mixed at 3000 rpm for 1 min and then centrifuged. The upper aqueous phase of each sample was transferred to a LC microvial and dried under a nitrogen gas flow. For both cecal and plasma samples, the residues were reconstituted in 100 µl of 10% methanol. 10 µl aliquots of the sample solutions and the standard solutions were injected into a C18 LC column (2.1 × 100 mm, 1.8 µm) to run UPLC-MRM/MS on a Waters Acquity UPLC system coupled to a Sciex QTRAP 6500 Plus mass spectrometer operated in the negative-ion mode for detection of nucleotides. The mobile phase was a tributylamine acetate buffer and acetonitrile for binary gradient elution (5% to 35% acetonitrile over 25 min), at 0.25 ml/min and 40°C. For quantitation of nucleosides and nucleobases, 10 µl aliquots of the sample solutions and the standard solutions were injected into

a polar C18 UPLC column (2.1 × 100 mm, 1.6 μm) to run UPLC-MRM/MS on the same LC-MS instrument operated in the positive-ion mode. The mobile phase was a 0.1% HFBA solution and acetonitrile for binary-solvent gradient elution (0% to 28% acetonitrile in 14 min), at 0.30 ml/min and 40°C.

Genes encoding for purine degradation

The quantification of genes encoding purine-degradation functions in bacterial genomes were predicted by BLASTP of the NCBI RefSeq Genome Database (refseq_genomes) using five *Enterocloster bolteae* genes found in multiple purine-utilizing organisms as queries (*ygeY*, encoding a Se-dependent hydrolase; *ygeX*/*dpaL*, encoding a diamino-propionate ammonia-lyase; *ssnA*, encoding an amino-hydrolase; *hydA*/*hyuA*, encoding a dihydro-pyrimidinase; *xdhD*, encoding a Se-dependent Xanthine DH). Purine-degrading bacteria were defined as bearing the five "probe" genes listed above with identity > 25% to the *E. bolteae* genes. We removed redundancy if two bacteria genomes had identical proteins encoded by all five purine degradation genes. To quantify the abundance of these genes in gut microbiomes of transplanted mice, we mapped metagenomic reads to genes for purine utilization functions using RSEM (ver. 1.3.1) (B. Li & Dewey, 2011). Relative abundance of microbial gene counts per million (CPM) were calculated using microbial gene expected counts divided by gene effective length then normalized by the total sum.

Human study

UA and coronary artery calcification (CAC) data were obtained from a pre-diabetes Swedish cohort participating in a study examining the link between the gut microbiota and type 2 diabetes (Wu et al., 2020). The cohort comprised men and women aged 50-64 years from the Gothenburg area, Sweden, who were recruited at random from the census register (n=988). Exclusion criteria were: known diabetes; inflammatory diseases, such as Crohn's disease,

ulcerative colitis, and rheumatic diseases; treatments with steroids or immunomodulatory drugs; cancer (unless relapse-free for the preceding 5 years); cognitive dysfunction; and treatment for infectious diseases and with antibiotics in the past three months. UA was measured using a photometric technique on a Roche Cobas analyzer. Coronary artery calcifications (CAC) were assessed by computed tomography (CT) scanning using a dual-source CT scanner equipped with a Stellar Detector (Siemens). CAC images were obtained using electrocardiogram-gated non-contrast cardiac CT imaging at 120 kV. All non-contrast image sets were reconstructed (B35f HeartView medium CaScore) and CAC were identified and scored using the syngo.via calcium scoring software (Volume Wizard; Siemens) to obtain a CAC score according to Agatston. Previously reported microbiome data34 was mapped against Unified Human Gastrointestinal Genome (UHGG) v1.0 catalogue using Kraken2 v.2.1.2 to examine association between UA levels and bacterial taxa. Xgboost and caret packages in R version 4.0.3 were used to select microbes associated with the serum urate levels based on regression analysis. These associations were further assessed for significance after adjustment for covariates (CACS, BMI, gender, triglycerides and HbA1c) in a mixed linear regression model.

Data analysis and statistical analysis

Data integration and statistical analysis were performed in R (ver. 3.6.3) or Prism 9. The data were expressed as box-and-whisker plots with individual data points, where the boxes indicate the median values and the interquartile ranges and the whiskers represent the minimum and maximum values. Significance was calculated by unpaired two-tailed Student's t test or one-way ANOVA with the Tukey post-tests. Details of these analyses (n, p values) are presented in figure legends. The correlation analysis was performed using Spearman's correlation by R function "cor.test()". For multiple testing, Benjamini-Hochberg FDR procedure was used to adjust p-values. Heatmap plots were performed using R package pheatmap (ver. 1.0.12). Purine metabolites include ones from class "purine nucleotides", "purine nucleosides",

“imidazopyrimidines”, and “azoles”. Pyrimidine metabolites include ones from class “pyrimidine nucleotides”, “pyrimidine nucleosides”, “diazines”, and “pyrazolopyrimidines”. A partial least squares discriminant analysis (PLS-DA) model was constructed using the function PLSR.Anal() in the MetanoAnalystR package and the nonlinear iterative partial least squares (NIPALS) algorithm to obtain the variable importance for the projection (VIP) of purine and pyrimidine metabolites (Pang et al., 2022). We performed the Leave-One-Out Cross-Validation (LOOCV) method for cross validation.

Data and code availability

The mouse metagenomics data, bacterial RNA-Seq data, Copro-Seq data, human metagenomics data were deposited in NCBI Sequence Read Archive under Bioproject: PRJNA904303, Bioproject: PRJNA911264, Bioproject: PRJNA903666, and European Genome-Phenome Archive under EGA: EGAS00001004480. Microscopy data reported in this paper will be shared by the lead contact upon request.

All original code is available here: https://github.com/qijunz/Purine_paper

Any additional information required to reanalyze the data reported in this paper including the custom Python scripts employed in the processing of metagenomic data is available from the lead contact upon request.

A.6 Contributions

K.K. and F.E.R. conceived the study. K.K. and R.L.K. performed microbiology and mouse studies and collected phenotypic and transcriptomic data. Q.Z. and M.P. conducted statistical analyses. M.M. and A.J.L. contributed key reagents. M.P., G.B., and F.B. conducted human studies. K.K., R.L.K., and F.E.R. wrote the manuscript. All authors read and approved the final manuscript.

A.7 Acknowledgements

We thank the University of Wisconsin Biotechnology Center DNA Sequencing Facility for providing sequencing and support services, the University of Wisconsin Mass Spectrometry Facility for the SCFA measurement, the Metabolomics Innovation Centre (TMIC) for purine metabolomics analysis, and Metabolon for untargeted metabolomics analysis. This work was partly supported by grants from the National Institutes of Health: NIH HL144651 (to F.E.R. and A.J.L.), NIH HL148577 (to F.E.R. and A.J.L.), and NIH HL147883 (to A.J.L.). This work was also supported by grants from a Transatlantic Networks of Excellence Award from Foundation Leducq (17CVD01 to F.B. and F.E.R.), from the Knut and Alice Wallenberg Foundation (2017.0026 to F.B.), from the Swedish Heart Lung Foundation (20210366 to F.B.), and from AFA insurances (160337 to F.B.). F.B. is the Torsten Söderberg Professor in Medicine and a Wallenberg Scholar.

A.8 References

Agarwal, V., Hans, N., & Messerli, F. H. (2013). Effect of Allopurinol on Blood Pressure: A Systematic Review and Meta-Analysis. *The Journal of Clinical Hypertension*, *15*(6), 435–442. https://doi.org/10.1111/j.1751-7176.2012.00701.x

Barba, M., Dutoit, R., Legrain, C., & Labedan, B. (2013). Identifying reaction modules in metabolic pathways: Bioinformatic deduction and experimental validation of a new putative route in purine catabolism. *BMC Systems Biology*, *7*(1), 99. https://doi.org/10.1186/1752-0509-7-99

Barnes, E. M., & Impey, C. S. (1974). The Occurence and Properties of Uric Acid Decomposing Anaerobic Bacteria in the Avian Caecum. *Journal of Applied Bacteriology*, *37*(3), 393–409. https://doi.org/10.1111/j.1365-2672.1974.tb00455.x

Bartlett, A., Padfield, D., Lear, L., Bendall, R., & Vos, M. (2022). A comprehensive list of bacterial pathogens infecting humans. *Microbiology*, *168*(12). https://doi.org/10.1099/mic.0.001269

Beker, B. M., Colombo, I., Gonzalez-Torres, H., & Musso, C. G. (2022). Decreasing microbiota-derived uremic toxins to improve CKD outcomes. *Clinical Kidney Journal*, *15*(12), 2214–2219. https://doi.org/10.1093/ckj/sfac154

Bennett, B. J., Davis, R. C., Civelek, M., Orozco, L., Wu, J., Qi, H., Pan, C., Packard, R. R. S., Eskin, E., Yan, M., Kirchgessner, T., Wang, Z., Li, X., Gregory, J. C., Hazen, S. L., Gargalovic, P. S., & Lusis, A. J. (2015). Genetic Architecture of Atherosclerosis in Mice: A Systems Genetics Analysis of Common Inbred Strains. *PLOS Genetics*, *11*(12), e1005711. https://doi.org/10.1371/journal.pgen.1005711

Bolte, L. A., Vich Vila, A., Imhann, F., Collij, V., Gacesa, R., Peters, V., Wijmenga, C., Kurilshikov, A., Campmans-Kuijpers, M. J. E., Fu, J., Dijkstra, G., Zhernakova, A., & Weersma, R. K. (2021). Long-term dietary patterns are associated with pro-inflammatory

and anti-inflammatory features of the gut microbiome. *Gut*, *70*(7), 1287–1298. https://doi.org/10.1136/gutjnl-2020-322670

Braga, T. T., Forni, M. F., Correa-Costa, M., Ramos, R. N., Barbuto, J. A., Branco, P., Castoldi, A., Hiyane, M. I., Davanso, M. R., Latz, E., Franklin, B. S., Kowaltowski, A. J., & Camara, N. O. S. (2017). Soluble Uric Acid Activates the NLRP3 Inflammasome. *Scientific Reports*, *7*(1), 39884. https://doi.org/10.1038/srep39884

Brandsma, E., Kloosterhuis, N. J., Koster, M., Dekker, D. C., Gijbels, M. J. J., Van Der Velden, S., Ríos-Morales, M., Van Faassen, M. J. R., Loreti, M. G., De Bruin, A., Fu, J., Kuipers, F., Bakker, B. M., Westerterp, M., De Winther, M. P. J., Hofker, M. H., Van De Sluis, B., & Koonen, D. P. Y. (2019). A Proinflammatory Gut Microbiota Increases Systemic Inflammation and Accelerates Atherosclerosis. *Circulation Research*, *124*(1), 94–100. https://doi.org/10.1161/CIRCRESAHA.118.313234

Chandra, H., Sharma, K. K., Tuovinen, O. H., Sun, X., & Shukla, P. (2021). Pathobionts: Mechanisms of survival, expansion, and interaction with host with a focus on *Clostridioides difficile*. *Gut Microbes*, *13*(1), 1979882. https://doi.org/10.1080/19490976.2021.1979882

Chen-Xu, M., Yokose, C., Rai, S. K., Pillinger, M. H., & Choi, H. K. (2019). Contemporary Prevalence of Gout and Hyperuricemia in the United States and Decadal Trends: The National Health and Nutrition Examination Survey, 2007–2016. *Arthritis & Rheumatology*, *71*(6), 991–999. https://doi.org/10.1002/art.40807

Choi, H. K. (2010). A prescription for lifestyle change in patients with hyperuricemia and gout. *Current Opinion in Rheumatology*, *22*(2), 165–172. https://doi.org/10.1097/BOR.0b013e328335ef38

Chu, Y., Sun, S., Huang, Y., Gao, Q., Xie, X., Wang, P., Li, J., Liang, L., He, X., Jiang, Y., Wang, M., Yang, J., Chen, X., Zhou, C., Zhao, Y., Ding, F., Zhang, Y., Wu, X., Bai, X., … Huang, R. (2021). Metagenomic analysis revealed the potential role of gut microbiome in

gout. *Npj Biofilms and Microbiomes*, *7*(1), 66. https://doi.org/10.1038/s41522-021-00235-2

Crosnier, C., Stamataki, D., & Lewis, J. (2006). Organizing cell renewal in the intestine: Stem cells, signals and combinatorial control. *Nature Reviews Genetics*, *7*(5), 349–359. https://doi.org/10.1038/nrg1840

Cusa, E., Obradors, N., Baldomà, L., Badía, J., & Aguilar, J. (1999). Genetic Analysis of a Chromosomal Region Containing Genes Required for Assimilation of Allantoin Nitrogen and Linked Glyoxylate Metabolism in *Escherichia coli*. *Journal of Bacteriology*, *181*(24), 7479–7484. https://doi.org/10.1128/JB.181.24.7479-7484.1999

Dankers, A. C. A., Mutsaers, H. A. M., Dijkman, H. B. P. M., Van Den Heuvel, L. P., Hoenderop, J. G., Sweep, F. C. G. J., Russel, F. G. M., & Masereeuw, R. (2013). Hyperuricemia influences tryptophan metabolism via inhibition of multidrug resistance protein 4 (MRP4) and breast cancer resistance protein (BCRP). *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1832*(10), 1715–1722. https://doi.org/10.1016/j.bbadis.2013.05.002

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., Collini, S., Pieraccini, G., & Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proceedings of the National Academy of Sciences*, *107*(33), 14691–14696. https://doi.org/10.1073/pnas.1005963107

Drivelegka, P., Forsblad-d'Elia, H., Angerås, O., Bergström, G., Schmidt, C., Jacobsson, L. T. H., & Dehlin, M. (2020). Association between serum level of urate and subclinical atherosclerosis: Results from the SCAPIS Pilot. *Arthritis Research & Therapy*, *22*(1), 37. https://doi.org/10.1186/s13075-020-2119-0

Dürre, P., & Andreesen, J. R. (1982). Anaerobic degradation of uric acid via pyrimidine derivatives by selenium-starved cells of Clostridium purinolyticum. *Archives of Microbiology*, *131*(3), 255–260. https://doi.org/10.1007/BF00405889

Dürre, P., & Andreesen, J. R. (1983). Purine and glycine metabolism by purinolytic clostridia. *Journal of Bacteriology*, *154*(1), 192–199. https://doi.org/10.1128/jb.154.1.192-199.1983

Errasti-Murugarren, E., Pastor-Anglada, M., & Casado, F. J. (2007). Role of CNT3 in the transepithelial flux of nucleosides and nucleoside-derived drugs. *The Journal of Physiology*, *582*(3), 1249–1260. https://doi.org/10.1113/jphysiol.2007.130138

Faith, J. J., McNulty, N. P., Rey, F. E., & Gordon, J. I. (2011). Predicting a Human Gut Microbiota's Response to Diet in Gnotobiotic Mice. *Science*, *333*(6038), 101–104. https://doi.org/10.1126/science.1206025

Falconi, C. A., Junho, C. V. D. C., Fogaça-Ruiz, F., Vernier, I. C. S., Da Cunha, R. S., Stinghen, A. E. M., & Carneiro-Ramos, M. S. (2021). Uremic Toxins: An Alarming Danger Concerning the Cardiovascular System. *Frontiers in Physiology*, *12*, 686249. https://doi.org/10.3389/fphys.2021.686249

Feig, D. I., Kang, D.-H., & Johnson, R. J. (2008). Uric Acid and Cardiovascular Risk. *New England Journal of Medicine*, *359*(17), 1811–1821. https://doi.org/10.1056/NEJMra0800885

Fernández-Calotti, P., Casulleras, O., Antolin, M., Guarner, F., & Pastor-Anglada, M. (2016). Galectin-4 interacts with the drug transporter human concentrative nucleoside transporter 3 to regulate its function. *The FASEB Journal*, *30*(2), 544–554. https://doi.org/10.1096/fj.15-272773

Fukuda, T., Majumder, K., Zhang, H., Turner, P. V., Matsui, T., & Mine, Y. (2016). Adenine Inhibits TNF-α Signaling in Intestinal Epithelial Cells and Reduces Mucosal Inflammation in a Dextran Sodium Sulfate-Induced Colitis Mouse Model. *Journal of Agricultural and Food Chemistry*, *64*(21), 4227–4234. https://doi.org/10.1021/acs.jafc.6b00665

Gaal, T., Bartlett, M. S., Ross, W., Turnbough, C. L., & Gourse, R. L. (1997). Transcription Regulation by Initiating NTP Concentration: rRNA Synthesis in Bacteria. *Science*, *278*(5346), 2092–2097. https://doi.org/10.1126/science.278.5346.2092

Haft, D. H., & Self, W. T. (2008). Orphan SelD proteins and selenium-dependent molybdenum hydroxylases. *Biology Direct*, *3*(1), 4. https://doi.org/10.1186/1745-6150-3-4

Hartwich, K., Poehlein, A., & Daniel, R. (2012). The Purine-Utilizing Bacterium Clostridium acidurici 9a: A Genome-Guided Metabolic Reconsideration. *PLoS ONE*, *7*(12), e51662. https://doi.org/10.1371/journal.pone.0051662

Iwadate, Y., & Kato, J. (2019). Identification of a Formate-Dependent Uric Acid Degradation Pathway in *Escherichia coli*. *Journal of Bacteriology*, *201*(11). https://doi.org/10.1128/JB.00573-18

Ju, C., Lai, R. W. C., Li, K. H. C., Hung, J. K. F., Lai, J. C. L., Ho, J., Liu, Y., Tsoi, M. F., Liu, T., Cheung, B. M. Y., Wong, I. C. K., Tam, L. S., & Tse, G. (2020). Comparative cardiovascular risk in users versus non-users of xanthine oxidase inhibitors and febuxostat versus allopurinol users. *Rheumatology*, *59*(9), 2340–2349. https://doi.org/10.1093/rheumatology/kez576

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., & Rey, F. E. (2018). Interactions between Roseburia intestinalis and diet modulate atherogenesis in a murine model. *Nature Microbiology*, *3*(12), 1461–1471. https://doi.org/10.1038/s41564-018-0272-x

Kasahara, K., Tanoue, T., Yamashita, T., Yodoi, K., Matsumoto, T., Emoto, T., Mizoguchi, T., Hayashi, T., Kitano, N., Sasaki, N., Atarashi, K., Honda, K., & Hirata, K. (2017). Commensal bacteria at the crossroad between cholesterol homeostasis and chronic inflammation in atherosclerosis. *Journal of Lipid Research*, *58*(3), 519–528. https://doi.org/10.1194/jlr.M072165

Keerman, M., Yang, F., Hu, H., Wang, J., Wang, F., Li, Z., Yuan, J., Yao, P., Zhang, X., Guo, H., Yang, H., & He, M. (2020). Mendelian randomization study of serum uric acid levels and diabetes risk: Evidence from the Dongfeng-Tongji cohort. *BMJ Open Diabetes Research & Care*, *8*(1), e000834. https://doi.org/10.1136/bmjdrc-2019-000834

Kessler, T., & Schunkert, H. (2021). Coronary Artery Disease Genetics Enlightened by Genome-Wide Association Studies. *JACC: Basic to Translational Science*, *6*(7), 610–623. https://doi.org/10.1016/j.jacbts.2021.04.001

Khosla, U. M., Zharikov, S., Finch, J. L., Nakagawa, T., Roncal, C., Mu, W., Krotova, K., Block, E. R., Prabhakar, S., & Johnson, R. J. (2005). Hyperuricemia induces endothelial dysfunction. *Kidney International*, *67*(5), 1739–1742. https://doi.org/10.1111/j.1523-1755.2005.00273.x

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kimura, Y., Yanagida, T., Onda, A., Tsukui, D., Hosoyamada, M., & Kono, H. (2020). Soluble Uric Acid Promotes Atherosclerosis via AMPK (AMP-Activated Protein Kinase)-Mediated Inflammation. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *40*(3), 570–582. https://doi.org/10.1161/ATVBAHA.119.313224

Koeth, R. A., Wang, Z., Levison, B. S., Buffa, J. A., Org, E., Sheehy, B. T., Britt, E. B., Fu, X., Wu, Y., Li, L., Smith, J. D., DiDonato, J. A., Chen, J., Li, H., Wu, G. D., Lewis, J. D., Warrier, M., Brown, J. M., Krauss, R. M., … Hazen, S. L. (2013). Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nature Medicine*, *19*(5), 576–585. https://doi.org/10.1038/nm.3145

Kratzer, J. T., Lanaspa, M. A., Murphy, M. N., Cicerchi, C., Graves, C. L., Tipton, P. A., Ortlund, E. A., Johnson, R. J., & Gaucher, E. A. (2014). Evolutionary history and metabolic

insights of ancient mammalian uricases. *Proceedings of the National Academy of Sciences*, *111*(10), 3763–3768. https://doi.org/10.1073/pnas.1320393111

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., Hagenbuch, B., Attie, A. D., & Rey, F. E. (2017). Host Genotype and Gut Microbiome Modulate Insulin Secretion and Diet-Induced Metabolic Phenotypes. *Cell Reports*, *18*(7), 1739–1750. https://doi.org/10.1016/j.celrep.2017.01.062

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lee, J. S., Wang, R. X., Alexeev, E. E., Lanis, J. M., Battista, K. D., Glover, L. E., & Colgan, S. P. (2018). Hypoxanthine is a checkpoint stress metabolite in colonic epithelial energy modulation and barrier function. *Journal of Biological Chemistry*, *293*(16), 6039–6051. https://doi.org/10.1074/jbc.RA117.000269

Lee, J. S., Wang, R. X., Goldberg, M. S., Clifford, G. P., Kao, D. J., & Colgan, S. P. (2020). Microbiota-Sourced Purines Support Wound Healing and Mucous Barrier Function. *iScience*, *23*(6), 101226. https://doi.org/10.1016/j.isci.2020.101226

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 1–16.

Li, Y., Jin, Z., Yu, X., Allewell, N. M., Tuchman, M., & Shi, D. (2011). The *ygeW* encoded protein from *Escherichia coli* is a knotted ancestral catabolic transcarbamylase. *Proteins: Structure, Function, and Bioinformatics*, *79*(7), 2327–2334. https://doi.org/10.1002/prot.23043

Li, Z., Pan, Q., Xiao, Y., Fang, X., Shi, R., Fu, C., Danchin, A., & You, C. (2019). Deciphering global gene expression and regulation strategy in *Escherichia coli* during carbon limitation. *Microbial Biotechnology*, *12*(2), 360–376. https://doi.org/10.1111/1751-7915.13343

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Liu, Y., Jarman, J. B., Low, Y. S., Huang, S., Chen, H., DeFeo, M. E., Sekiba, K., Hou, B.-H., Ganesan, C., Pao, A. C., Gombar, S., & Dodd, D. (2022). A widely distributed gene cluster compensates for uricase loss in hominids. *bioRxiv*. https://doi.org/10.1101/2022.07.24.501321

Lusis, A. J. (2000). Atherosclerosis. *Nature*, *407*(6801), 233–241. https://doi.org/10.1038/35025203

Mackenzie, I. S., Hawkey, C. J., Ford, I., Greenlaw, N., Pigazzani, F., Rogers, A., Struthers, A. D., Begg, A. G., Wei, L., Avery, A. J., Taggar, J. S., Walker, A., Duce, S. L., Barr, R. J., Dumbleton, J. S., Rooke, E. D., Townend, J. N., Ritchie, L. D., MacDonald, T. M., … Zutis, K. (2022). Allopurinol versus usual care in UK patients with ischaemic heart disease (ALL-HEART): A multicentre, prospective, randomised, open-label, blinded-endpoint trial. *The Lancet*, *400*(10359), 1195–1205. https://doi.org/10.1016/S0140-6736(22)01657-9

MAGIC, on behalf of Procardis Consortium, Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., Allen, H. L., Lindgren, C. M., Luan, J., Mägi, R., Randall, J. C., Vedantam, S., Winkler, T. W., Qi, L., Workalemahu, T., Heid, I. M., Steinthorsdottir, V., … Loos, R. J. F. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics*, *42*(11), 937–948. https://doi.org/10.1038/ng.686

Mars, R. A. T., Yang, Y., Ward, T., Houtti, M., Priya, S., Lekatz, H. R., Tang, X., Sun, Z., Kalari, K. R., Korem, T., Bhattarai, Y., Zheng, T., Bar, N., Frost, G., Johnson, A. J., van Treuren, W., Han, S., Ordog, T., Grover, M., … Kashyap, P. C. (2020). Longitudinal

Multi-omics Reveals Subset-Specific Mechanisms Underlying Irritable Bowel Syndrome. *Cell*, *182*(6), 1460-1473.e17. https://doi.org/10.1016/j.cell.2020.08.007

Martinon, F., Pétrilli, V., Mayor, A., Tardivel, A., & Tschopp, J. (2006). Gout-associated uric acid crystals activate the NALP3 inflammasome. *Nature*, *440*(7081), 237–241. https://doi.org/10.1038/nature04516

Méndez-Salazar, E. O., & Martínez-Nava, G. A. (2022). Uric acid extrarenal excretion: The gut microbiome as an evident yet understated factor in gout development. *Rheumatology International*, *42*(3), 403–412. https://doi.org/10.1007/s00296-021-05007-x

Modi, S. R., Collins, J. J., & Relman, D. A. (2014). Antibiotics and the gut microbiota. *Journal of Clinical Investigation*, *124*(10), 4212–4218. https://doi.org/10.1172/JCI72333

Murga-Garrido, S. M., Hong, Q., Cross, T.-W. L., Hutchison, E. R., Han, J., Thomas, S. P., Vivas, E. I., Denu, J., Ceschin, D. G., Tang, Z.-Z., & Rey, F. E. (2021). Gut microbiome variation modulates the effects of dietary fiber on host metabolism. *Microbiome*, *9*(1), 117. https://doi.org/10.1186/s40168-021-01061-6

Naes, S., Ab-Rahim, S., Mazlan, M., Amir Hashim, N. A., & Abdul Rahman, A. (2023). Increased ENT2 expression and its association with altered purine metabolism in cell lines derived from different stages of colorectal cancer. *Experimental and Therapeutic Medicine*, *25*(5), 212. https://doi.org/10.3892/etm.2023.11911

Nemet, I., Saha, P. P., Gupta, N., Zhu, W., Romano, K. A., Skye, S. M., Cajka, T., Mohan, M. L., Li, L., Wu, Y., Funabashi, M., Ramer-Tait, A. E., Naga Prasad, S. V., Fiehn, O., Rey, F. E., Tang, W. H. W., Fischbach, M. A., DiDonato, J. A., & Hazen, S. L. (2020). A Cardiovascular Disease-Linked Gut Microbial Metabolite Acts via Adrenergic Receptors. *Cell*, *180*(5), 862-877.e22. https://doi.org/10.1016/j.cell.2020.02.016

Org, E., Parks, B. W., Joo, J. W. J., Emert, B., Schwartzman, W., Kang, E. Y., Mehrabian, M., Pan, C., Knight, R., Gunsalus, R., Drake, T. A., Eskin, E., & Lusis, A. J. (2015). Genetic

and environmental control of host-gut microbiota interactions. *Genome Research*, *25*(10), 1558–1569. https://doi.org/10.1101/gr.194118.115

Pang, Z., Zhou, G., Ewald, J., Chang, L., Hacariz, O., Basu, N., & Xia, J. (2022). Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. *Nature Protocols*, *17*(8), 1735–1761. https://doi.org/10.1038/s41596-022-00710-w

Papakostas, K., Botou, M., & Frillingos, S. (2013). Functional Identification of the Hypoxanthine/Guanine Transporters YjcD and YgfQ and the Adenine Transporters PurP and YicO of Escherichia coli K-12. *Journal of Biological Chemistry*, *288*(52), 36827–36840. https://doi.org/10.1074/jbc.M113.523340

Pastor-Anglada, M., & Pérez-Torras, S. (2018). Emerging Roles of Nucleoside Transporters. *Frontiers in Pharmacology*, *9*, 606. https://doi.org/10.3389/fphar.2018.00606

Rahimi-Sakak, F., Maroofi, M., Rahmani, J., Bellissimo, N., & Hekmatdoost, A. (2019). Serum uric acid and risk of cardiovascular mortality: A systematic review and dose-response meta-analysis of cohort studies of over a million participants. *BMC Cardiovascular Disorders*, *19*(1), 218. https://doi.org/10.1186/s12872-019-1215-z

Rao, G. N., Corson, M. A., & Berk, B. C. (1991). Uric acid stimulates vascular smooth muscle cell proliferation by increasing platelet-derived growth factor A-chain expression. *Journal of Biological Chemistry*, *266*(13), 8604–8608. https://doi.org/10.1016/S0021-9258(18)93017-6

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Savaiano, D. A., & Clifford, A. J. (1981). Adenine, The Precursor of Nucleic Acids in Intestinal Cells Unable to Synthesize Purines De Novo. *The Journal of Nutrition*, *111*(10), 1816–1822. https://doi.org/10.1093/jn/111.10.1816

Schiefer-Ullrich, H., & Andreesen, J. R. (1985). Peptostreptococcus barnesae sp. Nov., a Gram-positive, anaerobic, obligately purine utilizing coccus from chicken feces. *Archives of Microbiology*, *143*(1), 26–31. https://doi.org/10.1007/BF00414763

Schiefer-Ullrich, H., Wagner, R., D□rre, P., & Andreesen, J. R. (1984). Comparative studies on physiology and taxonomy of obligately purinolytic clostridia. *Archives of Microbiology*, *138*(4), 345–353. https://doi.org/10.1007/BF00410902

Selvaraj, M. S., Li, X., Li, Z., Pampana, A., Zhang, D. Y., Park, J., Aslibekyan, S., Bis, J. C., Brody, J. A., Cade, B. E., Chuang, L.-M., Chung, R.-H., Curran, J. E., De Las Fuentes, L., De Vries, P. S., Duggirala, R., Freedman, B. I., Graff, M., Guo, X., … Natarajan, P. (2022). Whole genome sequence analysis of blood lipid levels in >66,000 individuals. *Nature Communications*, *13*(1), 5995. https://doi.org/10.1038/s41467-022-33510-7

Singh, G., Lingala, B., & Mithal, A. (2019). Gout and hyperuricaemia in the USA: Prevalence and trends. *Rheumatology*, *58*(12), 2177–2180. https://doi.org/10.1093/rheumatology/kez196

Singh, J. A., Ramachandaran, R., Yu, S., & Curtis, J. R. (2017). Allopurinol use and the risk of acute cardiovascular events in patients with gout and diabetes. *BMC Cardiovascular Disorders*, *17*(1), 76. https://doi.org/10.1186/s12872-017-0513-6

Song, Y., Liu, C., Molitoris, D. R., Tomzynski, T. J., Lawson, P. A., Collins, M. D., & Finegold, S. M. (2003). Clostridium bolteae sp. Nov., Isolated from Human Sources. *Systematic and Applied Microbiology*, *26*(1), 84–89. https://doi.org/10.1078/072320203322337353

Sorensen, L. B., & Levinson, D. J. (1975). Origin and Extrarenal Elimination of Uric Acid in Man. *Nephron*, *14*(1), 7–20. https://doi.org/10.1159/000180432

Stylianou, I. M., Bauer, R. C., Reilly, M. P., & Rader, D. J. (2012). Genetic Basis of Atherosclerosis: Insights From Mice and Humans. *Circulation Research*, *110*(2), 337–355. https://doi.org/10.1161/CIRCRESAHA.110.230854

Sun, H.-L., Pei, D., Lue, K.-H., & Chen, Y.-L. (2015). Uric Acid Levels Can Predict Metabolic Syndrome and Hypertension in Adolescents: A 10-Year Longitudinal Study. *PLOS ONE*, *10*(11), e0143786. https://doi.org/10.1371/journal.pone.0143786

Tang, W. H. W., Wang, Z., Levison, B. S., Koeth, R. A., Britt, E. B., Fu, X., Wu, Y., & Hazen, S. L. (2013). Intestinal Microbial Metabolism of Phosphatidylcholine and Cardiovascular Risk. *New England Journal of Medicine*, *368*(17), 1575–1584. https://doi.org/10.1056/NEJMoa1109400

Tilg, H., Zmora, N., Adolph, T. E., & Elinav, E. (2020). The intestinal microbiota fuelling metabolic inflammation. *Nature Reviews Immunology*, *20*(1), 40–54. https://doi.org/10.1038/s41577-019-0198-4

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A., Affourtit, J. P., Egholm, M., Henrissat, B., Heath, A. C., Knight, R., & Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature*, *457*(7228), 480–484. https://doi.org/10.1038/nature07540

Tydell, C. C., Yount, N., Tran, D., Yuan, J., & Selsted, M. E. (2002). Isolation, Characterization, and Antimicrobial Properties of Bovine Oligosaccharide-binding Protein. *Journal of Biological Chemistry*, *277*(22), 19658–19664. https://doi.org/10.1074/jbc.M200659200

Uo, T., Yoshimura, T., Nishiyama, T., & Esaki, N. (2002). Gene Cloning, Purification, and Characterization of 2,3-Diaminopropionate Ammonia-lyase from *Escherichia coli*. *Bioscience, Biotechnology, and Biochemistry*, *66*(12), 2639–2644. https://doi.org/10.1271/bbb.66.2639

Wakuda, H., Uchida, S., Ikeda, M., Tabuchi, M., Akahoshi, Y., Shinozuka, K., & Yamada, S. (2014). Is Hyperuricemia a Risk Factor for Arteriosclerosis? Uric Acid and Arteriosclerosis in Apolipoprotein E-Deficient Mice. *Biological and Pharmaceutical Bulletin*, *37*(12), 1866–1871. https://doi.org/10.1248/bpb.b14-00201

Walker, D. I., McQuillan, J., Taiwo, M., Parks, R., Stenton, C. A., Morgan, H., Mowlem, M. C., & Lees, D. N. (2017). A highly specific Escherichia coli qPCR and its comparison with existing methods for environmental waters. *Water Research*, *126*, 101–110. https://doi.org/10.1016/j.watres.2017.08.032

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*(7341), 57–63. https://doi.org/10.1038/nature09922

Weisman, A., Tomlinson, G. A., Lipscombe, L. L., Perkins, B. A., & Hawker, G. A. (2019). Association between allopurinol and cardiovascular outcomes and all-cause mortality in diabetes: A retrospective, population-based cohort study. *Diabetes, Obesity and Metabolism*, *21*(6), 1322–1329. https://doi.org/10.1111/dom.13656

Wolf, D., & Ley, K. (2019). Immunity and Inflammation in Atherosclerosis. *Circulation Research*, *124*(2), 315–327. https://doi.org/10.1161/CIRCRESAHA.118.313591

Wu, H., Tremaroli, V., Schmidt, C., Lundqvist, A., Olsson, L. M., Krämer, M., Gummesson, A., Perkins, R., Bergström, G., & Bäckhed, F. (2020). The Gut Microbiota in Prediabetes and Diabetes: A Population-Based Cross-Sectional Study. *Cell Metabolism*, *32*(3), 379-390.e3. https://doi.org/10.1016/j.cmet.2020.06.011

Xue, A., Wu, Y., Zhu, Z., Zhang, F., Kemper, K. E., Zheng, Z., Yengo, L., Lloyd-Jones, L. R., Sidorenko, J., Wu, Y., eQTLGen Consortium, Agbessi, M., Ahsan, H., Alves, I., Andiappan, A., Awadalla, P., Battle, A., Beutner, F., Bonder, M. J., … Yang, J. (2018). Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. *Nature Communications*, *9*(1), 2941. https://doi.org/10.1038/s41467-018-04951-w

Yun, Y., Yin, H., Gao, Z., Li, Y., Gao, T., Duan, J., Yang, R., Dong, X., Zhang, L., & Duan, W. (2017). Intestinal tract is an important organ for lowering serum uric acid in rats. *PLOS ONE*, *12*(12), e0190194. https://doi.org/10.1371/journal.pone.0190194

Zhu, J., Zeng, Y., Zhang, H., Qu, Y., Ying, Z., Sun, Y., Hu, Y., Chen, W., Yang, H., Yang, J., & Song, H. (2022). The Association of Hyperuricemia and Gout With the Risk of Cardiovascular Diseases: A Cohort and Mendelian Randomization Study in UK Biobank. *Frontiers in Medicine*, *8*, 817150. https://doi.org/10.3389/fmed.2021.817150

Zhu, W., Gregory, J. C., Org, E., Buffa, J. A., Gupta, N., Wang, Z., Li, L., Fu, X., Wu, Y., Mehrabian, M., Sartor, R. B., McIntyre, T. M., Silverstein, R. L., Tang, W. H. W., DiDonato, J. A., Brown, J. M., Lusis, A. J., & Hazen, S. L. (2016). Gut Microbial Metabolite TMAO Enhances Platelet Hyperreactivity and Thrombosis Risk. *Cell*, *165*(1), 111–124. https://doi.org/10.1016/j.cell.2016.02.011

A.9 Figures

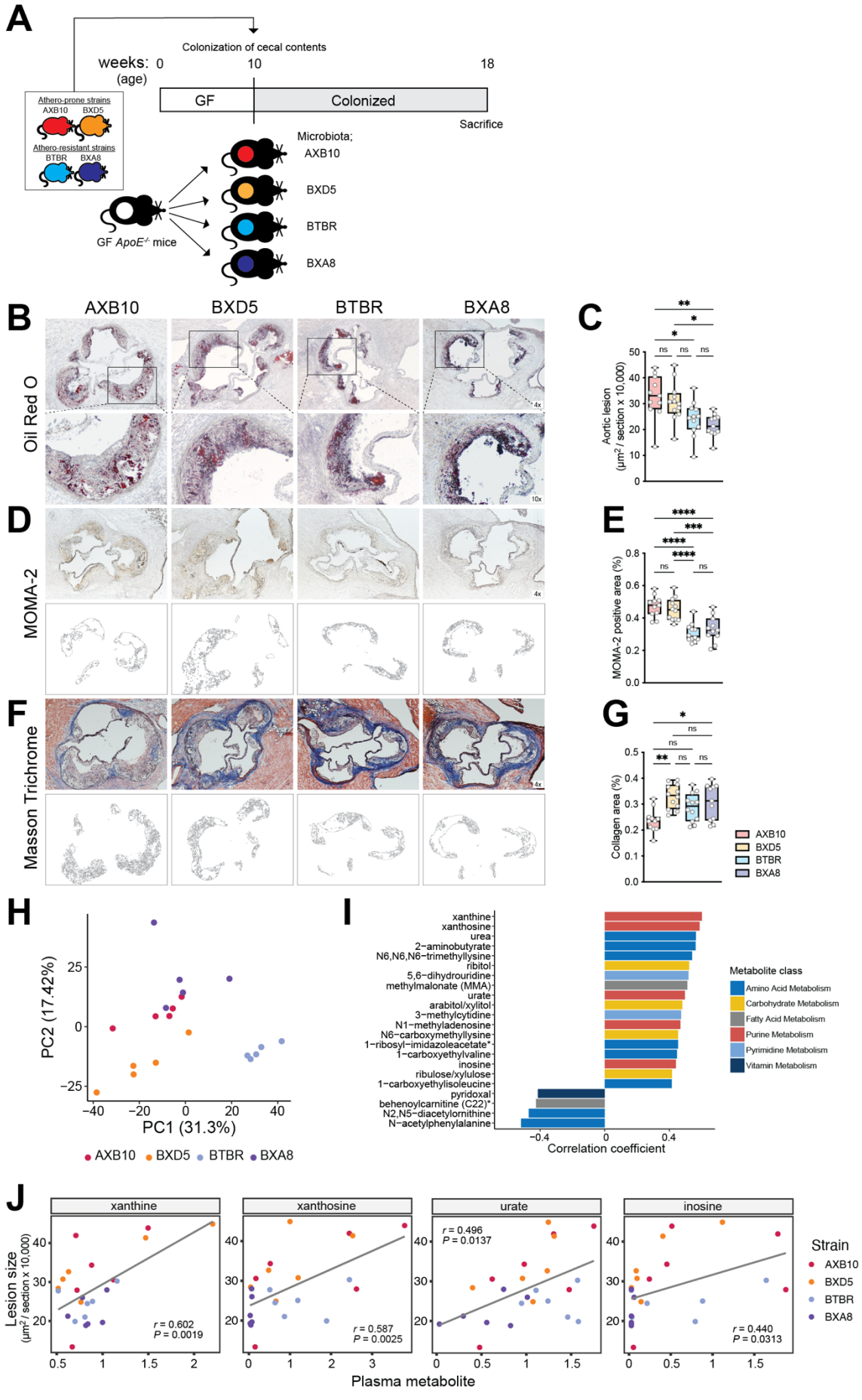

A
Colonization of cecal contents
weeks: 0
(age)
10
18
GF
Colonized
Sacrifice
Athero-prone strains
AXB10
BXD5
Athero-resistant strains
BTBR
BXA8
GF ApoE-/- mice
Microbiota;
AXB10
BXD5
BTBR
BXA8
B
AXB10
BXD5
BTBR
BXA8
Oil Red O
C
Aortic lesion (μm2 / section x 10,000)
D
MOMA-2
E
MOMA-2 positive area (%)
F
Masson Trichrome
G
Collagen area (%)
AXB10
BXD5
BTBR
BXA8
H
PC2 (17.42%)
PC1 (31.3%)
AXB10 BXD5 BTBR BXA8
I
xanthine
xanthosine
urea
2−aminobutyrate
N6,N6,N6−trimethyllysine
ribitol
5,6−dihydrouridine
methylmalonate (MMA)
urate
arabitol/xylitol
3−methylcytidine
N1−methyladenosine
N6−carboxymethyllysine
1−ribosyl−imidazoleacetate*
1−carboxyethylvaline
inosine
ribulose/xylulose
1−carboxyethylisoleucine
pyridoxal
behenoylcarnitine (C22)*
N2,N5−diacetylornithine
N−acetylphenylalanine
Correlation coefficient
Metabolite class
Amino Acid Metabolism
Carbohydrate Metabolism
Fatty Acid Metabolism
Purine Metabolism
Pyrimidine Metabolism
Vitamin Metabolism
J
xanthine
xanthosine
urate
inosine
Lesion size (μm2 / section x 10,000)
r = 0.602
P = 0.0019
r = 0.587
P = 0.0025
r = 0.496
P = 0.0137
r = 0.440
P = 0.0313
Plasma metabolite
Strain
AXB10
BXD5
BTBR
BXA8

Figure A.1 Plasma levels of purines are associated with atherosclerosis burden in transplanted gnotobiotic *ApoE* KO mice

(A) Experimental design.

(B–G) Representative sections and quantitative analysis of oil red O staining (B and C), MOMA-2 staining (D and E), and Masson's trichrome staining (F and G) in the aortic sinus (n = 11 for AXB10, n = 12 for BXD5, n = 12 for BTBR, n = 11 for BXA8). The data are expressed as box-and-whisker plots with individual data points, where the boxes indicate the median values and the interquartile ranges, and the whiskers represent the minimum and maximum values. Significance calculated by one-way ANOVA with the Tukey post-tests is indicated as follows: ∗p value < 0.05; ∗∗p value < 0.01; ∗∗∗p value <0.001; ∗∗∗∗p value < 0.0001.

(H) Principal component analysis of gut microbial functions from transplanted mice as determined by metagenomic analysis.

(I) Plasma metabolites positively or negatively associated with atherosclerotic lesion size, according to Spearman correlation analysis.

(J) Scatterplots showing associations between purines (relative mass spectrometry scaled intensities) and atherosclerosis lesion size (×104 μm2). GF; germ-free, *ApoE*; Apolipoprotein E, Chol; cholesterol, MOMA; monocytes and macrophage.

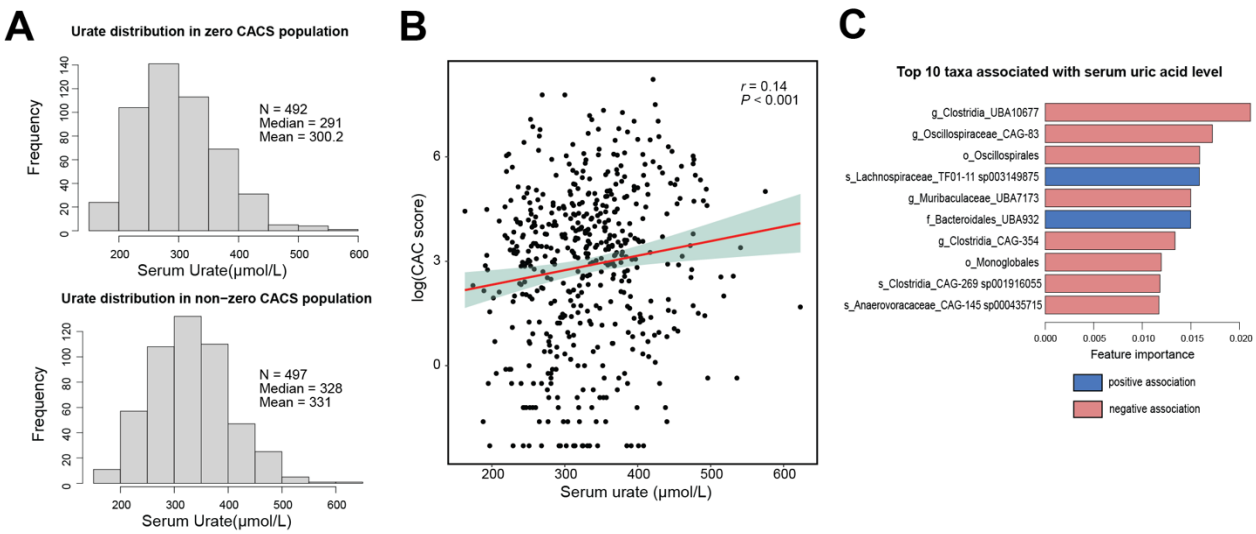


Figure A.2 Plasma UA levels are positively associated with coronary artery calcium (CAC) score in a human cohort

(A) Distribution of uric acid levels in serum from individuals with CAC score = 0 and CAC score > 0.

(B) Spearman correlation analysis between serum uric acid levels and CAC score.

(C) Top 10 taxa associated with serum UA levels. Blue and red bars show positive and negative associations, respectively.

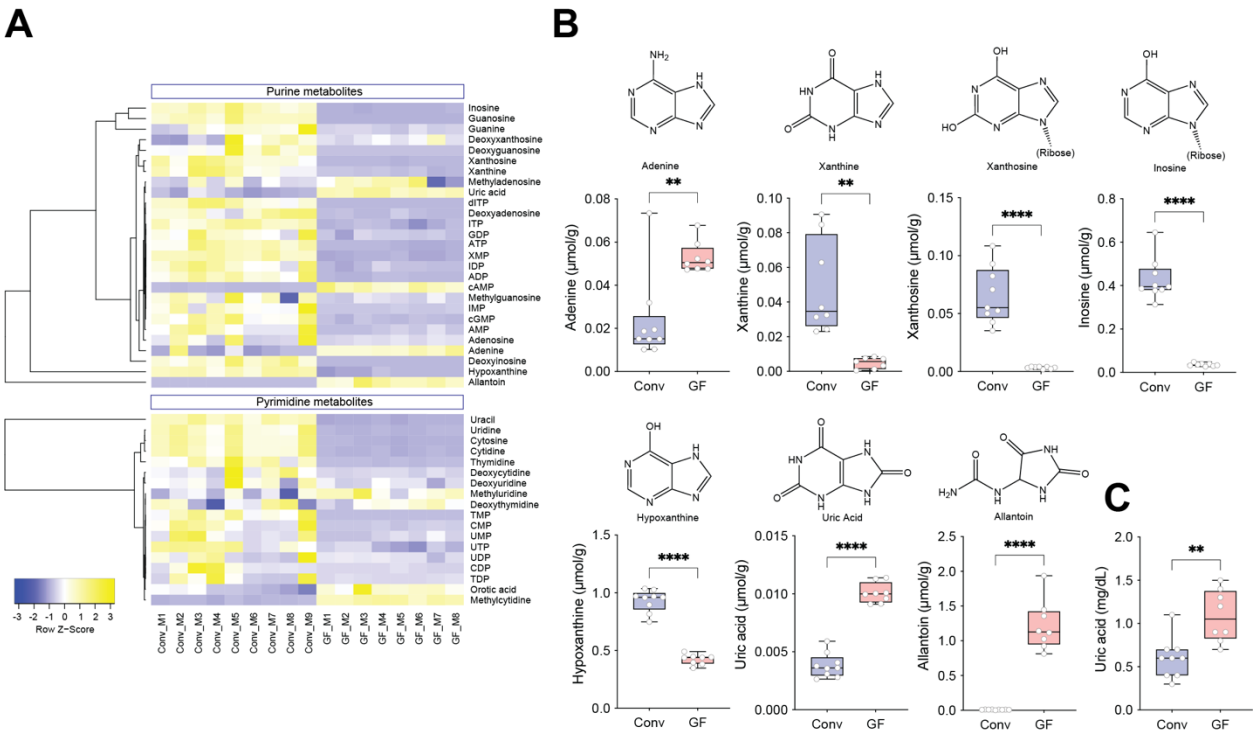


Figure A.3 Gut microbiome modulates purines in cecum and circulation

(A) Heatmap of purines and related metabolites in cecal contents from Conv (n = 9) and GF (n = 8) mice.

(B and C) (B) Values for adenine, xanthine, xanthosine, inosine, hypoxanthine, uric acid, and allantoin in cecal contents measured by LC-MS/MS, and (C) plasma uric acid levels analyzed by an enzymatic assay are shown using box-and-whisker plots with individual data points, where the boxes indicate the median values and the interquartile ranges and the whiskers represent the minimum and maximum values. Significance was calculated by unpaired two-tailed Student's t test and is designated as follows: ∗∗p value < 0.01; ∗∗∗∗p value < 0.0001. Conv, conventionally raised; GF; germ-free.

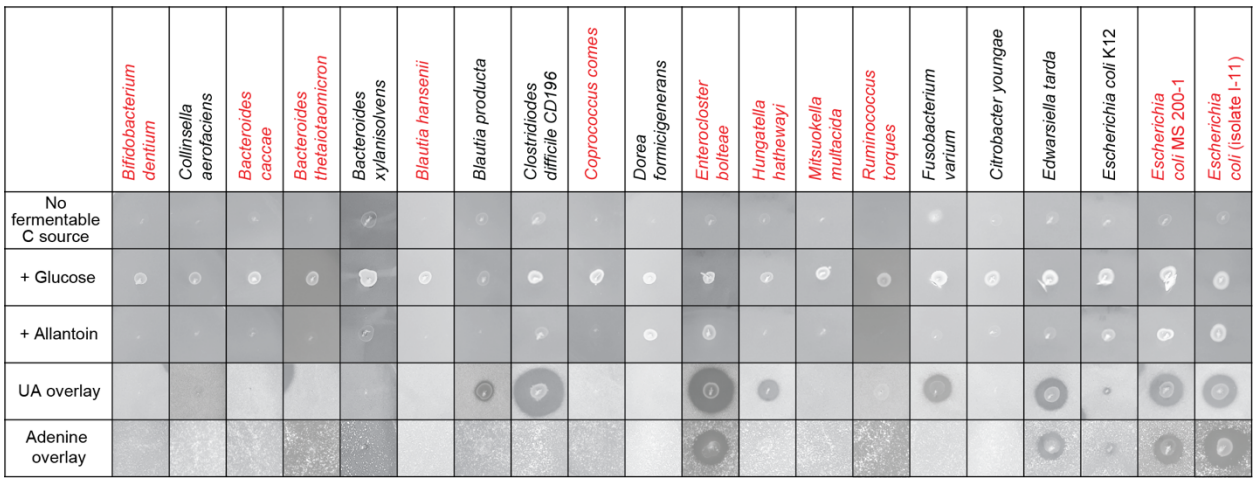


Figure A.4 Gut bacterial isolates use purines as carbon and energy sources

Anaerobic growth of bacterial strains on plates containing soluble (glucose and allantoin), and insoluble (uric acid and adenine) substrates. As detailed in Methods, plates were inoculated with 4 µL of dense overnight cultures grown in rich medium then incubated for 2 days (no fermentable substrate, glucose, allantoin or uric acid conditions) or 7 days (adenine). Growth is indicated by the appearance of cell patches and a zone of clearing for the overlay plates. Details about strains are specified in the key resources table, and a summary of all tested strains is presented in Supplementary Figure A.4. Strains indicated in red were used for colonization of gnotobiotic mice.

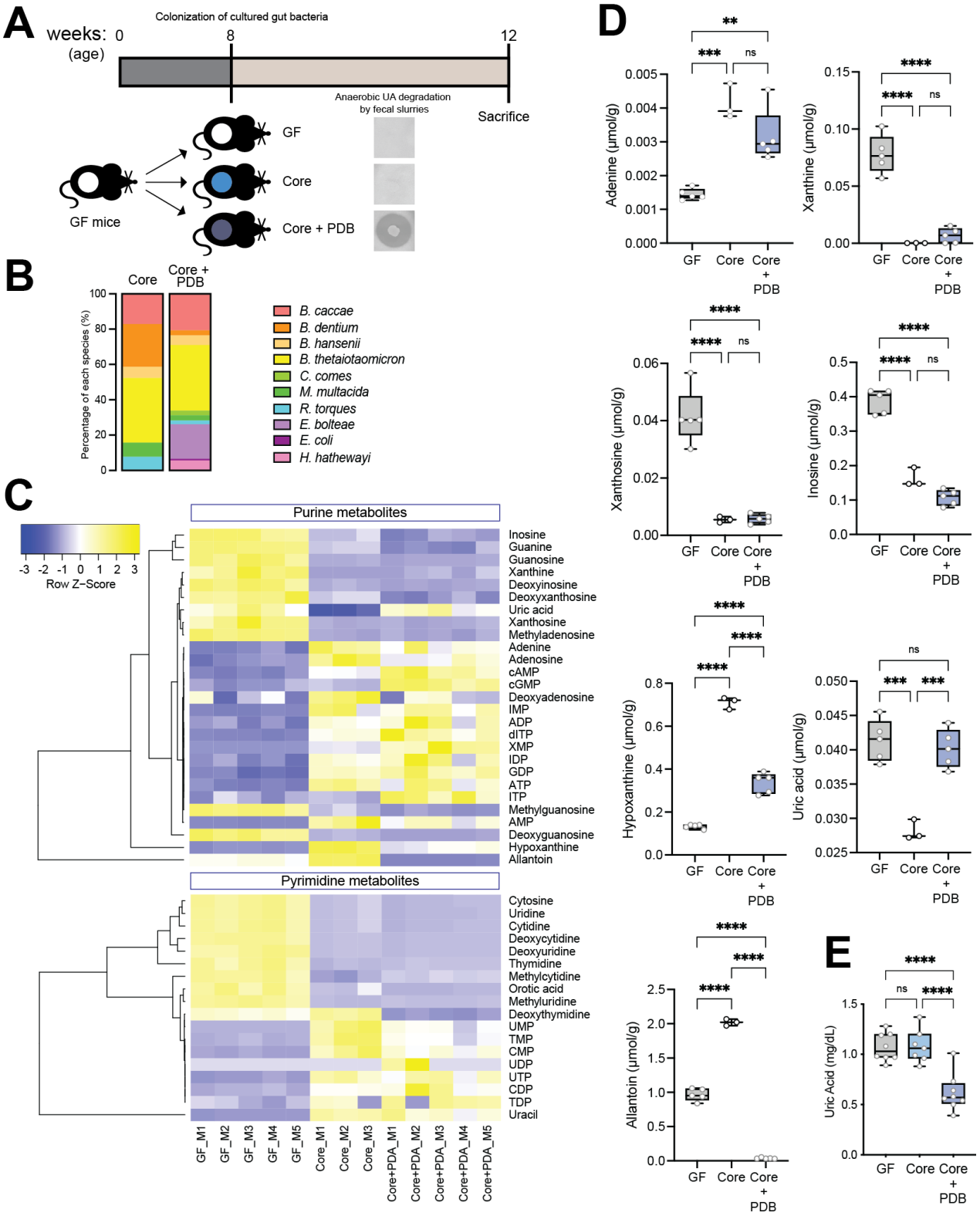

A
weeks: 0
(age)
8
12
Colonization of cultured gut bacteria
Anaerobic UA degradation by fecal slurries
Sacrifice
GF
Core
Core + PDB
GF mice
B
Core
Core + PDB
Percentage of each species (%)
B. caccae
B. dentium
B. hansenii
B. thetaiotaomicron
C. comes
M. multacida
R. torques
E. bolteae
E. coli
H. hathewayi
C
Row Z-Score
Purine metabolites
Inosine
Guanine
Guanosine
Xanthine
Deoxyinosine
Deoxyxanthosine
Uric acid
Xanthosine
Methyladenosine
Adenine
Adenosine
cAMP
cGMP
Deoxyadenosine
IMP
ADP
dITP
XMP
IDP
GDP
ATP
ITP
Methylguanosine
AMP
Deoxyguanosine
Hypoxanthine
Allantoin
Pyrimidine metabolites
Cytosine
Uridine
Cytidine
Deoxycytidine
Deoxyuridine
Thymidine
Methylcytidine
Orotic acid
Methyluridine
Deoxythymidine
UMP
TMP
CMP
UDP
UTP
CDP
TDP
Uracil
GF_M1
GF_M2
GF_M3
GF_M4
GF_M5
Core_M1
Core_M2
Core_M3
Core+PDA_M1
Core+PDA_M2
Core+PDA_M3
Core+PDA_M4
Core+PDA_M5
D
Adenine (μmol/g)
Xanthine (μmol/g)
Xanthosine (μmol/g)
Inosine (μmol/g)
Hypoxanthine (μmol/g)
Uric acid (μmol/g)
Allantoin (μmol/g)
GF
Core
Core + PDB
ns
E
Uric Acid (mg/dL)

Figure A.5 Purine-degrading bacteria (PDB) modulate abundance of purines in cecum and circulation

(A) Experimental design. Anaerobic uric acid degradation by fecal samples from different groups is indicated using uric acid overlay plates as detailed in Methods.

(B) Community profiling by sequencing (COPRO-seq) analysis of fecal samples from gnotobiotic B6 mice colonized with the core community (n = 4) or the core plus PDB community (n = 3). The bar charts show the abundance of each species in each community.

(C) Heatmap of purines and related metabolites in cecal contents from GF (n = 5), core (n = 3) and core plus PDB (n = 5) mice analyzed by LC-MS/MS.

(D and E) (D) Values for adenine, xanthine, xanthosine, inosine, hypoxanthine, uric acid, and allantoin in cecal contents of the three mouse cohorts analyzed by targeted metabolomics and (E) plasma uric acid levels analyzed by enzymatic assay were expressed as box-and-whisker plots with individual data points, where the boxes indicate the median values and the interquartile ranges and the whiskers represent the minimum and maximum values. Significance was calculated by one-way ANOVA test with the Tukey post-tests and is designated as follows: ∗∗ p value < 0.01; ∗∗∗p value < 0.001; ∗∗∗∗p value < 0.0001.

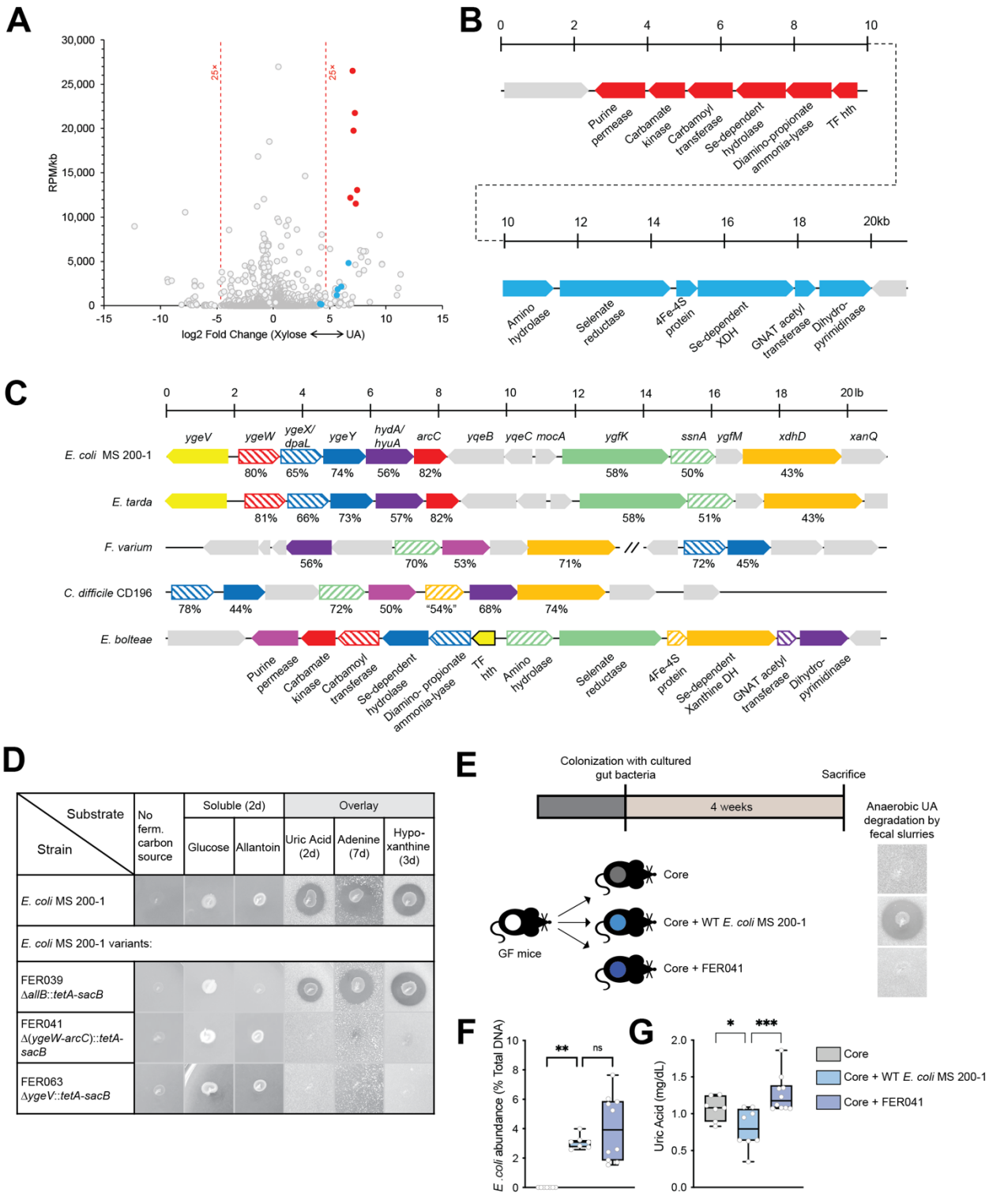

A
RPM/kb
log2 Fold Change (Xylose ⟷ UA)
25x
25x
B
Purine permease
Carbamate kinase
Carbamoyl transferase
Se-dependent hydrolase
Diamino-propionate ammonia-lyase
TF hth
Amino hydrolase
Selenate reductase
4Fe-4S protein
Se-dependent XDH
GNAT acetyl transferase
Dihydro-pyrimidinase
C
E. coli MS 200-1
E. tarda
F. varium
C. difficile CD196
E. bolteae
ygeV
ygeW
ygeX/ dpaL
ygeY
hydA/ hyuA
arcC
yqeB
yqeC
mocA
ygfK
ssnA
ygfM
xdhD
xanQ
Purine permease
Carbamate kinase
Carbamoyl transferase
Se-dependent hydrolase
Diamino- propionate ammonia-lyase
TF hth
Amino hydrolase
Selenate reductase
4Fe-4S protein
Se-dependent Xanthine DH
GNAT acetyl transferase
Dihydro-pyrimidinase
D
Substrate
Strain
No ferm. carbon source
Soluble (2d)
Glucose
Allantoin
Overlay
Uric Acid (2d)
Adenine (7d)
Hypo-xanthine (3d)
E. coli MS 200-1
E. coli MS 200-1 variants:
FER039 ΔallB::tetA-sacB
FER041 Δ(ygeW-arcC)::tetA-sacB
FER063 ΔygeV::tetA-sacB
E
Colonization with cultured gut bacteria
Sacrifice
4 weeks
Anaerobic UA degradation by fecal slurries
GF mice
Core
Core + WT E. coli MS 200-1
Core + FER041
F
E. coli abundance (% Total DNA)
**
ns
G
Uric Acid (mg/dL)
*
***
Core
Core + WT E. coli MS 200-1
Core + FER041

Figure A.6 Identification of a gene cluster necessary for anaerobic bacterial growth on purines

(A) Comparison of transcriptional profiles for Enterocloster bolteae showing differentially expressed genes (FDR < 0.01) and reads per million (RPM)/gene size (kb) for cultures grown on xylose + NH_4Cl (xylose, upregulated genes to the left) or uric acid (upregulated genes to the right), highlighting genes encoding two adjacent predicted operons encoding functions likely necessary for uric acid metabolism (blue and red circles). Descriptions of additional upregulated genes including those encoding micronutrient transport, one glycine-cleavage system and a probable bifurcating hydrogenase system as well as genes upregulated during growth on xylose are described in Supplementary Figure A.7.

(B) Diagram of the adjacent upregulated predicted operons, including genes CGC65_RS20560-RS20625, color-matched with the filled circles in (A).

(C) Representative alignments of chromosomal regions from multiple organisms that anaerobically catabolize uric acid. The genetic regions from *E. bolteae* shown in (B) are compared with those from *Clostridioides difficile* CD196 (CD196_RS16070–RS16115), *Fusobacterium varium* (C4N18_RS01955–RS01995) and (C4N18_RS03270–RS03290), *Edwardsiella tarda* (ETATCC_RS03320–RS03390) and *E. coli* MS 200-1 (HMPREF9553_RS03160–RS03225). Matched genes are color-coded, and the percent similarities of the encoded proteins are indicated. Although selected genes appear to be conserved, their organization differs in different organisms, and in the case of F. varium do not occur in a contiguous chromosomal region.

(D) Growth of *E. coli* MS 200-1 wild-type and deletion variants (FER039 [Δ*allB*::*tetA-sacB*], FER041 [Δ(*ygeW-arcC*)::*tetA-sacB*], and FER063 [Δ*ygeV*::*tetA-sacB*]) on plates lacking a carbon source, supplemented with glucose or allantoin, or prepared with overlays containing saturating amounts of uric acid, adenine, or hypoxanthine.

(E) In vivo experimental design. Mice were colonized with the core community as in Figure 5 and with either *E. coli* MS 200-1 wild-type or the deletion variant FER041 (Δ(*ygeW-arcC*)::*tetA-sacB*).

Anaerobic uric acid degradation by fecal samples from different groups is indicated using uric acid overlay plates.

(F and G) (F) The levels of fecal E. coli in the bacterial communities were assessed by qPCR and (G) plasma uric acid levels measured by HPLC were expressed as box-and-whisker plots with individual data points, where the boxes indicate the median values and the interquartile ranges and the whiskers represent the minimum and maximum values. Significance was calculated by one-way ANOVA test with the Tukey post-tests and is designated as follows: $*$p value < 0.05; $**$p value < 0.01; $***$p value < 0.001.

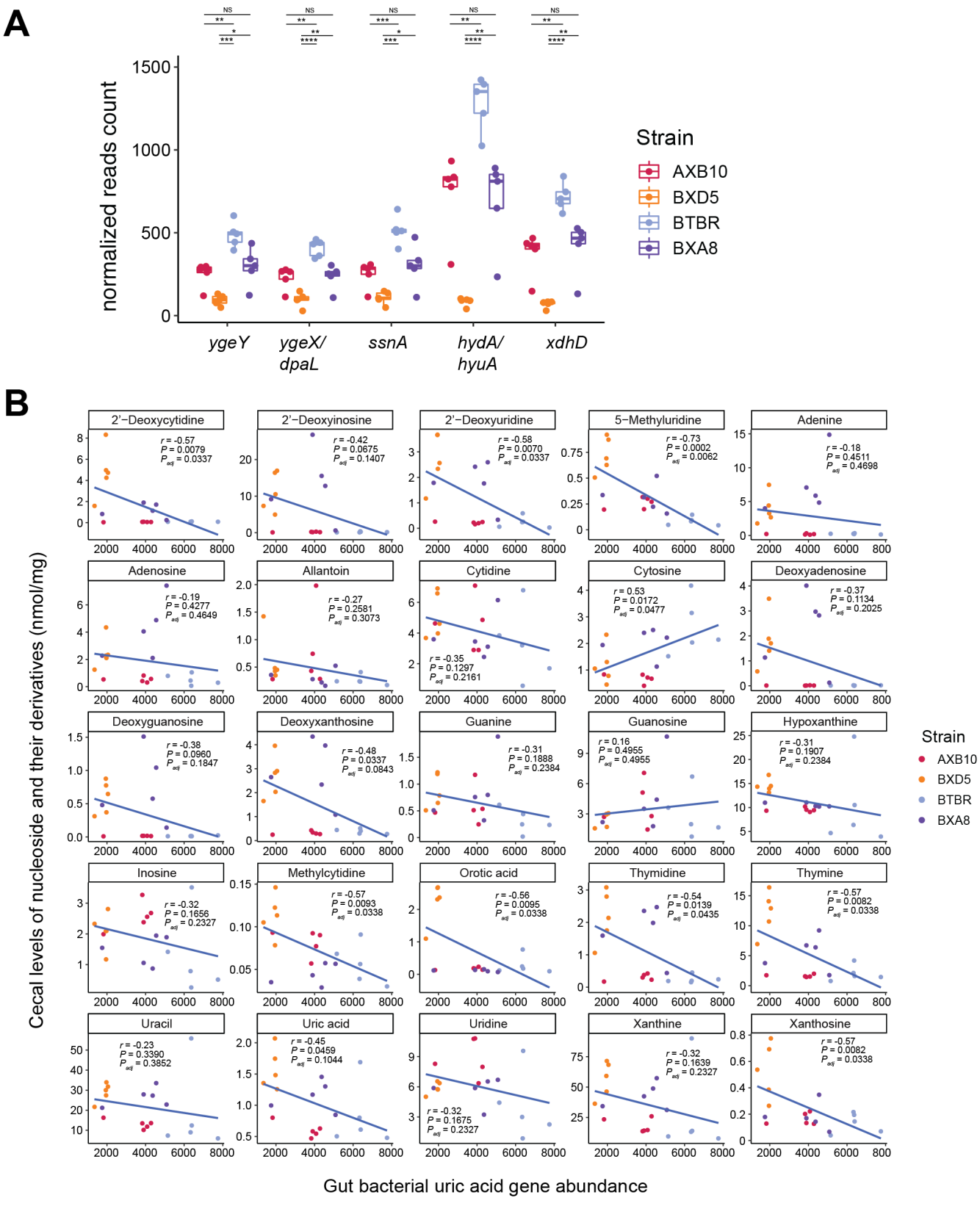
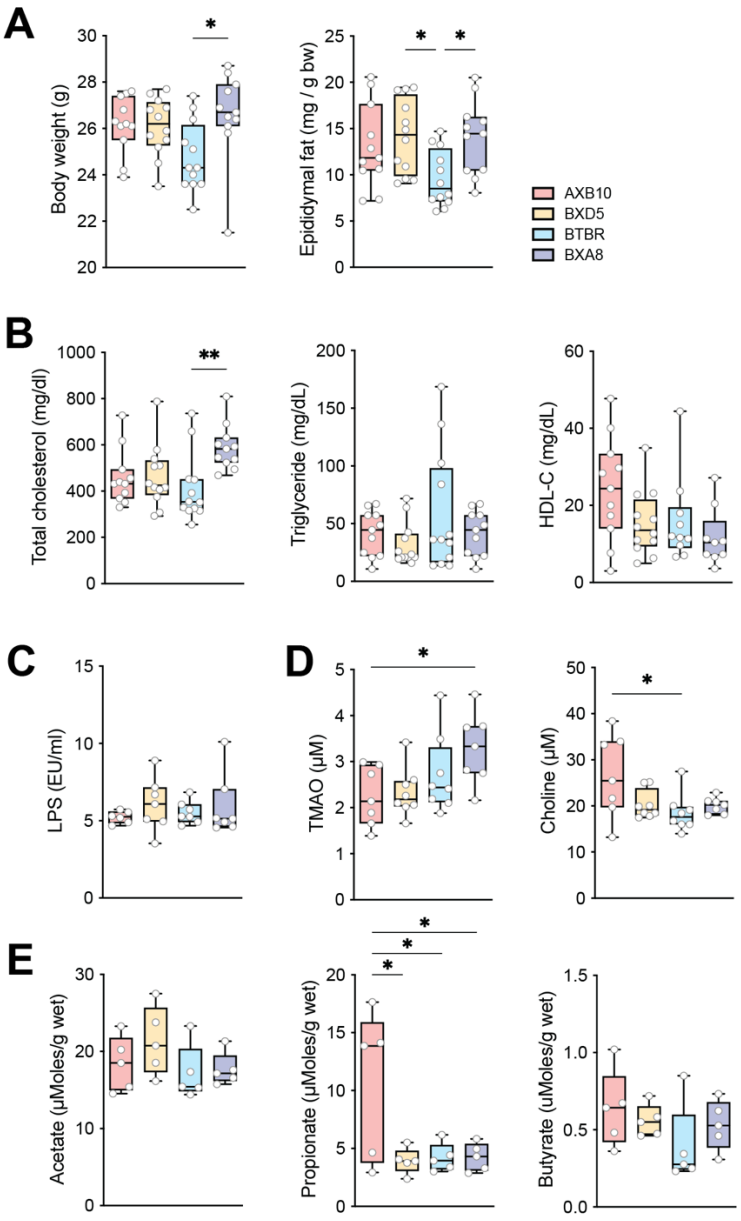

A
normalized reads count
1500
1000
500
0
ygeY
ygeX/ dpaL
ssnA
hydA/ hyuA
xdhD
Strain
AXB10
BXD5
BTBR
BXA8
B
Cecal levels of nucleoside and their derivatives (nmol/mg)
2'−Deoxycytidine
r = -0.57 P = 0.0079 Padj = 0.0337
2'−Deoxyinosine
r = -0.42 P = 0.0675 Padj = 0.1407
2'−Deoxyuridine
r = -0.58 P = 0.0070 Padj = 0.0337
5−Methyluridine
r = -0.73 P = 0.0002 Padj = 0.0062
Adenine
r = -0.18 P = 0.4511 Padj = 0.4698
Adenosine
r = -0.19 P = 0.4277 Padj = 0.4649
Allantoin
r = -0.27 P = 0.2581 Padj = 0.3073
Cytidine
r = -0.35 P = 0.1297 Padj = 0.2161
Cytosine
r = 0.53 P = 0.0172 Padj = 0.0477
Deoxyadenosine
r = -0.37 P = 0.1134 Padj = 0.2025
Deoxyguanosine
r = -0.38 P = 0.0960 Padj = 0.1847
Deoxyxanthosine
r = -0.48 P = 0.0337 Padj = 0.0843
Guanine
r = -0.31 P = 0.1888 Padj = 0.2384
Guanosine
r = 0.16 P = 0.4955 Padj = 0.4955
Hypoxanthine
r = -0.31 P = 0.1907 Padj = 0.2384
Inosine
r = -0.32 P = 0.1656 Padj = 0.2327
Methylcytidine
r = -0.57 P = 0.0093 Padj = 0.0338
Orotic acid
r = -0.56 P = 0.0095 Padj = 0.0338
Thymidine
r = -0.54 P = 0.0139 Padj = 0.0435
Thymine
r = -0.57 P = 0.0082 Padj = 0.0338
Uracil
r = -0.23 P = 0.3390 Padj = 0.3852
Uric acid
r = -0.45 P = 0.0459 Padj = 0.1044
Uridine
r = -0.32 P = 0.1675 Padj = 0.2327
Xanthine
r = -0.32 P = 0.1639 Padj = 0.2327
Xanthosine
r = -0.57 P = 0.0082 Padj = 0.0338
Strain
AXB10
BXD5
BTBR
BXA8
Gut bacterial uric acid gene abundance

Figure A.7 Correlation between cecal purines and gut bacterial genes encoding uric acid degradation in transplanted *ApoE* KO mice

(A) Abundance of genes encoding anaerobic purine degradation in gut metagenomes from gnotobiotic mice transplanted with cecal contents from strains with disparate atherosclerosis phenotypes (see Figure A.1). Data are shown as box-and-whisker plots with individual data points, where the boxes indicate the median values and the interquartile ranges and the whiskers represent the minimum and maximum values. Differences between groups were evaluated using unpaired two-tailed Welch's t test. $*$ p value < 0.05; $**$ p value < 0.01; $***$p value < 0.001; $****$p value < 0.0001.

(B) Correlation between abundance of nucleosides and their derivatives and bacterial functions was performed using Spearman correlation.



Supplementary Figure A.1 Body weight, lipid profile, and microbial metabolites from transplanted mice, related to Figure A.1

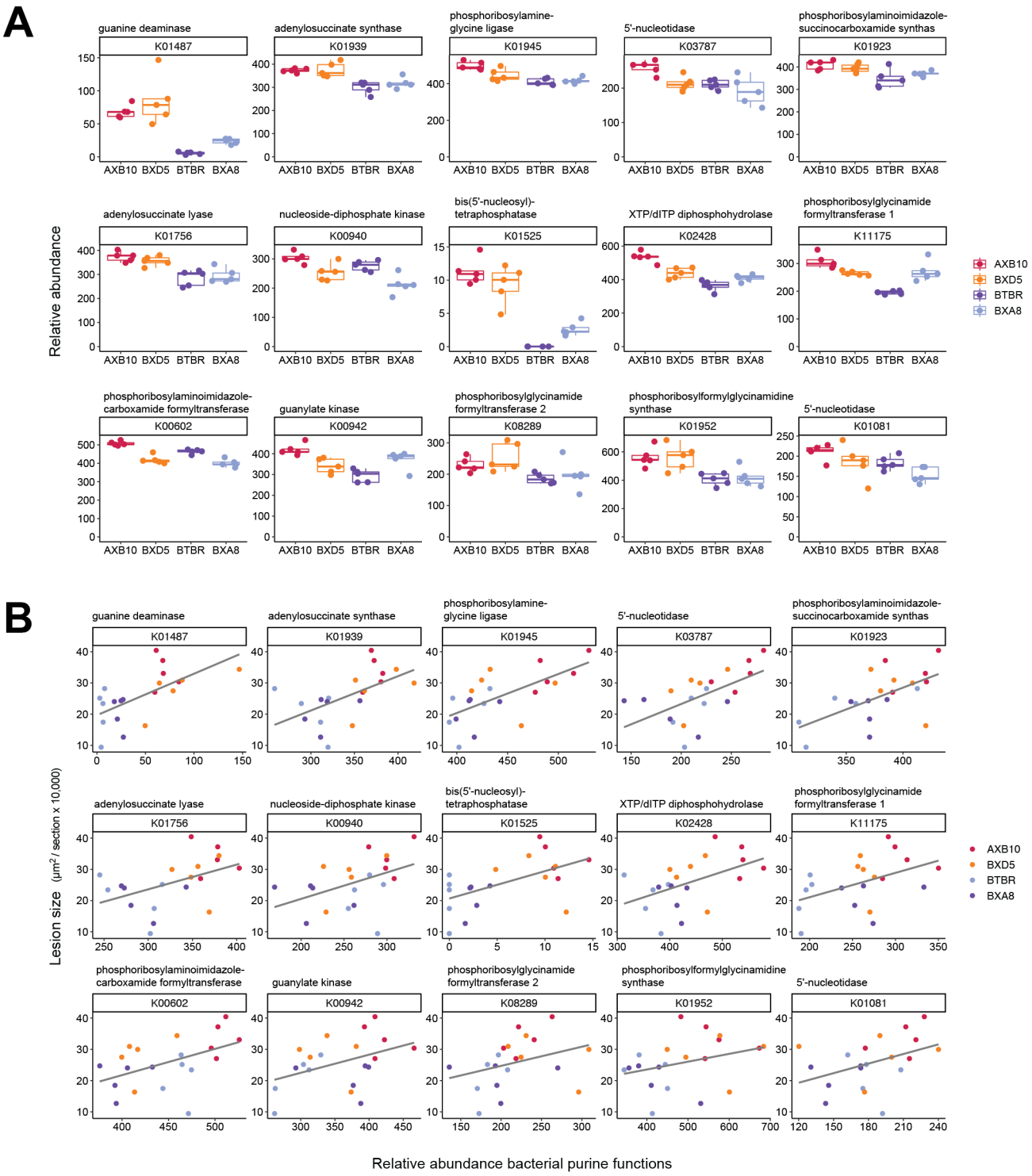
A) Body weight and epididymal fat weight collected at the end of the experiment.

B) Total cholesterol, triglyceride, and HDL-cholesterol in plasma.

C) Plasma lipopolysaccharide (LPS) levels.

D) Plasma TMAO and choline levels.

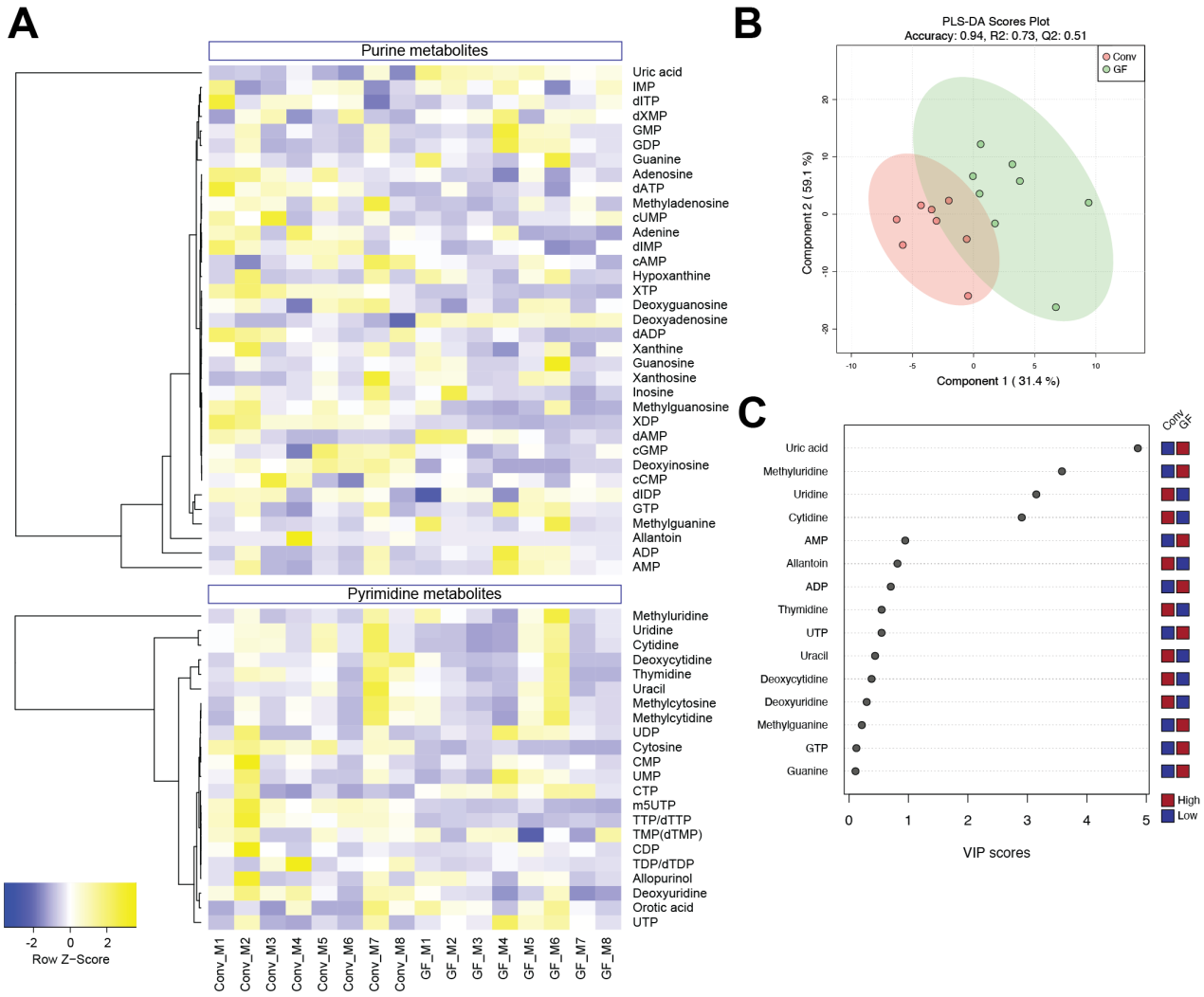
E) Cecal acetate, propionate, and butyrate levels. Data are shown as box-and-whisker plots with individual data points, where the boxes indicate the median values and the interquartile ranges and the whiskers represent the minimum and maximum values. Significance was calculated by one-way ANOVA with the Tukey post-tests and is reported as follows: *, p-value of <0.05; **, p-value of <0.01.



Supplementary Figure A.2 Metagenomic analysis of transplanted *ApoE* knockout mice, related to Figure A.1

A) Relative abundance of molecular functions (KEGG Orthology, KO) involved in purine metabolism in transplanted *ApoE* knockout mice. Data are shown as box and-whisker plots with individual data points, where the boxes indicate the median values and the interquartile ranges and the whiskers represent the minimum and maximum values.

B) Spearman correlation between bacterial KO involved in purine and atherosclerosis lesion size.



Supplementary Figure A.3 Targeted purine metabolite quantitation in plasma samples from Conv and GF mice, related to Figure A.3

A) Heatmap of purines and related metabolites in plasma samples from Conv (n=8) and GF (n=8) mice analyzed by LC-MS/MS.

B) PLS-DA plot based on the data derived from purine metabolites in plasma samples from Conv and GF mice.

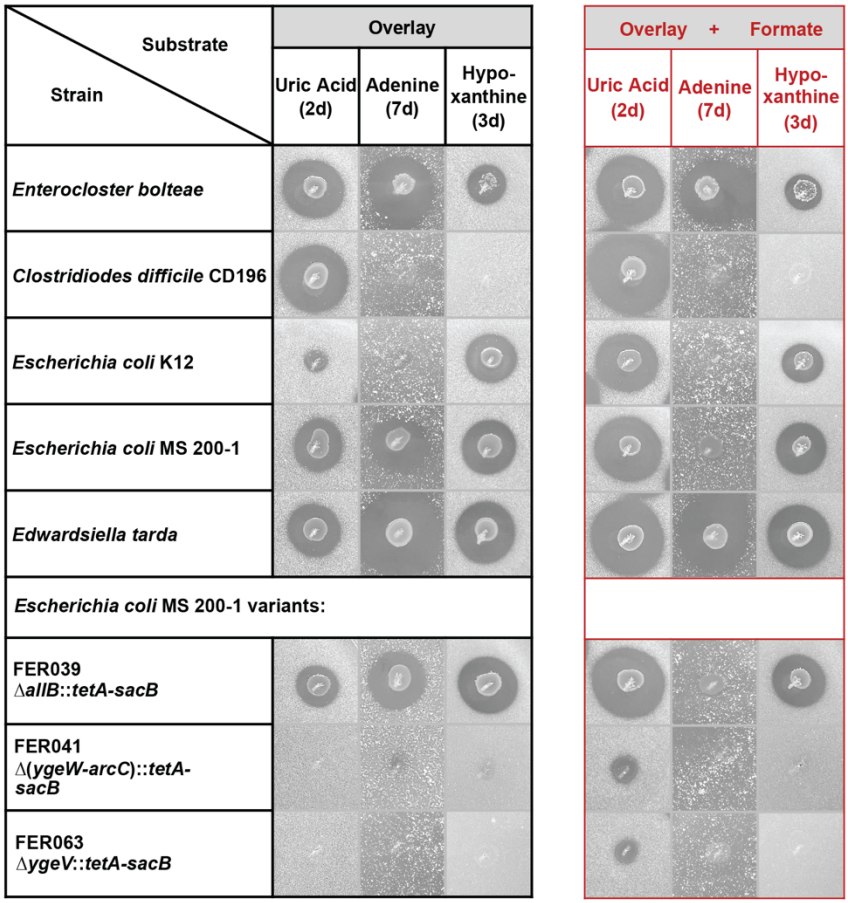
C) VIP plot indicating the most discriminating metabolites in descending order of importance. The colored boxes on the right indicate the relative concentrations of the corresponding metabolite in

each group. Conv; Conventionally-raised, GF; germ-free, PLSDA; Partial Least Squares Discriminant Analysis, VIP; variable importance of projection.

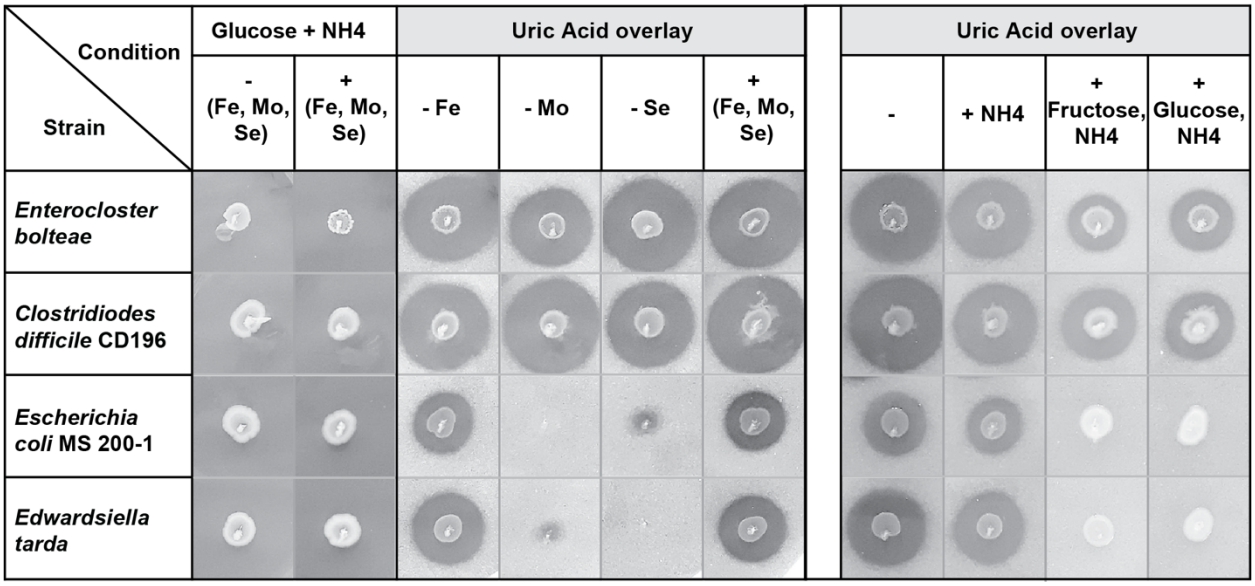
Supplementary Figure A.4 Screen of gut bacterial isolates for growth on purines, related to Figure A.4

Overnight bacterial cultures were spotted (4 µl) onto medium 26B agar plates, and plates containing soluble additions (NH_4 = 10 mM NH_4Cl, Glucose = 25 mM glucose, Allantoin = 25 mM Allantoin) or overlays containing saturating levels of uric acid (UA), Adenine, UA plus formate (25 mM) or UA plus glucose (25 mM), as detailed in Methods. Plates were incubated anaerobically at 37°C for 2 (all except with adenine overlay) or 7 days (adenine overlay). * indicates no test performed.

A


Substrate
Strain
Overlay
Uric Acid (2d)
Adenine (7d)
Hypo-xanthine (3d)
Overlay + Formate
Uric Acid (2d)
Adenine (7d)
Hypo-xanthine (3d)
Enterocloster bolteae
Clostridiodes difficile CD196
Escherichia coli K12
Escherichia coli MS 200-1
Edwardsiella tarda
Escherichia coli MS 200-1 variants:
FER039 ΔallB::tetA-sacB
FER041 Δ(ygeW-arcC)::tetA-sacB
FER063 ΔygeV::tetA-sacB

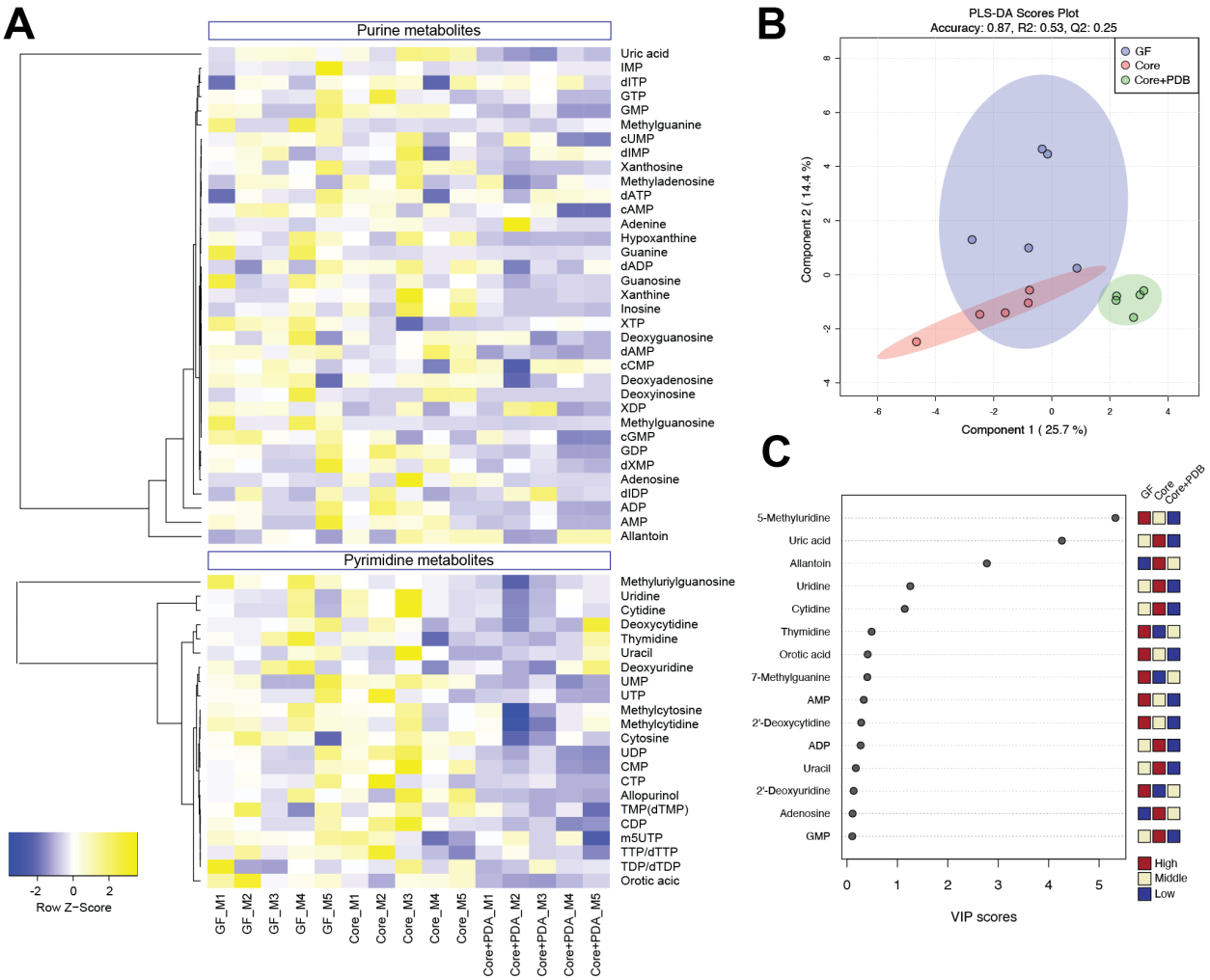
B


Condition
Strain
Glucose + NH4
- (Fe, Mo, Se)
+ (Fe, Mo, Se)
Uric Acid overlay
- Fe
- Mo
- Se
+ (Fe, Mo, Se)
Uric Acid overlay
-
+ NH4
+ Fructose, NH4
+ Glucose, NH4
Enterocloster bolteae
Clostridiodes difficile CD196
Escherichia coli MS 200-1
Edwardsiella tarda

Supplementary Figure A.5 Environmental factors influence purine utilization, related to Figure A.4

A) Formate. Several Firmicutes and Proteobacteria species were spotted onto purine (Uric Acid, Adenine, Hypoxanthine) overlay plates and otherwise identical media supplemented with filter-sterilized formate (pH 7, 25 mM in both the base and overlay layers). As previously reported for cell suspensions of *Escherichia coli* K12, UA utilization was enhanced in the presence of formate. Similar results are evident for *E. coli* MS 200-1 and the purine-utilizing *allB* variant (FER039) as well as for *Edwardsiela tarda*. The enhanced UA utilization is less pronounced with the two tested Firmicutes (*Enterocloster bolteae* and *Clostridiodes difficile*), and is not evident for any strain for adenine and hypoxanthine, which are more reduced than UA. The slight utilization of UA in the presence of formate by the variants FER041 and FER063 suggests the presence of a second, perhaps adventitious, UA utilization system.

B) Trace minerals and sugars. Trace minerals: Plates containing 20 mM glucose + 10 mM NH_4Cl, or bilayer uric acid overlay plates were prepared with different trace mineral compositions: i) containing all trace element additions (see Methods with 2.5 µM Fe, 5 µM Mo and 0.5 µM Se) or ii) lacking the addition of the indicated trace element. No attempt was made to rigorously remove the "missing" minerals, and plates were prepared with otherwise standard cysteine·HCl-reduced, phosphate-buffered medium 26B containing 0.1% yeast extract and Difco Bacto agar. Sugars: Standard bilayer uric acid plates were prepared (see Methods) or supplemented with filter-sterilized stocks of NH_4Cl (to 10 mM), fructose and NH4Cl (to 40 and 10 mM, respectively), or glucose + NH_4Cl (to 40 and 10 mM, respectively). Plates were spotted with 4 µl of cultures freshly-grown in rich medium and incubated anaerobically for 2 days at 37°C.



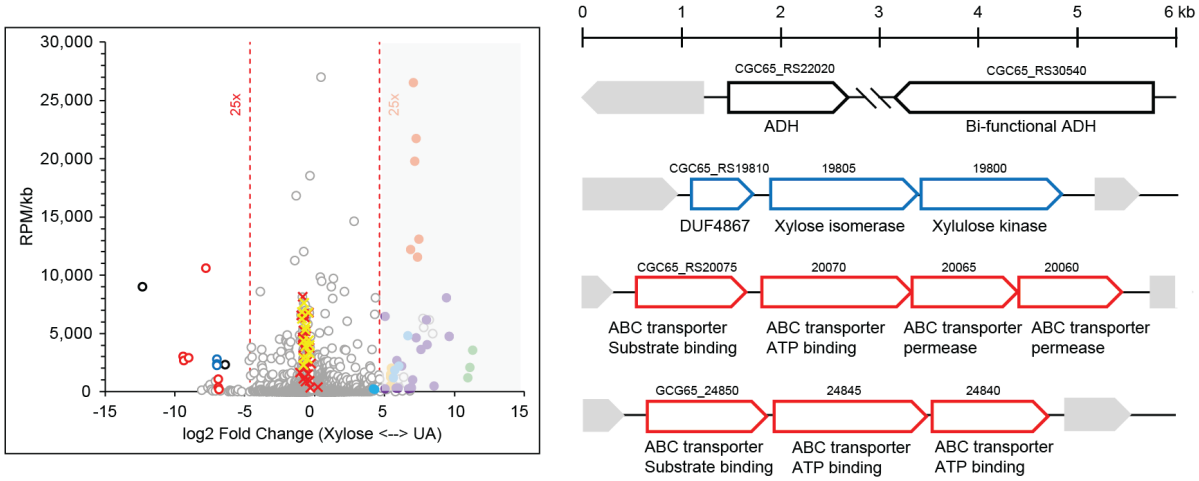
Supplementary Figure A.6 Targeted purine metabolite quantitation in plasma samples from GF and gnotobiotic mice, related to Figure A.5

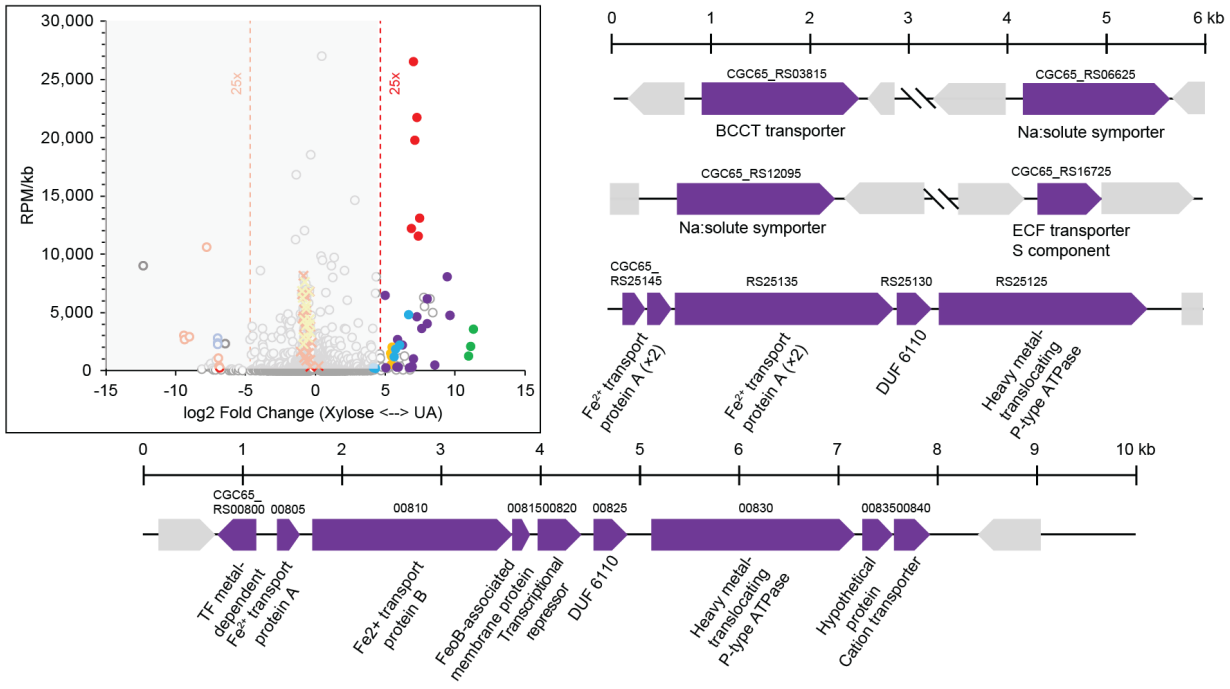
A) Heatmap of purines and related metabolites in plasma samples from GF (n=5), 'core' (n=3) and 'core plus PDB' (n=5) mice analyzed by LC-MS/MS.

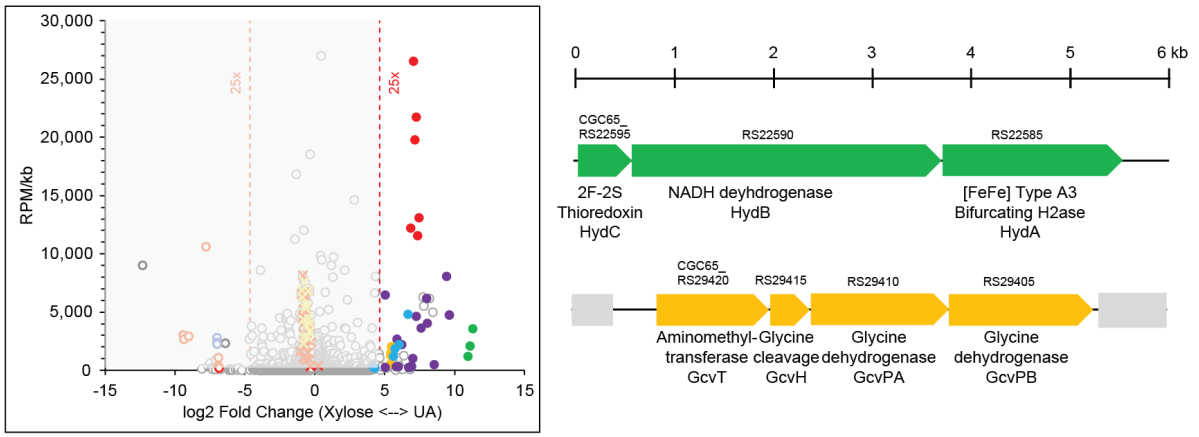
B) PLS-DA plot based on the data derived from purine metabolites in plasma samples from GF, 'core' and 'core plus PDB' mice.

C) Variable Importance Projection plot indicating the most discriminating metabolites in descending order of importance. The colored boxes on the right indicate the relative concentrations of the corresponding metabolite in each group. GF; germ-free, PDB; purine-

degrading bacteria, PLS-DA; Partial Least Squares Discriminant Analysis, VIP; variable importance of projection.


RPM/kb
log2 Fold Change (Xylose <--> UA)
25x
0 1 2 3 4 5 6 kb
CGC65_RS22020
CGC65_RS30540
ADH
Bi-functional ADH
CGC65_RS19810
19805
19800
DUF4867
Xylose isomerase
Xylulose kinase
CGC65_RS20075
20070
20065
20060
ABC transporter Substrate binding
ABC transporter ATP binding
ABC transporter permease
ABC transporter permease
GCG65_24850
24845
24840
ABC transporter Substrate binding
ABC transporter ATP binding
ABC transporter ATP binding


RPM/kb
log2 Fold Change (Xylose <--> UA)
25x
0 1 2 3 4 5 6 kb
CGC65_RS03815
CGC65_RS06625
BCCT transporter
Na:solute symporter
CGC65_RS12095
CGC65_RS16725
Na:solute symporter
ECF transporter S component
CGC65_RS25145
RS25135
RS25130
RS25125
Fe2+ transport protein A (x2)
Fe2+ transport protein A (x2)
DUF 6110
Heavy metal-translocating P-type ATPase
0 1 2 3 4 5 6 7 8 9 10 kb
CGC65_RS00800
00805
00810
0081500820
00825
00830
0083500840
TF metal-dependent Fe2+ transport protein A
Fe2+ transport protein B
FeoB-associated membrane protein
Transcriptional repressor
DUF 6110
Heavy metal-translocating P-type ATPase
Hypothetical protein
Cation transporter


RPM/kb
log2 Fold Change (Xylose <--> UA)
25x
0 1 2 3 4 5 6 kb
CGC65_RS22595
RS22590
RS22585
2F-2S Thioredoxin HydC
NADH deyhdrogenase HydB
[FeFe] Type A3 Bifurcating H2ase HydA
CGC65_RS29420
RS29415
RS29410
RS29405
Aminomethyl-transferase GcvT
Glycine cleavage GcvH
Glycine dehydrogenase GcvPA
Glycine dehydrogenase GcvPB

Supplementary Figure A.7 Comparison of transcriptional profiles for *Enterocloster bolteae* grown on xylose vs. uric acid, related to Figure A.6

Plot showing differentially-expressed genes (FDR < 0.01) and reads per million (RPM)/ gene size (kb) for *Enterocloster bolteae* grown on xylose + NH4Cl (upregulated genes to the left) or uric acid (upregulated genes to the right). Genes encoding 30S and 50S RNA Polymerase (RNAP) subunits are indicated near the center of the figure (yellow and red “x” symbols, respectively) with a slight bias (1.6-fold) towards the xylose substrate side (left) in good agreement with the faster growth rate observed on this substrate and growth rate-limiting nature of RNAP subunit expression. Growth on xylose + NH_4Cl elicited high expression of genes for sugar transport functions, an operon encoding xylose-utilization proteins, and alcohol dehydrogenases, the latter consistent with the accumulation of ethanol in these cultures (not shown). In addition to the two operons described in the manuscript, growth on uric acid also induced high expression of micronutrient transport functions, one of three glycine cleavage systems, and a bifurcating hydrogenase system. Relevant genes are indicated in the right-hand panels, color-coded according to the expression plots shown on the left.

APPENDIX B: Dissecting the impact of dietary fiber type on atherosclerosis in mice colonized with different gut microbial communities

The work presented in this appendix has been published:

*Hutchison ER[*], Ksahara, K, **Zhang Q**, Vivas EI, Cross TW, Rey FE[+]*

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

Data and supplementary information available online

B.1 Abstract

Dietary fiber consumption has been linked with improved cardiometabolic health, however, human studies have reported large interindividual variations in the observed benefits. We tested whether the effects of dietary fiber on atherosclerosis are influenced by the gut microbiome. We colonized germ-free *ApoE*$^{-/-}$ mice with fecal samples from three human donors (DonA, DonB, and DonC) and fed them diets supplemented with either a mix of 5 fermentable fibers (FF) or non-fermentable cellulose control (CC) diet. We found that DonA-colonized mice had reduced atherosclerosis burden with FF feeding compared to their CC-fed counterparts, whereas the type of fiber did not affect atherosclerosis in mice colonized with microbiota from the other donors. Microbial shifts associated with FF feeding in DonA mice were characterized by higher relative abundances of butyrate-producing taxa, higher butyrate levels, and enrichment of genes involved in synthesis of B vitamins. Our results suggest that atheroprotection in response to FF is not universal and is influenced by the gut microbiome.

B.2 Introduction

Individual responses to the same diet or therapeutic drugs are often inconsistent and not universal. This notion is a fundamental principle of precision medicine and nutrition (Denson et al., 2019; Ordovas & Berciano, 2020). Many factors influence how a subject responds to a given treatment including genetics, diet, and sex. Recently, it has become apparent that the gut microbiome is a major contributor to the observed interpersonal variation in responsiveness (Deehan et al., 2020; Hughes et al., 2019; Leshem et al., 2020; Zeevi et al., 2015). It is now widely recognized that the gut microbiome plays a significant role in health and its composition is highly variable among individuals (Turnbaugh et al., 2007). Dietary components, from foodstuffs to orally administered drugs, come in close contact with resident microbes along the gastrointestinal tract. The gut microbiome collectively encodes >100-fold more genes than the human genome, including a rich array of enzymes with the potential to metabolize these ingested compounds and modulate their bioavailability, activity, and ultimately their effects on the host (Javdan et al., 2020; Zimmermann et al., 2019a, 2019b). Indeed, gut microbes have received considerable attention in recent years for their capacity to modulate responses to bioactive compounds (Kolodziejczyk et al., 2019) ranging from antihypertensive drugs to immunosuppressants for organ transplants (Koppel et al., 2018; Lee et al., 2015). Gaining a better understanding of which interventions are most sensitive to microbiome variation is critical for the effective implementation of precision medicine.

Cardiovascular disease (CVD) is the leading cause of death in the United States and accounts for over a third of all deaths globally (Ahmad et al., 2021). Atherosclerosis is the most common manifestation of CVD and is driven by inflammatory processes that result in the formation of macrophage-dense, fatty plaques within the arterial wall (Russell, 1999). There is increasing evidence that the gut microbiome plays an important role in modulating atherosclerosis development. Epidemiological studies have identified differences in the microbiomes of individuals with coronary artery disease compared to healthy individuals (Jie et al., 2017; Kelly et

al., 2016; Tang & Hazen, 2017). Furthermore, several microbial metabolites arising from specific dietary components have been shown to modulate atherosclerosis progression in humans and animal models through a variety of mechanisms. For example, trimethylamine *N*-oxide, a microbial derivative of choline, is associated with increased risk of major cardiovascular events in humans (Wang et al., 2011); the microbial metabolite indole-3-propionic acid, which is derived from tryptophan, protects against atherosclerosis progression by promoting cholesterol efflux (Xue et al., 2022); and short-chain fatty acids (SCFAs), which are produced via fermentation of dietary fiber, have been shown to ameliorate atherosclerosis by limiting dietary cholesterol absorption (propionate) and reducing inflammation and gut permeability (butyrate) (Aguilar et al., 2014; Haghikia et al., 2022; Kasahara et al., 2018). Indeed, diet has long been known to play a major role in both the promotion and prevention of atherosclerosis (Kritchevsky, 1978; Torres et al., 2015). For example, it is well-established that foods such as whole-grain cereals and legumes which are rich in dietary fiber, are protective against CVD (Higginson & Pepler, 1954; Threapleton et al., 2013). However, inconsistent responses to a number of dietary and pharmacological interventions for CVD have been observed between individuals (Healey et al., 2017; Weeke & Roden, 2013). Most studies linking dietary fiber to improved cardiovascular health are assessed using population averages (Kirk et al., 2021) and do not account for individual characteristics. Therefore, the causes behind these inconsistencies are understudied.

Dietary fibers are oligo- or polysaccharides that resist degradation by host enzymes and are available to be metabolized by microbes in the distal gut. Dietary fibers vary widely in structure and composition and are often subdivided according to their biochemical properties. One such division is drawn by whether they can be fermented by gut microbes. Thus, fermentable fibers are dietary fibers that can be metabolized by intestinal microbes, while non-fermentable fibers resist intestinal fermentation (Cummings, 1984). By this definition, fermentability is not a static or inherent property of any given fiber since it is context-dependent and is contingent on the presence of specific microbes that can degrade the fiber in question and the host environment

(e.g., transit time, rumination). Fermentation of dietary fiber in the gastrointestinal tract results in the production of SCFAs, the most abundant of which are acetate, propionate, and butyrate. SCFAs have been linked to improved cardiometabolic health (Haghikia et al., 2022; Kasahara et al., 2018) and are hypothesized to mediate some of the atheroprotective effects associated with dietary fiber consumption (Ohira et al., 2017). However, multiple studies have shown that SCFA production is dependent on microbiome structure and is highly variable among individuals. For example, McOrist et al. found individualized responses in butyrate production to a resistant starch (RS) dietary supplement (McOrist et al., 2011). Additionally, we recently found that various fermentable fibers (pectin, inulin, fructo-oligosaccharide (FOS), RS-2, and RS-4) elicited disparate responses in SCFA production when fed to mice colonized with different microbial communities (Murga-Garrido et al., 2021).

Given the personalized nature of the gut microbiome and the fact that gut microbes are necessary for metabolizing dietary fiber, we hypothesized that the atheroprotective effects attained in response to a given dietary fiber are modulated by the gut microbiome composition of the consumer. To test this, we colonized groups of germ-free (GF) *apolipoprotein E* deficient (*ApoE*$^{-/-}$) mice with fecal microbial communities from one of three human donors, each of which exhibited divergent microbial compositions and SCFA-producing capacities. Colonized mice were fed a diet containing either a mixture of fermentable fibers (FF) or a non-fermentable cellulose control diet (CC). We found that protection from atherosclerosis by FF consumption was not universal, but was instead modulated by the resident microbiome. We also observed that atheroprotection was associated with increased butyrate production and enrichment for bacterial genes involved in pathways for carbohydrate metabolism and vitamin synthesis.

B.3 Results

Engraftment of donor communities prior to dietary treatment

Germ-free (GF) female *ApoE*$^{-/-}$ mice were colonized with fecal samples from one of three human donors (Supplementary Figure B.1). These samples were selected from a repository of fecal specimens previously collected from adults in their mid-seventies (Herd et al., 2014) and were chosen based on (i) their divergent community structure as assessed by unweighted UniFrac distances of 16S rRNA profiles and (ii) their capacity to generate differing levels of SCFAs when engrafted in GF mice consuming a semi-purified diet containing an assortment of fibers (Murga-Garrido et al., 2021). After colonization, mice were maintained on the FF diet for two weeks to allow for stabilization of engrafted communities before the dietary treatment phase (Supplementary Figure B.1). Fecal samples were collected at this point to assess bacterial engraftment via 16S rRNA V4 amplicon sequencing. The engraftment efficiency (genus level) was 64% in DonA-, 65% DonB-, and 72% DonC-colonized mice, respectively (Supplementary Figure B.2f). When calculated as the percentage of donor genera detected in at least one of the recipient mice, efficiencies were 90, 88, and 87% in DonA-, DonB-, and DonC-colonized mice, respectively (Supplementary Figure B.2c–e). These engraftment efficiencies are in line with previous studies (Goodman et al., 2011; Turnbaugh, Ridaura, et al., 2009). Physiological, anatomical, and behavioral differences between human donors and recipient mice along with differences in diet likely explain why only a fraction of the donor bacteria engrafted.

A few genera were only detected in the recipient mice but were unique to each of the three donor groups, suggesting that these taxa may have been present in low abundance in the human donor samples rather than the result of contamination. Principal coordinate analysis (PCoA) of weighted UniFrac distances of fecal samples collected prior to dietary treatment showed uniform engraftment between mice bound for the two diets (FF-bound and CC-bound) within all treatment groups (all pairwise PERMANOVA adjusted $P > 0.1$, Supplementary Figure B.2a). However, comparisons using unweighted UniFrac distances (sensitive to presence/absence of taxa)

showed a significant difference in community structure in DonA-colonized mice between FF-bound and CC-bound communities (adjusted $P = 0.0012$, Supplementary Figure B.2b). This was driven by 9 genera that were detected in one diet-bound group but not the other (Supplementary Figure B.2c). Eleven weeks after dietary treatment, cecal samples were collected and used to assess terminal microbial communities. By the end of the experiment, 5 of the 9 missing genera were no longer detected in cecal contents of mice on either diet, while 4 genera (*Clostridium*, *Faecalibacterium*, *Gemmiger*, and an undetermined *Ruminococcus* genus) were found only in FF-fed mice (Supplementary Figure B.2c). This introduces the possibility that the differences observed in the assembled communities between dietary groups for DonA mice are the result of inconsistent engraftment rather than an effect of diet. Alternatively, since microbial communities undergo considerable fluctuations in the period after colonization (Chung et al., 2012), it is possible that these missing taxa were present in the CC-bound mice, but below detectable levels. The latter scenario is supported by the fact that (i) all of the missing taxa were detected in the human donor sample used to inoculate all DonA mice, and (ii) similar FF-diet-driven patterns were observed with *Faecalibacterium* and *Gemmiger* abundances in a previous study (Murga-Garrido et al., 2021) that used the same donor feces and the same diets. These findings highlight the importance of reporting pre-treatment engraftment data in mouse transplant studies such that the conclusions can be appropriately contextualized.

Diet-induced shifts in microbiota composition are largely community-specific

Two weeks after colonization, mice were placed into their dietary treatment groups and were maintained on their respective diets for 11 weeks. PCoA analysis of weighted and unweighted UniFrac distances (Figure B.1a,b) of cecal bacterial communities assessed at the completion of the study shows that mice were highly distinguishable by dietary treatment within each donor group. When using unweighted UniFrac distances, ordination shows that donor group had a stronger effect on community composition than diet (Figure B.1a). PCoA of weighted

UniFrac distances, which factors in the abundance of each taxon, shows clear distinctions by diet, but more overlap between donor groups (Figure B.1b), suggesting that FF consumption elicits distinct changes of some low-abundance, phylogenetically-related taxa across the three communities. Alpha diversity (Shannon) was higher in mice colonized with DonA consuming FF relative to CC consumption but was not significantly impacted by diet in DonB- or DonC-colonized mice (Figure B.1f). Similarly, observed amplicon sequence variant (ASV) richness was significantly increased with the FF diet in DonA mice but was reduced by FF feeding in DonB mice (Figure B.1g). FF-consumption resulted in increased DNA concentration in cecal content, a proxy for microbial biomass (Contijoch et al., 2019), compared to CC-fed mice (Supplementary Figure B.3). This effect was observed in all three donor groups and suggests that a diet rich in fermentable fiber generally increases microbial biomass.

Bacterial communities were dominated by Firmicutes and Bacteroidetes and had detectable levels of Proteobacteria, and Verrucomicorbia in all treatment groups (Figure B.1d). Actinobacteria were detected in FF-fed DonA-colonized mice, CC-fed DonB-colonized mice, and both diets for DonC-colonized mice. FF feeding lowered the Bacteroidetes to Firmicutes ratio for all donor groups (Figure B.1e), but the only significant reduction was observed in DonA-colonized mice. There was generally little consistency in the diet-associated enrichment patterns observed across donor groups, even when considering phylum level changes, suggesting a lack of a universal response to dietary treatment (Figure B.1c–g).

We used the Microbiome Multivariable Association with Linear Models (MaAsLin 2) (Mallick et al., 2021) to analyze enrichment patterns at the genus level caused by diet for each donor group. Most genera that were significantly different (adjusted $P < 0.1$) were unique to each donor group: *Blautia* was the only genus that was significantly enriched by FF-consumption across all donor groups, while *Eggerthella, Butyricimonas*, and *Parabacteroides* were the only genera to exhibit universal enrichment in response to the CC diet (Figure B.1c). One obvious explanation for the lack of universality is the fact that each donor group possess different

collections of microbes. However, even in the cases of genera that were present across multiple donors, the directionality in response to diet tended to vary by community ([Lachn] *Clostridium*, *Bacteroides*, *Akkermansia*, [Lachn] Undetermined, [Rumin] *Clostridium*, *Oscillospira*, [Erysi] *Clostridium*, [Lachn] *Ruminococcus*). For example, the *Bacteroides* genus was significantly enriched by CC feeding in DonC-colonized mice, by FF feeding in DonB-colonized mice and unaffected by diet in DonA-colonized animals (Figure B.1c). Interestingly, the relative abundance of *Akkermansia* was significantly increased by CC feeding in DonB mice, by the FF diet in DonC mice, and was not affected by diet in DonA-colonized mice. *Akkermansia muciniphila*, the most prevalent species in this genus, feeds on host mucins which can be glycosylated by fiber-degrading bacteria thereby influencing *Akkermansia* levels (Earley et al., 2019; Zhang et al., 2022). Although *Akkermansia* has been reported to thrive on diets poor in fermentable fiber (Kim et al., 2020), our data indicate that microbiota composition can influence *Akkermansia* response to specific fiber sources. It is also possible that the different donor groups possess distinct strains of *Akkermansia muciniphilia* that are themselves differentially affected by diet. Together, these results suggest that microbial responses to dietary fiber are context-dependent and are likely impacted by the composition and metabolic capabilities of the broader community.

Fermentable fiber impacts atherosclerosis progression in a donor-dependent manner

FF-fed mice colonized with DonA had significantly reduced lipid deposition in atherosclerotic plaques and a trend towards reduced plaque area ($P = 0.070$) compared to their CC-fed counterparts, while there were no differences observed between diets in either DonB- or DonC-colonized mice (Figure B.2a–c). To further characterize atherosclerosis disease status, lesions were assessed for macrophage infiltration by immunohistology with MOMA-2 antibodies. There were no statistically significant differences observed between diets in lesion MOMA-2 density in any of the donor groups (Figure B.2a,d), but there was a trend of reduced density with

FF feeding in DonA mice ($P = 0.11$). Previous studies report inconsistent results regarding of the effect of inulin (a component fiber in the FF diet) on atherosclerosis in mice (Rault-Nania et al., 2006). Rault-Nania et al. found that inulin ameliorated atherosclerosis in *ApoE*$^{-/-}$ mice, whereas Hoving and colleagues found that inulin exacerbated atherosclerosis in *APOE*3-Leiden* mice. While this discrepancy could be due to the different diets and/or mouse models used in these studies, our results support the notion that the gut microbiome modulates the atheroprotective effect of fermentable dietary fiber, providing a possible explanation for these conflicting findings.

The atheroprotective effect of the FF diet is not associated with changes in plasma lipids or alterations in the expression of aortic immune markers

To test whether fermentable fiber consumption altered lipid composition in circulation, we measured lipid levels in the plasma of the mice described above. No statistical differences were observed between diets for any of the donor groups in plasma levels of total cholesterol, HDL cholesterol, or triglycerides (Figure B.2e–g). We also assessed aortic expression levels of *Abca1* and *Abcg1* mRNA by RT-qPCR as markers of reverse cholesterol transport but did not observe any statistically significant differences between diets within any of the donor groups (Supplementary Figure B.4d,e). These results suggest that the atheroprotective effect of FF consumption observed in DonA mice was not mediated by major alterations in plasma lipids or cholesterol homeostasis. To test whether atheroprotection was associated with changes in vascular inflammation status, we measured aortic expression of the inflammatory markers *Tnf-α*, *Il1-β*, and *Vcam-1*, which are commonly associated with atherosclerosis progression (Kasahara et al., 2018; Libby, 2012). There were no significant differences observed in expression of these markers between diets in any of the donor groups (Supplementary Figure B.4a–c). These data suggest that the atheroprotective effect of FF-feeding in DonA mice may be independent of these immune processes and reverse cholesterol transport, although further analyses are needed to fully rule out these factors.

Cecal SCFA profiles are altered by diet in a donor-dependent manner

Given the variability in microbiome composition between diets, we next tested whether there were differences in fiber fermentation capacities between donor groups. In line with the microbiome composition patterns above, diet-induced shifts in SCFA profiles were highly dependent on donor group (Figure B.3a–d). Acetate, the most abundant SCFA, was increased in FF-fed mice colonized with both DonA and DonB communities (Figure B.3a). Propionate was increased by FF-feeding only in DonA mice, whereas diet did not affect propionate levels in the other two donor groups (Figure B.3b). FF-feeding resulted in substantially elevated cecal butyrate levels in DonA mice, but reduced butyrate levels in DonB-colonized mice compared to CC-fed counterparts (Figure B.3c). Cecal butyrate concentrations were not different between diets in DonC-colonized mice. The branched-chain fatty acids isobutyrate and isovalerate, which are primarily produced via protein fermentation, were not affected by diet within any of the donor groups (Figure B.3e,f). A lack of differences in branched-chain fatty acids is consistent with the fact that the FF and CC diets are isoproteic. Total SCFA concentrations (i.e., the sum of acetate, propionate, and butyrate) within donor groups were increased by FF-feeding in DonA and DonB, but not DonC (Figure B.3d). These results reflect the shifts observed in SCFA-producing microbiota. Among the genera that were increased by FF feeding in DonA mice only were *Clostridium*, *Oscillospira*, *Ruminococcus*, *Gemmiger*, and *Faecalibacterium*, (Figure B.1c) all of which contain butyrate-producing species (Vital et al., 2014). Notably, most of these genera were also present in the other donor groups but were not enriched by FF-feeding. This could be due to complex, community-level interactions (e.g., competition) influencing responses of individual genera to dietary fiber, or strain-level differences in response to diet, or both.

Our data are consistent with previously reported studies suggesting that butyrate-induced protection against atherosclerosis in mice occurs in the absence of major changes in plasma lipid levels (Aguilar et al., 2014; Kasahara et al., 2018). Similar to our results, Kasahara et al. reported

that the introduction of a butyrate-producing microbe in mice colonized with a simplified bacterial community led to reductions in plaque burden and macrophage infiltration in *ApoE*$^{-/-}$ mice without significant changes in cholesterol homeostasis. However, Kasahara et al. also detected butyrate-induced reductions in aortic expression of inflammatory markers *Tnf-α*, *Il1-β*, and *Vcam-1*, which we did not observe in our study. Additionally, a recent study found that propionate consumption protected against atherosclerosis by inhibiting cholesterol uptake in the intestine (Haghikia et al., 2022). Although we observed increased cecal propionate levels in DonA-colonized mice consuming the FF diet, we did not detect differences in plasma cholesterol levels. The concentration and site within the gastrointestinal tract where propionate accumulates (i.e., greater concentration in the small intestine when consumed orally vs. greater concentration in the large intestine when produced via fiber fermentation) may influence its effect on cholesterol absorption. Together, these findings suggest that SCFA-production capacity is dependent on both accessible dietary fiber and microbial community composition. These results also validate the notion that the abundance of butyrate-producing microbes is associated with cecal levels of butyrate.

Bacterial functional profiles were modulated by dietary fiber in a donor-specific manner

We next sought to identify links between the functional potential of the microbiome and atheroprotection by examining changes in microbial metagenomic profiles. We performed shotgun sequencing of DNA isolated from cecal contents (average of 29.4 ± 7.7 million paired-end reads/sample; $n = 5$/diet-donor group). Sequence data were analyzed with HUMAnN3 to generate metagenomic functional profiles that included KEGG orthology (KO) abundances for each mouse. Hierarchical clustering of KO profiles using Bray–Curtis dissimilarity shows that the treatment groups are different from one another, but the effect of diet on clustering patterns varied by donor (Figure B.4a). Mice colonized with DonA and DonC communities clustered closer by diet than by donor group, suggesting a significant level of FF-influenced overlap in KO profiles between these two donor groups. DonB mice, on the other hand, clustered separately from all

other mice but sub-clustered by diet (Figure B.4a). Differential abundances of individual KOs between FF- and CC-fed mice within each donor group were calculated with MaAsLin 2. This analysis revealed that 2676 KOs were significantly different (adjusted $P < 0.05$) between diets in at least one donor group (DonA = 971; DonB = 1964; DonC = 1419). Of these, only 67 (2.5%) were enriched by FF-feeding across all donor groups, whereas 79 (3%) were enriched by CC-feeding across all donor groups, suggesting that most of the diet-induced changes in functional profiles were donor-specific. DonA- and DonC-colonized mice shared the most FF-enriched KOs with 326, while DonA- and DonB-colonized mice shared 99, and animals colonized with DonB- and DonC-colonized mice shared 81 KOs.

We next aimed to gain further insight into how dietary treatment affected the metabolic pathways of the cecal microbial communities in each donor group. We were specifically interested in identifying pathways that might help explain the atheroprotection associated with FF-feeding in DonA. We used the MicrobiomeAnalyst KEGG pathway tool (Chong et al., 2020) to conduct pathway enrichment analysis in the KOs that were overrepresented by FF-feeding relative to their counterparts consuming the CC diet. Similar to the taxonomy results discussed above, we observed a lack of universality among the metagenomic changes in response to diet. Of the 38 KEGG pathways that were detected as significantly overrepresented (adjusted $P < 0.1$) by FF-feeding in at least one donor group, only three (*Pyruvate metabolism*, *Amino sugar and nucleotide sugar metabolism*, and *Biosynthesis of amino acids*) were observed across all three donor groups (Figure B.4b). In DonA-colonized mice, 27 pathways were significantly overrepresented in the FF diet relative to CC diet. These included pathways involved in vitamin synthesis (*Thiamine biosynthesis*; *Folate biosynthesis*; *Porphyrin metabolism* [vitamin B12]), SCFA synthesis (*Butanoate metabolism*; *Propionate metabolism*), and amino acid metabolism (*Lysine biosynthesis*; *Cysteine and methionine metabolism*; *Histidine metabolism*; *Phenylalanine, tyrosine and tryptophan biosynthesis*; *Valine, leucine and isoleucine biosynthesis*; *Glycine, serine and threonine metabolism*; *Biosynthesis of amino acids*) (Figure B.4b). Interestingly, the dietary

effects on the enrichment of pathways involved the synthesis of acetate (*Carbon metabolism, Glyoxylate and dicarboxylate metabolism*, and *Pyruvate metabolism*), propionate (*Propionate metabolism*), and butyrate (*Butanoate metabolism*) corresponded very closely with the cecal SCFA levels described above (Figure B.4b and Figure B.3a–c).

Both folate and vitamin B12 are involved in the detoxification of homocysteine, a metabolite of methionine metabolism that has been linked to cardiovascular disease. A study involving *ApoE*$^{-/-}$ mice with hyperhomocysteinemia found that supplementation with a mixture of folate, vitamin B12, and vitamin B6 protected against atherosclerosis (Hofmann et al., 2001). Moreover, a recent metagenomic study in humans showed that patients with CVD ($n = 218$) had decreased abundance of genes encoding for components of the folate biosynthesis pathway than healthy patients ($n = 187$) (Jie et al., 2017). Interestingly, the authors of that study also found that CVD was associated with lower abundances of propionate and butyrate synthesis genes. To gain a more detailed picture of the metagenomic dynamics of these pathways, we compared the differential abundances of the individual KOs involved in folate biosynthesis (KEGG map00790) and anaerobic cobalamin (vitamin B12) biosynthesis (KEGG M00924). In agreement with our enrichment analysis, most of the differentially abundant KOs in both pathways were significantly upregulated by FF feeding in DonA-colonized mice, but not in the other groups (Figure B.4c). Folate and vitamin B12 were supplied in the FF and CC diets at the same inclusion rate (AIN-93 vitamin mix, Supplementary Table 1), but it is possible that some amount of additional vitamin availability via microbial biosynthesis may have a physiological effect in host homocysteine metabolism. These results suggest that microbial production of vitamins B12 and folate may act as a potential mediator of the atheroprotection associated with FF diet in mice colonized with this community.

To uncover associations between atheroprotection and fiber metabolism, we determined the level and type of carbohydrate-active enzyme (CAZyme) families between dietary treatments within each donor group using cecal metagenomic data. We detected a number of CAZyme

families that were highly abundant in all donor groups and largely unaffected by diet (Figure B.5a,b, Supplementary Figure B.5). Differential abundance analysis revealed that the CAZyme families which were most significantly affected by diet were about 100-fold lower than the highest abundance CAZymes (Figure B.5c). Given the differences in cecal SCFA levels between treatment groups, this suggests that these highly-differential, low-abundance CAZymes have an outsized impact on the dynamics of SCFA metabolism. To highlight the most differentially abundant CAZyme families, we compared the abundances of the CAZyme families that were most affected by diet (top 10% by MaAsLin 2 effect size within each donor group, Figure B.5c). The vast majority of FF-enriched CAZymes in DonA-colonized mice were significantly correlated (Spearman, $P < 0.05$) with cecal butyrate levels, potentially linking them to butyrate production (Figure B.5c). One such CAZyme family, GH59, encompasses β-galactosidases which free terminal β-D-galactose monomers from galactan side chains of pectin (Cankar et al., 2014). Interestingly, many commonly cited CAZyme families involved in inulin, pectin, RS-2/4, and scFOS were not found among the most highly differentially abundant CAZymes in our dataset. It is possible that the inclusion of multiple fermentable fibers creates competition among microbes that are specialized for each fiber type, reducing the magnitude of changes detected in fiber-specific CAZymes.

B.4 Discussion

In summary, we showed that the gut microbiome regulates the effect of dietary fiber on atherosclerosis development in gnotobiotic *ApoE*$^{-/-}$mice colonized with different human fecal communities. We found that diet-induced shifts in microbial composition, metabolic potential, and metabolic output (SCFAs) varied among the different donor groups. Our results showed that atheroprotection was associated with increased cecal butyrate levels and abundances of butyrate-producing organisms. Additionally, shotgun metagenomic sequencing revealed donor-dependent shifts in genes involved in carbohydrate metabolism, SCFA production, and vitamin synthesis. These data support the notion that diet-associated shifts in the gut microbiota are not solely a function of diet but are instead the result of complex interactions between diet and the larger gut microbial community structure and functional network. These results are also in line with previous work showing that butyrate is atheroprotective without modifying cholesterol metabolism (Aguilar et al., 2014; Kasahara et al., 2018).

The current study has some limitations that should be addressed. First, we observed an imperfect engraftment efficiency from human donor to mouse recipient. As discussed above, we detected differences in the pre-treatment engraftment patterns of DonA mice. Our data suggest that this difference in detection was likely a consequence of the stochasticity of microbial communities shortly (two weeks) after colonization and not a result of differences in inoculation or contamination. Nonetheless, this discrepancy introduces the possibility that the differences observed in atherosclerosis within DonA mice were due to inconsistent engraftment rather than response to diet. Another limitation is that our study only used three human donors. A much larger and more diverse cohort of donors would be needed to fully appreciate the breadth of cardiometabolic responses to these diets, but the limited group used here is sufficient to demonstrate that the athero-modulatory effect of dietary fiber is microbiota-dependent. This study is additionally limited by the use of only female mice, precluding us from testing the effect of sex.

Finally, the CC and FF diets used in this study differed slightly in their starch content, which may contribute to the differences described above.

Despite these limitations, the work presented here suggests that microbiome variation modulates responses to dietary fiber consumption, which can differentially impact the development of atherosclerosis. Together, these results support the notion that dietary interventions are not universally efficacious and should be tailored to individuals. More research is needed to understand the relevant mechanisms and the metabolic and ecological dynamics that govern the microbiome-dependent individual responses to diet.

B.5 Methods

Germ-free animals

All animals in the current study were handled and maintained in accordance with the University of Wisconsin–Madison, standards for animal welfare and all protocols were approved by the university's Animal Care and Use Committee. Germ-free (GF) *ApoE*$^{-/-}$ mice (derived GF from B6.129P2-Apoe^{tm1Unc}/J; Jax 002052) were housed in a controlled environment within gnotobiotic isolators under a 12-h light/dark cycle and received autoclaved water and chow (LabDiet 5021; LabDiet, St. Louis, MO) *ad libitum*. Mice were housed with Alpha-dri® (Shepherd Specialty Papers, Kalamazoo, MI) bedding and were enriched with paper huts (Bio-Huts, Bio-Serv, Flemington, NJ) and ALPHA-twist^{TM} (Shepherd Specialty Papers). The GF status of the isolators was evaluated monthly via PCR using universal 16S rRNA primers with fecal DNA as well as a growth test of feces in rich media incubated at 37 °C aerobically and anaerobically for 7 days.

Selection of human donors

Human fecal samples used in this study were collected from participants as part of the Wisconsin Longitudinal Study (WLS) (Herd et al., 2014, 2018) and stored −80 °C. WLS data and specimen collection were approved by the UW-Madison Internal Review Board (2014-1066, 2015-0955) and written informed consent was obtained in the original study (Herd et al., 2018). In a previous publication from our group (Murga-Garrido et al., 2021), a subset of candidate WLS specimens were selected based on their distinct bacterial community structures and then subsequently transplanted into GF mice to measure cecal SCFA profiles. In the current study, we used this information to select three human donor samples: microbiota from donor WLS-sample-8 (referred to here as DonA); donor WLS-sample-1 (referred to here as DonB) and donor WLS-sample-5 (referred to here as DonC). In our previous study (Murga-Garrido et al., 2021), we found that gnotobiotic mice consuming a semi-purified diet containing an assortment of fibers that

included resistant starch type 2 and 4, short-chain fructo-oligosaccharides, inulin, and pectin colonized with DonA, DonB or DonC accumulated different levels of SCFA. Mice colonized with DonA accumulated the highest levels of cecal butyrate (~1.5 mM) among all of the donors tested, whereas DonB-colonized mice accumulated significantly lower levels of cecal butyrate (~ 0.6 mM), and mice colonized with DonC showed the highest cecal propionate levels (~10 mM) and intermediate butyrate levels (~1.0 mM) (Murga-Garrido et al., 2021). All of the WLS specimens used in the current study were obtained from subjects that self-reported consuming a western-style diet, were overweight (BMI > 25), and were not diagnosed with diabetes, cancer, or heart disease (Herd et al., 2014; Murga-Garrido et al., 2021). Identifiable information of WLS participants was blinded to the researchers in the current study.

Colonization of gnotobiotic mice with human feces and dietary treatment

At 6 weeks of age, mice were transferred to ventilated cages on an Allentown Sentry SPP IVC rack system (Allentown Inc., Allentown, NJ) and placed on irradiated FF diet which contained 10% total fiber (wt/wt) composed of 5 fermentable fibers (inulin, pectin, short-chain FOS, RS-2 and RS-4; Supplementary Figure B.1, Supplementary Table 1). Inclusion rates of each fermentable fiber source were individually adjusted based on purity and ash content to achieve an effective inclusion rate of 2% of dietary fiber from each fiber source. One week later, GF mice were colonized with microbiota from one of the WLS fecal specimens (DonA, DonB, or DonC) by a single oral gavage of a fecal slurry or by cohousing. Slurries were prepared anaerobically by homogenizing ~200 mg of frozen human feces in 5 mL of pre-reduced Mega Media (Murga-Garrido et al., 2021) in an anaerobic chamber, and then were immediately used to gavage recipient mice using syringes flushed with anaerobic atmosphere. A subset of GF mice was colonized by cohousing together with mice that had been gavage-colonized with human feces 4 weeks prior. Cohousing is an effective strategy for colonizing germ-free mice and is similar to gavage in terms of microbiota colonization and phenotype transfer (Bokoliya et al., 2021; Hansen

et al., 2012). We were unable to detect differences in cecal microbial profiles, nor did we observe significant differences in phenotypes between cohoused mice and their gavage-colonized counterparts (Supplementary Figure B.2a,b, Supplementary Table 2). Therefore, we considered all mice within the same treatment group as biological replicates regardless of colonization method. Operating under the rationale that a diet with a greater diversity of fiber sources would promote colonization of more microbes, mice were maintained on the FF diet for an additional two weeks to allow colonization to stabilize before the dietary treatment phase. Upon dietary treatment, mice either continued the FF diet or were switched to the CC diet containing 10% cellulose (Supplementary Table 1), a non-fermentable fiber control. All experimental diets in this study were vacuum packed and irradiation-sterilized by the manufacturer. The FF diet and CC diet differed only in their fiber sources. Our experimental design (Supplementary Figure B.1) resulted in six treatment groups (three donors and two diets; $n = 7–10$ mice per treatment group), each of which were conducted in two separately-caged cohorts to account for cage effects. After 11 weeks of dietary treatment, mice were sacrificed at 20 weeks of age after 4 h of fasting.

Atherosclerotic lesion analysis

Upon sacrifice, the heart was perfused with PBS buffer before being cut laterally at the mid heart and the ascending aorta to capture the aortic sinus. This tissue was embedded in OCT compound, frozen on dry ice, and stored at −80 °C until further processing. To characterize atherosclerotic plaques in the aortic sinus, the embedded tissue was sectioned on a cryostat (CM1950, Leica, Deer Park, IL) and collected on slides in 100 µm intervals, moving proximally from the base of the aortic root toward the ascending aorta. This resulted in slides containing eight equidistant sections (10 µm thickness) spanning 700 µm of the aortic sinus (0, 100, 200, 300, 400, 500, 600, 700 µm from the base of the aortic root). Formalin-fixed slides from each mouse were rinsed with 60% isopropanol for 1 min, stained for lipids using Oil Red O for 15 min, and counter-stained with hematoxylin for 1 min. Macrophage infiltration assessment of

atherosclerotic plaques was conducted by incubating formalin-fixed slides (same sectioning pattern as above) overnight with macrophage antibodies (MOMA-2, 1:50; ab33451, Abcam, Cambridge, UK) followed by incubation with secondary antibodies (1:400; ab6733, Abcam) for 1 h and streptavidin horseradish peroxidase (1:500; P0397, Agilent, Santa Clara, CA) for 15 min. Sections were then washed with PBS and counterstained with DAB for 15 s and hematoxylin for 5 s. Images of all stained sections were digitally captured and then analyzed on ImageJ (National Institutes of Health, Bethesda, MD) to measure lipid-positive area, total plaque area, and MOMA-2 positive area. Plaque areas and lipid-positive areas for each mouse are expressed as averages across all eight sections. To calculate macrophage infiltration, the three sections with the largest visible lesions were selected from each mouse and their MOMA-2 positive area densities were averaged. One sample was lost during processing, so sample sizes for atherosclerosis characterization ranged from 7–10 samples per treatment group.

Cecal short-chain fatty acids

SCFA levels in cecal contents were measured using headspace gas chromatography. Samples were prepared by adding 20–150 mg of frozen cecal contents to vials (Restek, Bellefonte, PA) containing *N* µL of water (where *N* equals 300 minus the mg of cecal content) along with 2 g of NaH_2SO_4 and 1 mL of chilled 60 µM 2-butanol as an internal standard. The preparations were immediately sealed in a GC sampling vial then allowed to sit overnight at RT. Standards for acetate, propionate, isobutyrate, butyrate, isovalerate, valerate and were combined at known concentrations (pH 7.0) and serially diluted to generate a standard curve. Vials were loaded into a HS-20 headspace sampler (Shimadzu, Columbia, OH), shaken for 20 min at 80 °C and injected onto an SH-Stabilwax 30 m column (227-36246-01, Shimadzu) connected to a flame ionization detector on a CG-2010 Plus GC (Shimadzu). Running conditions were as follows: the sample vial was equilibrated at 80 kPa for 3 min before injection; injection was performed using a 2 mL injection loop, a 12 s loading period with the transfer line at 150 °C, 1:15 split ratio, and a

N_2 column flow of 1.2 mL/min; the column temperature was maintained at 40 °C for 2 min and then increased to 200 °C at a rate of 20 °C/min, held for 2 min, and then reduced to 120 °C (−20 °C/min), reduced to 40 °C (−40 °C/min) and held at 40 °C for 1 min. Areas under the curve for each target compound were calculated with Shimadzu Lab Solution software (version 5.92) and normalized by the sample mass and the dilution factor and converted to $\mu mol \cdot g^{-1}$ using a standard curve.

Plasma triglyceride and cholesterol measurements

Blood was collected from mice while under isoflurane-induced anesthesia by cardiac puncture using an EDTA-rinsed syringe. Blood cells were separated by centrifugation then plasma was collected and stored at −80 °C. Plasma levels of triglycerides, total cholesterol, and high-density lipoprotein (HDL) cholesterol, were measured using commercially available colorimetric assay kits from Waco Diagnostics (Cat. No. 994-02891, 99902601, 997-01301, respectively; Fujifilm, Tokyo, Japan) in accordance with manufacturer's instructions.

Quantitative real-time PCR

Total RNA was extracted from frozen aorta using TRIzol reagent (Invitrogen/Thermo Fisher Scientific, Waltham, MA) with 2-min of bead-beating (BioSpec Products, Barlesville, OK) at RT in tubes containing 1.0 g of 1 mm diameter zirconium beads (BioSpec Products) and cleaned with the Qiagen RNeasy mini kit (Qiagen, Hilden, Germany). Template cDNA was synthesized using 125 ng of purified RNA in 20 µL reaction volumes. cDNA was diluted 1:1 with water then 1 µL was mixed with SYBR qPCR Mastermix (Bio-Rad, Hercules, CA) and combined with the appropriate primers (400 nM) and water for a total reaction volume of 10 µL. A list of primers is shown in Supplementary Table 3. The cycling protocol was performed using Mastercycler® nexus (Eppendorf, Hamburg, Germany) as follows: 30 s at 95 ˚C, followed by 35 cycles of 10 s at 95 ˚C, 30 s at 60 ˚C. A melt curve was conducted from 65 ˚C to 95 ˚C at

increments of 0.5 ˚C at 5 sec/step. All reactions were run in duplicate and delta-delta-Ct values were calculated relative to the endogenous control (*Gapdh*).

16S rRNA gene sequencing

DNA was extracted from all mouse cecal content and feces as well as human fecal slurries using a phenol-chloroform extraction method that included a bead-beating step (Turnbaugh, Hamady, et al., 2009). 16S rRNA gene (V4) amplification was done by PCR involving unique barcodes (8-bp) both on the forward and reverse primers what are fused to Illumina adapters (Kozich et al., 2013). The V4 amplicons from each sample were combined and submitted for sequencing on an Illumina MiSeq run (2 × 250 bp, Illumina, San Diego, CA) at the University of Wisconsin, Madison Biotechnology Center's DNA Sequencing Facility. Samples with less than 20% of the average sample read count were excluded, resulting in the removal of four samples from further analysis. The remaining samples ranged from 33,869 to 133,538 paired-end reads with an average of 77,704 paired-end reads per sample. Qiime2 (version 2019.10) was used to generate amplicon sequence variant (ASV) tables and taxonomy tables from the 16S rRNA reads. Demultiplexed reads were trimmed and filtered for quality with the Qiime2 DADA2 plug-in (Callahan et al., 2016). ASVs were annotated to the genus level with SILVA reference database (Quast et al., 2012) (version 132) using Naïve Bayes classifier in Qiime2. All subsequent analysis were conducted in R. ASV-level feature counts normalized by converting to relative abundance. ASVs were filtered to include only those with at least 0.0001 average relative abundance across all cecal samples. For genus-level analysis, a cutoff was set at 0.0005 average relative abundance across all samples. These cutoffs were also applied to the fecal sample ASV and genera profiles (human inoculum fecal slurries and pre-dietary treatment mouse feces). Engraftment efficiency was calculated as the number of features that were detected in the pre-treatment mouse feces of at least one animal per donor- or donor/diet-group divided by the number of features detected in the donor fecal slurry. Positive detection was defined as any

feature (ASV or genus) that was found to be greater than 0.0001 relative abundance in a given sample.

Shotgun sequencing

Genomic DNA isolated from the cecal contents of 5 mice per treatment group was used for metagenomic analysis. Libraries were prepared using the Illumina TruSeq PCR-free kit following vendor protocols and sequenced at the University of Wisconsin Biotechnology Center's DNA Sequencing Facility. All samples were run on a single NovaSeq6000 2 × 150 S4-Flowcell lane. The resulting sequences were trimmed for quality using Trimmomatic (version 0–39) and then aligned against reference host genomes (*Mus musculus* GRCm38_Rel98) with bowtie2 (version 2.3.4) to remove host reads (average host alignment rate was 5.3%) leaving only high-quality, non-host reads. Cleaning ultimately resulted in an average of 29.4 million paired-end reads per sample.

Functional annotation

Reads remaining after trimming and removal of host sequences were concatenated into a single fastq file and fed into HUMAnN3 (version 3.0.0.alpha.4) for functional annotation. This resulted in a UniRef90 (Suzek et al., 2015) gene family abundance table in reads per kilobase, and a relative abundance table of microbial taxa for each mouse. The UniRef90 gene family abundance tables were converted to KO counts-per-million (CPM) abundance tables with the human_regroup_table and human_renorm_table functions. Differential abundance analysis was conducted on KOs that were present in at least 25% of samples.

CAZyme annotation

CAZyme profiles of each sample were predicted using run_dbcan (version 2.0.11). We assembled cleaned (trimmed, host-free) reads into contigs with metaSPAdes (version 3.14.0) with

multiple k-mer sizes (metaspades.py -k 21, 33, 55, 77). Contigs shorter than 500 bp were discarded from further processing. Open reading frames (ORFs; i.e., microbial metagenes), were predicted from assembled contigs via Prodigal (version 2.6.3) using Hidden Markov Model (HMM) with default parameters. All predicted genes shorter than 100 bp were discarded from further processing. Nucleotide ORF sequences were converted amino acid sequences and were used as input for run_dbcan (version 2.0.11) to predict CAZyme profiles. CAZyme annotation was accepted if an ORF was annotated by >= 2 tools (DIAMOND, HMMER, Hotpep). This resulted in a table indicating the presence or absence of each CAZyme family in the CAZyme database. To estimate CAZyme abundance, each CAZyme family was assigned a count-per-million (CPM) value of its associated ORF as predicted by Prodigal. If multiple CAZymes were predicted from the same ORF, they were all assigned the ORF's CPM value.

Microbiome analysis

PCoA plots and diversity measures were generated using 16S rRNA ASV profiles with the phyloseq (version 1.40.0) package in R. All pairwise PERMANOVA tests were conducted between dietary groups within each donor group using the pairwiseAdonis (version 0.4) R package with 9999 permutations. Feature-level differential abundance analysis of 16S rRNA amplicon taxonomy, shotgun metagenomic KO abundances and shotgun metagenomic CAZyme abundances were conducted using the MaAsLin 2 function within the MaAsLin 2 (version 1.10.0) R package with default settings. To assess KO pathway enrichment, sets of KOs that were significantly upregulated by FF feeding (MaAsLin 2 differential abundance, adjusted $P < 0.1$) were generated for each donor group and used as input for the PerformKOEnrichAnalysis_KO01100 function within the MicrobiomeAnalyistR (version 0.0.0.9000) package in R with default parameters. This resulted in a list of KEGG pathways that were significantly ($P < 0.05$) overrepresented in FF-fed mice within each donor group. CAZymes families were considered to

be highly differential if they belonged to the top 10% of CAZymes ordered by MaAsLin 2 effect size within each donor group, regardless of direction (absolute value of effect size).

Statistical analysis

Unless otherwise noted, comparisons of means were conducted using a two-sided Wilcoxon rank-sum test between dietary groups within each donor group. Equal variance was determined for all means tests (Levene's test, $P > 0.05$) except comparisons of atherosclerotic plaques (size, lipid content, macrophage infiltration) and cecal SCFA levels which were found to have significantly different variances between diets (Levene's test, $P < 0.05$). Correlations between CAZymes and cecal SCFA levels were conducted using all mice to calculate Spearman's rank correlation coefficient. *P*-value adjustment for PERMANOVA was done using the Bonferroni method, while all other adjusted *P*-values were calculated using the Benjamini–Hochberg method.

Data availability

Sequencing data reported in this study is available at the European Nucleotide Archive (ENA) under the study accession number PRJEB58699.

B.6 Contributions

E.R.H., T.-W.L.C., and F.E.R. conceived of the study. E.R.H., K.K., and E.I.V. performed the mouse experiments and collected tissues. E.R.H. conducted analysis of atherosclerosis phenotype. E.R.H. and Q.Z. conducted gut metagenomic analysis. E.R.H. and F.E.R. prepared the manuscript. This manuscript was approved by all authors.

B.7 Acknowledgements

The authors thank Dr. Barbara Mickelson (Envigo) for assistance with diets. We also thank the University of Wisconsin–Madison Biotechnology Center's DNA Sequencing Facility for providing sequencing and support services. We thank the University of Wisconsin–Madison's Center for High Throughput Computing (CHTC) in the Department of Computer Sciences for providing computational resources, support, and assistance. This work was partly supported by grants from NIH HL144651 (F.E.R.), HL148577 (F.E.R.), and EB030340 (F.E.R.). This work was also supported by a grant from a Transatlantic Networks of Excellence Award from the Leducq Foundation (17CVD01). ERH was supported in part by the Metabolism and Nutrition Training Program NIH T32 (DK007665) and by the University of Wisconsin–Madison Food Research Institute (Robert H. and Carol L. Deibel Distinguished Graduate Fellowship in Probiotic Research). T.-W.L.C. was supported by the National Institutes of Health, under Ruth L. Kirschstein National Research Service Award T32 HL007936 from the National Heart Lung and Blood Institute to the University of Wisconsin–Madison Cardiovascular Research Center.

B.8 References

Aguilar, E. C., Leonel, A. J., Teixeira, L. G., Silva, A. R., Silva, J. F., Pelaez, J. M. N., Capettini, L. S. A., Lemos, V. S., Santos, R. A. S., & Alvarez-Leite, J. I. (2014). Butyrate impairs atherogenesis by reducing plaque inflammation and vulnerability and decreasing NFκB activation. *Nutrition, Metabolism and Cardiovascular Diseases*, *24*(6), 606–613. https://doi.org/10.1016/j.numecd.2014.01.002

Ahmad, F. B., Cisewski, J. A., Miniño, A., & Anderson, R. N. (2021). *Provisional Mortality Data—United States, 2020*. *70*(14).

Bokoliya, S. C., Dorsett, Y., Panier, H., & Zhou, Y. (2021). Procedures for Fecal Microbiota Transplantation in Murine Microbiome Studies. *Frontiers in Cellular and Infection Microbiology*, *11*, 711055. https://doi.org/10.3389/fcimb.2021.711055

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Cankar, K., Kortstee, A., Toonen, M. A. J., Wolters-Arts, M., Houbein, R., Mariani, C., Ulvskov, P., Jorgensen, B., Schols, H. A., Visser, R. G. F., & Trindade, L. M. (2014). Pectic arabinan side chains are essential for pollen cell wall integrity during pollen development. *Plant Biotechnology Journal*, *12*(4), 492–502. https://doi.org/10.1111/pbi.12156

Chong, J., Liu, P., Zhou, G., & Xia, J. (2020). Using MicrobiomeAnalyst for comprehensive statistical, functional, and meta-analysis of microbiome data. *Nature Protocols*, *15*(3), 799–821. https://doi.org/10.1038/s41596-019-0264-1

Chung, H., Pamp, S. J., Hill, J. A., Surana, N. K., Edelman, S. M., Troy, E. B., Reading, N. C., Villablanca, E. J., Wang, S., Mora, J. R., Umesaki, Y., Mathis, D., Benoist, C., Relman, D. A., & Kasper, D. L. (2012). Gut Immune Maturation Depends on Colonization with a

Host-Specific Microbiota. *Cell*, *149*(7), 1578–1593. https://doi.org/10.1016/j.cell.2012.04.037

Contijoch, E. J., Britton, G. J., Yang, C., Mogno, I., Li, Z., Ng, R., Llewellyn, S. R., Hira, S., Johnson, C., Rabinowitz, K. M., Barkan, R., Dotan, I., Hirten, R. P., Fu, S.-C., Luo, Y., Yang, N., Luong, T., Labrias, P. R., Lira, S., … Faith, J. J. (2019). Gut microbiota density influences host physiology and is shaped by host and microbial factors. *eLife*, 26.

Cummings, J. H. (1984). Cellulose and the human gut. *Gut*, *25*(8), 805–810. https://doi.org/10.1136/gut.25.8.805

Deehan, E. C., Yang, C., Perez-Muñoz, M. E., Nguyen, N. K., Cheng, C. C., Triador, L., Zhang, Z., Bakal, J. A., & Walter, J. (2020). Precision Microbiome Modulation with Discrete Dietary Fiber Structures Directs Short-Chain Fatty Acid Production. *Cell Host & Microbe*, *27*(3), 389-404.e6. https://doi.org/10.1016/j.chom.2020.01.006

Denson, L. A., Curran, M., McGovern, D. P. B., Koltun, W. A., Duerr, R. H., Kim, S. C., Sartor, R. B., Sylvester, F. A., Abraham, C., De Zoeten, E. F., Siegel, C. A., Burns, R. M., Dobes, A. M., Shtraizent, N., Honig, G., Heller, C. A., Hurtado-Lorenzo, A., & Cho, J. H. (2019). Challenges in IBD Research: Precision Medicine. *Inflammatory Bowel Diseases*, *25*(Supplement_2), S31–S39. https://doi.org/10.1093/ibd/izz078

Earley, H., Lennon, G., Balfe, Á., Coffey, J. C., Winter, D. C., & O'Connell, P. R. (2019). The abundance of Akkermansia muciniphila and its relationship with sulphated colonic mucins in health and ulcerative colitis. *Scientific Reports*, *9*(1), 15683. https://doi.org/10.1038/s41598-019-51878-3

Goodman, A. L., Kallstrom, G., Faith, J. J., Reyes, A., Moore, A., Dantas, G., & Gordon, J. I. (2011). Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proceedings of the National Academy of Sciences*, *108*(15), 6252–6257. https://doi.org/10.1073/pnas.1102938108

Haghikia, A., Zimmermann, F., Schumann, P., Jasina, A., Roessler, J., Schmidt, D., Heinze, P., Kaisler, J., Nageswaran, V., Aigner, A., Ceglarek, U., Cineus, R., Hegazy, A. N., Van Der Vorst, E. P. C., Döring, Y., Strauch, C. M., Nemet, I., Tremaroli, V., Dwibedi, C., … Landmesser, U. (2022). Propionate attenuates atherosclerosis by immune-dependent regulation of intestinal cholesterol metabolism. *European Heart Journal*, *43*(6), 518–533. https://doi.org/10.1093/eurheartj/ehab644

Hansen, C. H. F., Nielsen, D. S., Kverka, M., Zakostelska, Z., Klimesova, K., Hudcovic, T., Tlaskalova-Hogenova, H., & Hansen, A. K. (2012). Patterns of Early Gut Colonization Shape Future Immune Responses of the Host. *PLoS ONE*, *7*(3), e34043. https://doi.org/10.1371/journal.pone.0034043

Healey, G. R., Murphy, R., Brough, L., Butts, C. A., & Coad, J. (2017). Interindividual variability in gut microbiota and host response to dietary interventions. *Nutrition Reviews*, *75*(12), 1059–1080. https://doi.org/10.1093/nutrit/nux062

Herd, P., Carr, D., & Roan, C. (2014). Cohort Profile: Wisconsin longitudinal study (WLS). *International Journal of Epidemiology*, *43*(1), 34–41. https://doi.org/10.1093/ije/dys194

Herd, P., Schaeffer, N. C., DiLoreto, K., Jacques, K., Stevenson, J., Rey, F., & Roan, C. (2018). The Influence of Social Conditions Across the Life Course on the Human Gut Microbiota: A Pilot Project With the Wisconsin Longitudinal Study. *The Journals of Gerontology: Series B*, *73*(1), 124–133. https://doi.org/10.1093/geronb/gbx029

Higginson, J., & Pepler, W. J. (1954). Fat Intake, Serum Cholesterol Concentration, and Atherosclerosis in the South African Bantu. Part II. Atherosclerosis and Coronary Artery Disease. *Journal of Clinical Investigation*, *33*(10), 1366–1371. https://doi.org/10.1172/JCI103013

Hofmann, M. A., Lalla, E., Lu, Y., Gleason, M. R., Wolf, B. M., Tanji, N., Ferran, L. J., Kohl, B., Rao, V., Kisiel, W., Stern, D. M., & Schmidt, A. M. (2001). Hyperhomocysteinemia

enhances vascular inflammation and accelerates atherosclerosis in a murine model. *Journal of Clinical Investigation*, *107*(6), 675–683. https://doi.org/10.1172/JCI10588

Hughes, R. L., Marco, M. L., Hughes, J. P., Keim, N. L., & Kable, M. E. (2019). The Role of the Gut Microbiome in Predicting Response to Diet and the Development of Precision Nutrition Models—Part I: Overview of Current Methods. *Advances in Nutrition*, *10*(6), 953–978. https://doi.org/10.1093/advances/nmz022

Javdan, B., Lopez, J. G., Chankhamjon, P., Lee, Y.-C. J., Hull, R., Wu, Q., Wang, X., Chatterjee, S., & Donia, M. S. (2020). Personalized Mapping of Drug Metabolism by the Human Gut Microbiome. *Cell*, *181*(7), 1661-1679.e22. https://doi.org/10.1016/j.cell.2020.05.001

Jie, Z., Xia, H., Zhong, S.-L., Feng, Q., Li, S., Liang, S., Zhong, H., Liu, Z., Gao, Y., Zhao, H., Zhang, D., Su, Z., Fang, Z., Lan, Z., Li, J., Xiao, L., Li, J., Li, R., Li, X., … Kristiansen, K. (2017). The gut microbiome in atherosclerotic cardiovascular disease. *Nature Communications*, *8*(1), 845. https://doi.org/10.1038/s41467-017-00900-1

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., & Rey, F. E. (2018). Interactions between Roseburia intestinalis and diet modulate atherogenesis in a murine model. *Nature Microbiology*, *3*(12), 1461–1471. https://doi.org/10.1038/s41564-018-0272-x

Kelly, T. N., Bazzano, L. A., Ajami, N. J., He, H., Zhao, J., Petrosino, J. F., Correa, A., & He, J. (2016). Gut Microbiome Associates With Lifetime Cardiovascular Disease Risk Profile Among Bogalusa Heart Study Participants. *Circulation Research*, *119*(8), 956–964. https://doi.org/10.1161/CIRCRESAHA.116.309219

Kim, Y., Hwang, S. W., Kim, S., Lee, Y.-S., Kim, T.-Y., Lee, S.-H., Kim, S. J., Yoo, H. J., Kim, E. N., & Kweon, M.-N. (2020). Dietary cellulose prevents gut inflammation by modulating lipid metabolism and gut microbiota. *Gut Microbes*, *11*(4), 944–961. https://doi.org/10.1080/19490976.2020.1730149

Kirk, D., Catal, C., & Tekinerdogan, B. (2021). Precision nutrition: A systematic literature review. *Computers in Biology and Medicine*, *133*, 104365. https://doi.org/10.1016/j.compbiomed.2021.104365

Kolodziejczyk, A. A., Zheng, D., & Elinav, E. (2019). Diet–microbiota interactions and personalized nutrition. *Nature Reviews Microbiology*, *17*(12), 742–753. https://doi.org/10.1038/s41579-019-0256-8

Koppel, N., Bisanz, J. E., Pandelia, M.-E., Turnbaugh, P. J., & Balskus, E. P. (2018). Discovery and characterization of a prevalent human gut bacterial enzyme sufficient for the inactivation of a family of plant toxins. *eLife*, *7*, e33953. https://doi.org/10.7554/eLife.33953

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*(17), 5112–5120. https://doi.org/10.1128/AEM.01043-13

Kritchevsky, D. (1978). Fiber, lipids, and atherosclerosis. *The American Journal of Clinical Nutrition*, *31*(10), 65S-74S. https://doi.org/10.1093/ajcn/31.10.S65

Lee, J. R., Muthukumar, T., Dadhania, D., Taur, Y., Jenq, R. R., Toussaint, N. C., Ling, L., Pamer, E., & Suthanthiran, M. (2015). Gut Microbiota and Tacrolimus Dosing in Kidney Transplantation. *PLOS ONE*, *10*(3), e0122399. https://doi.org/10.1371/journal.pone.0122399

Leshem, A., Segal, E., & Elinav, E. (2020). The Gut Microbiome and Individual-Specific Responses to Diet. *mSystems*, *5*(5), e00665-20. https://doi.org/10.1128/mSystems.00665-20

Libby, P. (2012). Inflammation in Atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *32*(9), 2045–2051. https://doi.org/10.1161/ATVBAHA.108.179705

Mallick, H., Rahnavard, A., McIver, L. J., Ma, S., Zhang, Y., Nguyen, L. H., Tickle, T. L., Weingart, G., Ren, B., Schwager, E. H., Chatterjee, S., Thompson, K. N., Wilkinson, J. E., Subramanian, A., Lu, Y., Waldron, L., Paulson, J. N., Franzosa, E. A., Bravo, H. C., & Huttenhower, C. (2021). Multivariable association discovery in population-scale meta-omics studies. *PLOS Computational Biology*, *17*(11), e1009442. https://doi.org/10.1371/journal.pcbi.1009442

McOrist, A. L., Miller, R. B., Bird, A. R., Keogh, J. B., Noakes, M., Topping, D. L., & Conlon, M. A. (2011). Fecal Butyrate Levels Vary Widely among Individuals but Are Usually Increased by a Diet High in Resistant Starch1,2. *The Journal of Nutrition*, *141*(5), 883–889. https://doi.org/10.3945/jn.110.128504

Murga-Garrido, S. M., Hong, Q., Cross, T.-W. L., Hutchison, E. R., Han, J., Thomas, S. P., Vivas, E. I., Denu, J., Ceschin, D. G., Tang, Z.-Z., & Rey, F. E. (2021). Gut microbiome variation modulates the effects of dietary fiber on host metabolism. *Microbiome*, *9*(1), 117. https://doi.org/10.1186/s40168-021-01061-6

Ohira, H., Tsutsui, W., & Fujioka, Y. (2017). Are Short Chain Fatty Acids in Gut Microbiota Defensive Players for Inflammation and Atherosclerosis? *Journal of Atherosclerosis and Thrombosis*, *24*(7), 660–672. https://doi.org/10.5551/jat.RV17006

Ordovas, J. M., & Berciano, S. (2020). Personalized nutrition and healthy aging. *Nutrition Reviews*, *78*(Supplement_3), 58–65. https://doi.org/10.1093/nutrit/nuaa102

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596. https://doi.org/10.1093/nar/gks1219

Rault-Nania, M.-H., Gueux, E., Demougeot, C., Demigné, C., Rock, E., & Mazur, A. (2006). Inulin attenuates atherosclerosis in apolipoprotein E-deficient mice. *British Journal of Nutrition*, *96*(5), 840–844. https://doi.org/10.1017/BJN20061913

Russell, R. (1999). Atherosclerosis—An Inflammatory Disease. *The New England Journal of Medicine*.

Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B., Wu, C. H., & the UniProt Consortium. (2015). UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*, *31*(6), 926–932. https://doi.org/10.1093/bioinformatics/btu739

Tang, W. H. W., & Hazen, S. L. (2017). The Gut Microbiome and Its Role in Cardiovascular Diseases. *Circulation*, *135*(11), 1008–1010. https://doi.org/10.1161/CIRCULATIONAHA.116.024251

Threapleton, D. E., Greenwood, D. C., Evans, C. E. L., Cleghorn, C. L., Nykjaer, C., Woodhead, C., Cade, J. E., Gale, C. P., & Burley, V. J. (2013). Dietary fibre intake and risk of cardiovascular disease: Systematic review and meta-analysis. *BMJ*, *347*(dec19 2), f6879–f6879. https://doi.org/10.1136/bmj.f6879

Torres, N., Guevara-Cruz, M., Velázquez-Villegas, L. A., & Tovar, A. R. (2015). Nutrition and Atherosclerosis. *Archives of Medical Research*, *46*(5), 408–426. https://doi.org/10.1016/j.arcmed.2015.05.010

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A., Affourtit, J. P., Egholm, M., Henrissat, B., Heath, A. C., Knight, R., & Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature*, *457*(7228), 480–484. https://doi.org/10.1038/nature07540

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The Human Microbiome Project. *Nature*, *449*(7164), 804–810. https://doi.org/10.1038/nature06244

Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R., & Gordon, J. I. (2009). The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized

Gnotobiotic Mice. *Science Translational Medicine*, *1*(6). https://doi.org/10.1126/scitranslmed.3000322

Vital, M., Howe, A. C., & Tiedje, J. M. (2014). Revealing the Bacterial Butyrate Synthesis Pathways by Analyzing (Meta)genomic Data. *mBio*, *5*(2), e00889-14. https://doi.org/10.1128/mBio.00889-14

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*(7341), 57–63. https://doi.org/10.1038/nature09922

Weeke, P., & Roden, D. M. (2013). Pharmacogenomics and Cardiovascular Disease. *Current Cardiology Reports*, *15*(7), 376. https://doi.org/10.1007/s11886-013-0376-0

Xue, H., Chen, X., Yu, C., Deng, Y., Zhang, Y., Chen, S., Chen, X., Chen, K., Yang, Y., & Ling, W. (2022). Gut Microbially Produced Indole-3-Propionic Acid Inhibits Atherosclerosis by Promoting Reverse Cholesterol Transport and Its Deficiency Is Causally Related to Atherosclerotic Cardiovascular Disease. *Circulation Research*, *131*(5), 404–420. https://doi.org/10.1161/CIRCRESAHA.122.321253

Zeevi, D., Korem, T., Zmora, N., Israeli, D., Rothschild, D., Weinberger, A., Ben-Yacov, O., Lador, D., Avnit-Sagi, T., Lotan-Pompan, M., Suez, J., Mahdi, J. A., Matot, E., Malka, G., Kosower, N., Rein, M., Zilberman-Schapira, G., Dohnalová, L., Pevsner-Fischer, M., … Segal, E. (2015). Personalized Nutrition by Prediction of Glycemic Responses. *Cell*, *163*(5), 1079–1094. https://doi.org/10.1016/j.cell.2015.11.001

Zhang, Y., Hu, J., Tan, H., Zhong, Y., & Nie, S. (2022). Akkermansia muciniphila, an important link between dietary fiber and host health. *Current Opinion in Food Science*, *47*, 100905. https://doi.org/10.1016/j.cofs.2022.100905

Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., & Goodman, A. L. (2019a). Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature*, *570*(7762), 462–467. https://doi.org/10.1038/s41586-019-1291-3

Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., & Goodman, A. L. (2019b). Separating host and microbiome contributions to drug pharmacokinetics and toxicity. *Science*, *363*(6427), eaat9931. https://doi.org/10.1126/science.aat9931

B.9 Figures

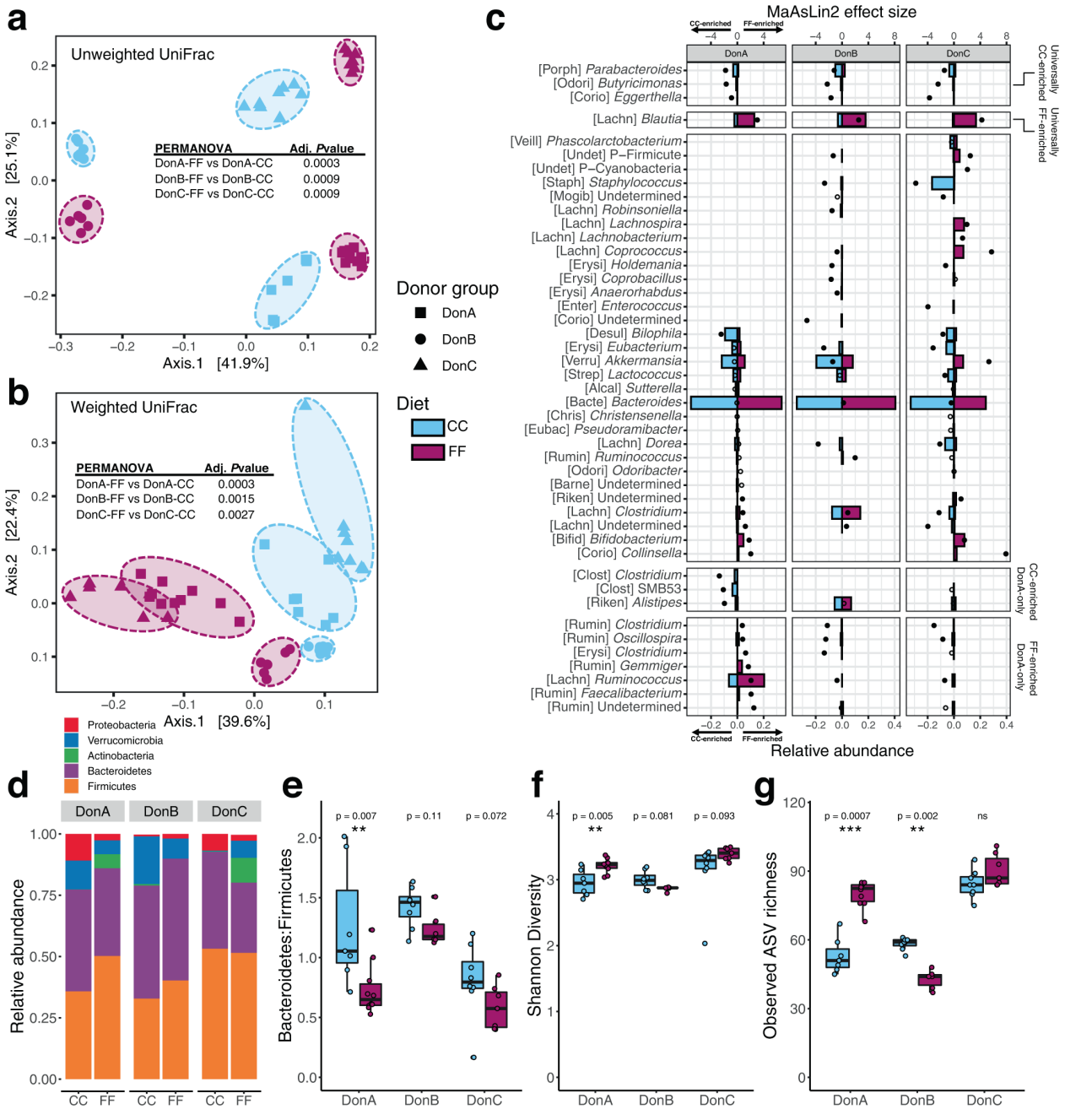


Figure B.1 Effects of dietary fiber on microbial community structure in gnotobiotic mice colonized with different human communities. **a**, **b** Principal Coordinate Analysis of unweighted and weighted UniFrac distances of 16S rRNA V4 ASVs. Animals colonized with fecal samples from three different donors: DonA, DonB, and DonC. **c** Relative abundances of genus-

level taxa for the two diets (bottom *x*-axis, colored bars) along with between-diet differential abundance (MaAsLin 2) effect sizes (top *x*-axis, dots). Significant differences in abundance for taxa within each donor group are denoted by a solid dot (adjusted $P < 0.1$) and open circles denote no significant change (adjusted $P > 0.1$). The dot's orientation relative to the origin represents the effect of diet on the abundance of each taxa (negative values correspond to CC abundances, positive values correspond to FF abundances). The first 5 letters of the family encompassing each taxon is shown in brackets; if the family is undetermined the taxon phylum is listed instead and noted with a "P-". **d** Relative abundance of Phylum-level taxa as a function of diet and donor group. **e** Bacteroidetes to Firmicutes ratio. **f**, **g** Shannon diversity index and observed richness. Box and whisker plots denote the interquartile range, median, and spread of points within 1.5 times the interquartile range along with individual data points; magenta = Fermentable Fiber (FF), blue = Cellulose Control (CC). Comparisons of means ($n = 7–10$/diet/donor group) conducted with Wilcoxon test, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

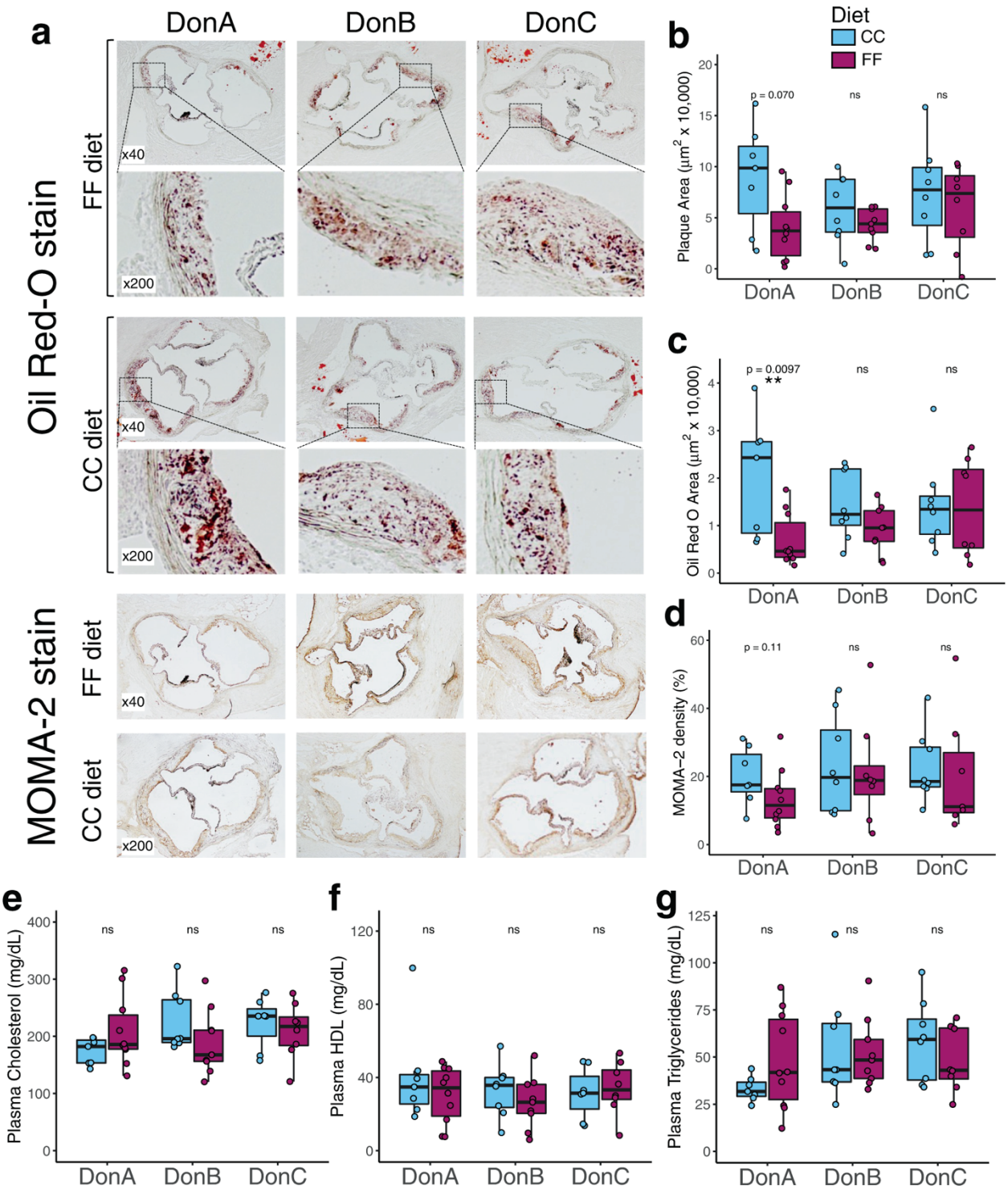


Figure B.2 Atherosclerosis response to dietary fiber in mice colonized with different human communities. Atherosclerosis was measured in GF *ApoE*$^{-/-}$ mice colonized with human

fecal communities DonA, DonB, and DonC and fed either fermentable fiber diet, or a cellulose control diet. **a** Representative Oil Red O staining and MOMA-2 antibody staining of aortic sinus cross-sections content. Quantification of plaque average area (**b**), lipid positive area (**c**) or MOMA-2 positive area (**d**). Plasma levels of total cholesterol (**e**), HDL-cholesterol (**f**) and triglycerides (**g**). Box and whisker plots denote the interquartile range, median, and spread of points within 1.5 times the interquartile range along with individual data points; magenta = Fermentable Fiber (FF), blue = Cellulose Control (CC). Comparisons of means between diets within each donor group (n = 7–10/diet/donor group) were conducted using a Wilcoxon test with appropriate correction for equal variance assumption (Levenes' test), *$P < 0.05$, **$P < 0.01$.

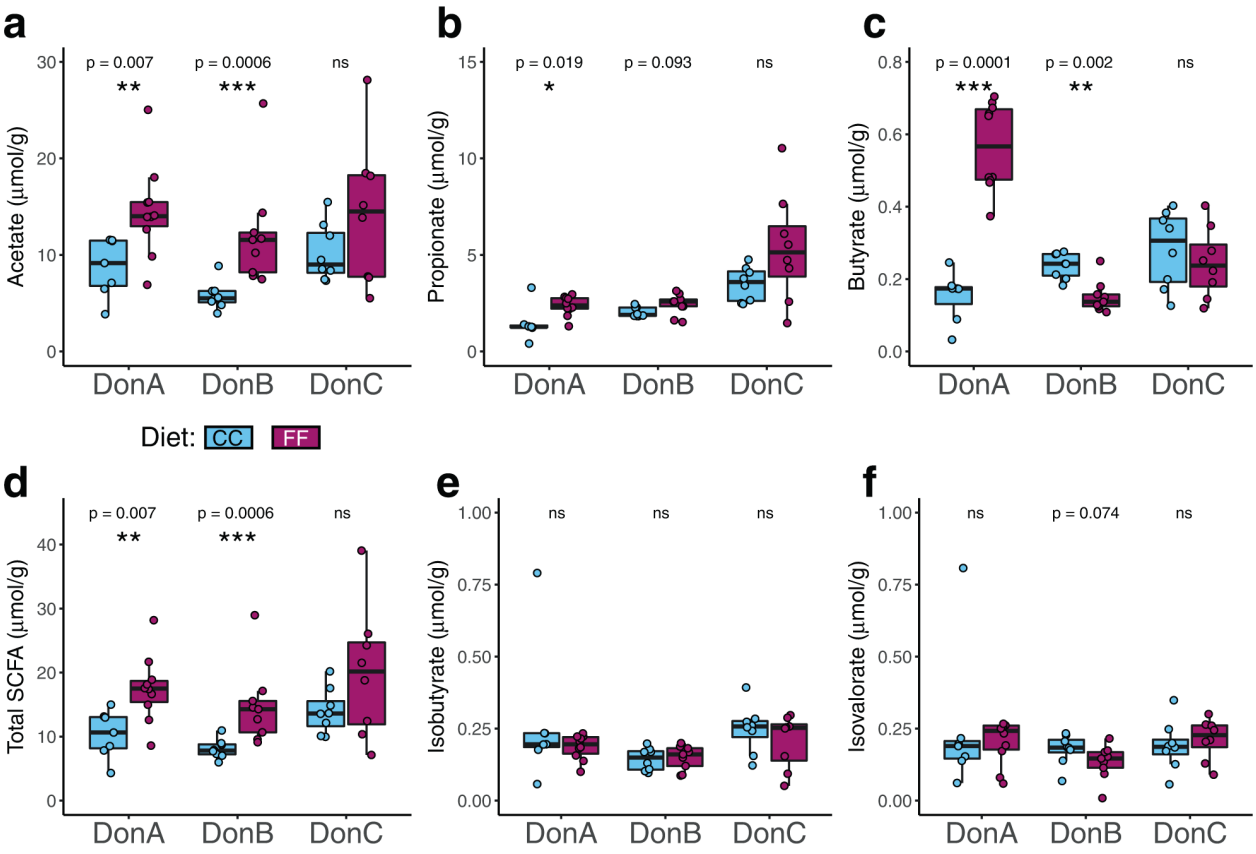


Figure B.3 Effect of dietary fiber on cecal levels of SCFAs and branched-chain fatty acids. Cecal levels of acetate (**a**), propionate (**b**), butyrate (**c**), total SCFAs (sum of acetate, propionate, and butyrate) (**d**), isobutyrate (**e**) and isovalerate (**f**). Concentrations are expressed per gram of cecal content wet weight. Box and whisker plots denote the interquartile range, median, and spread of points within 1.5 times the interquartile range along with individual data points. Comparisons of means between diets within each donor group ($n = 7–10$/diet/donor group) were conducted using a Wilcoxon test, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

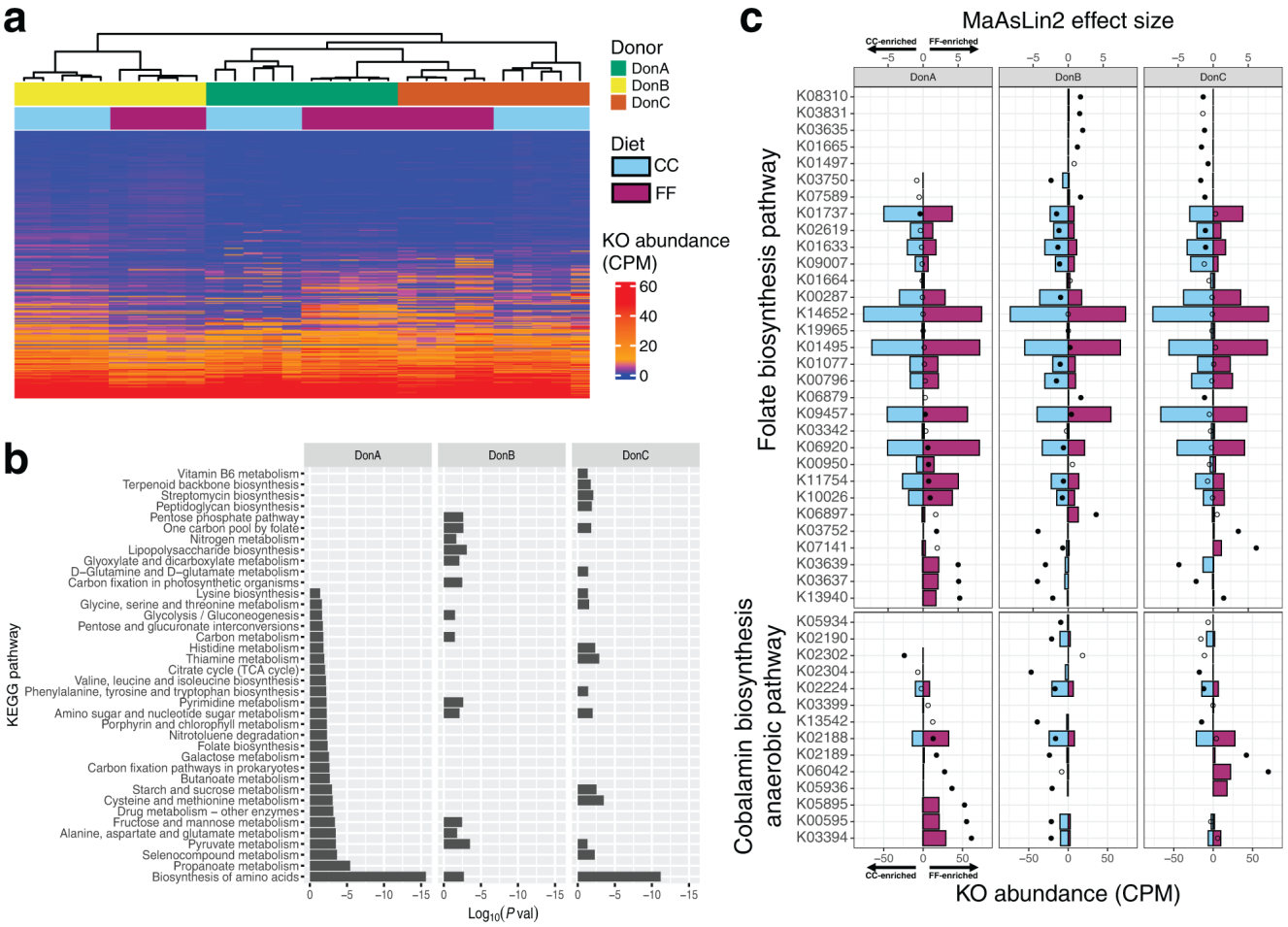


Figure B.4 Effect of dietary fiber on KEGG orthology profiles and metagenomic pathway enrichment across donor groups. a Heatmap of KO profiles expressed in CPM for each mouse. Individual profiles were clustered using the spearman hierarchical clustering method. **b** Overrepresented KEGG pathways and their significance (represented as the $\log_{10}$ of the enrichment *P*-value) identified by pathway enrichment analysis using lists of KOs from each donor group that were significantly upregulated (MaAsLin 2 differential abundance, adjusted $P < 0.1$). **c** MaAsLin 2 differential abundance (bottom axis, colored bars) and effect size (top axis, solid dot = adjusted $P < 0.1$, open dot = adjusted $P > 0.1$) of KOs involved in folate biosynthesis and cobalamin (vitamin B12) biosynthesis. Negative values reflect KO abundance (CPM) in CC-fed mice and effect sizes (MaAsLin 2 coefficient) favoring the CC condition, while

positive values indicate KOs abundances and effect sizes in the FF condition ($n = 5$/diet/donor group).

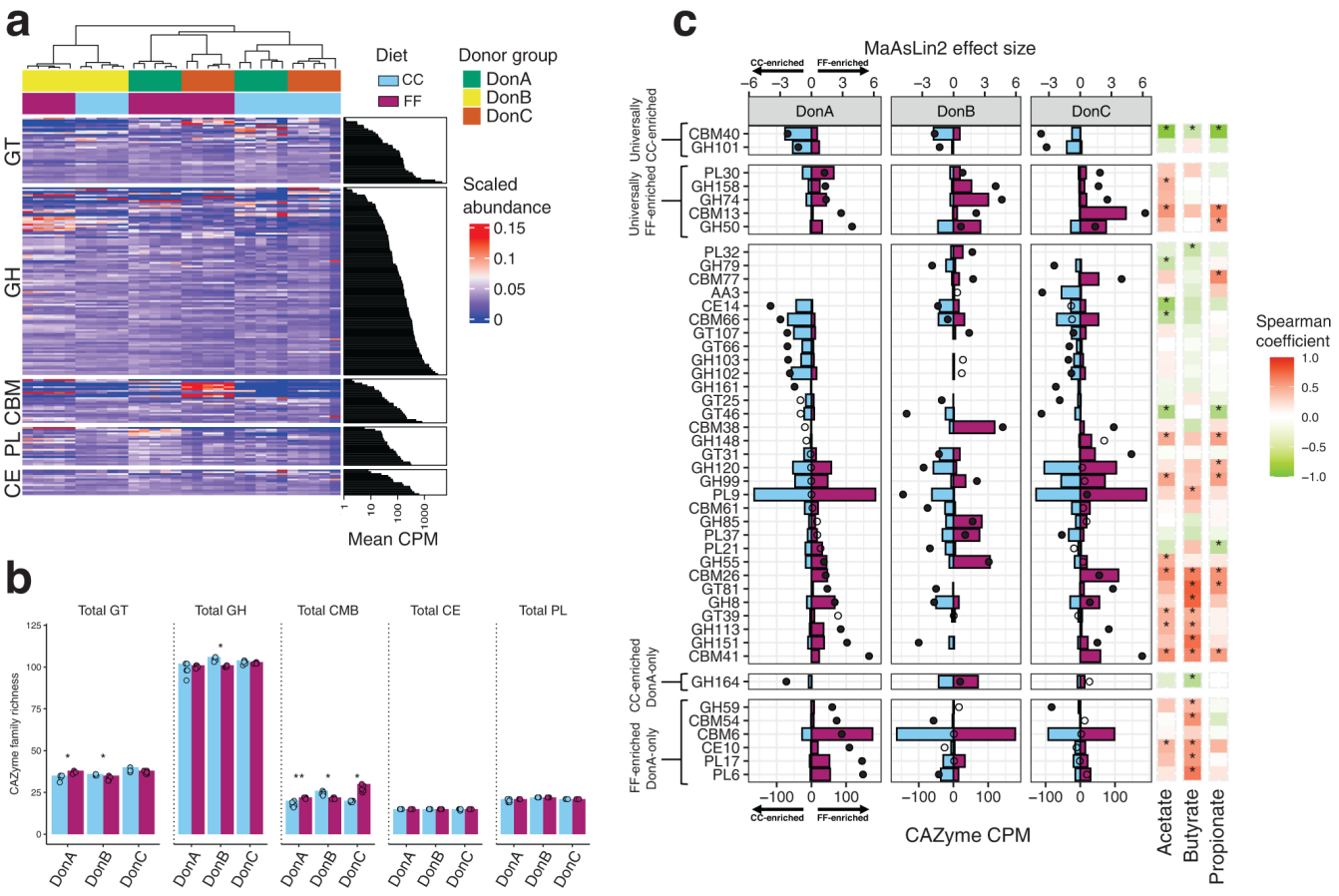
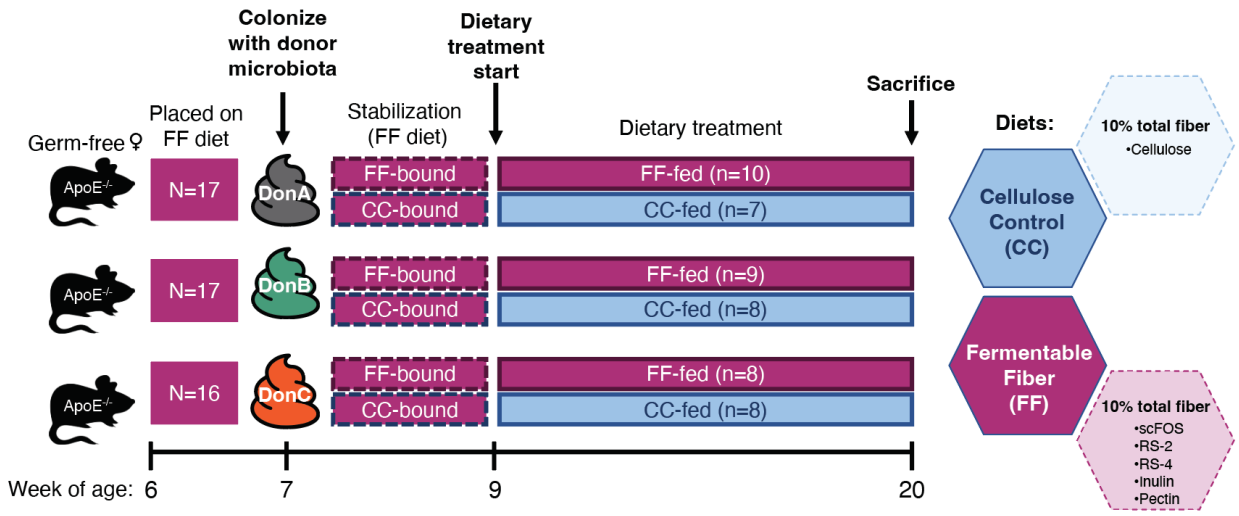
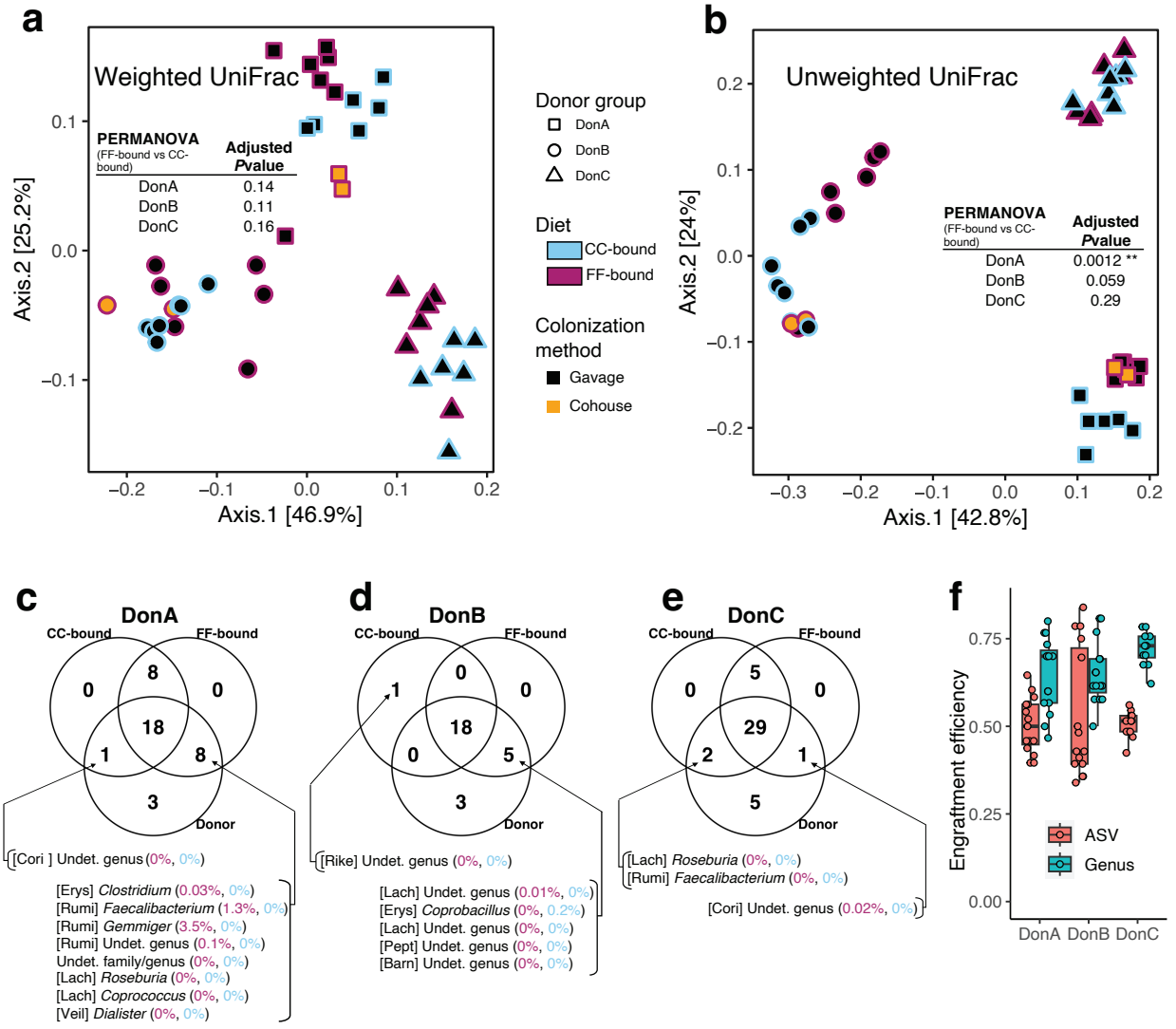


Figure B.5 Effect of dietary fermentable fiber on microbial Carbohydrate Active EnZyme (CAZyme) profiles. a Heatmap of CAZyme family profiles arranged by category (Glycoside hydrolase = GH; glycosyltransferases = GT; carbohydrate binding modules = CBM; carbohydrate esters = CE; and polysaccharide lyases = PL) along with the mean counts for each CAZyme family expressed in counts-per-million (CPM). Mice profiles were clustered using the Spearman hierarchical clustering method. **b** Comparison of CAZyme family richness between diets (total number of CAZyme families detected within each CAZyme category) by donor group. Barplots denote the mean with individual data points; comparisons of means between dietary groups were conducted using the Wilcoxon test; $*P < 0.05$, $**P < 0.01$. **c** MaAsLin 2 differential abundance (bottom axis, colored bars) and effect size (top axis, solid dot = adjusted $P < 0.1$, open dot = adjusted $P > 0.1$) of the top 10% most differentially abundant CAZymes ($n = 5$/diet/donor

group). The right panel depicts a heatmap of Spearman correlation coefficients between each corresponding CAZyme family and cecal SCFA levels across all mice, $*P < 0.05$.

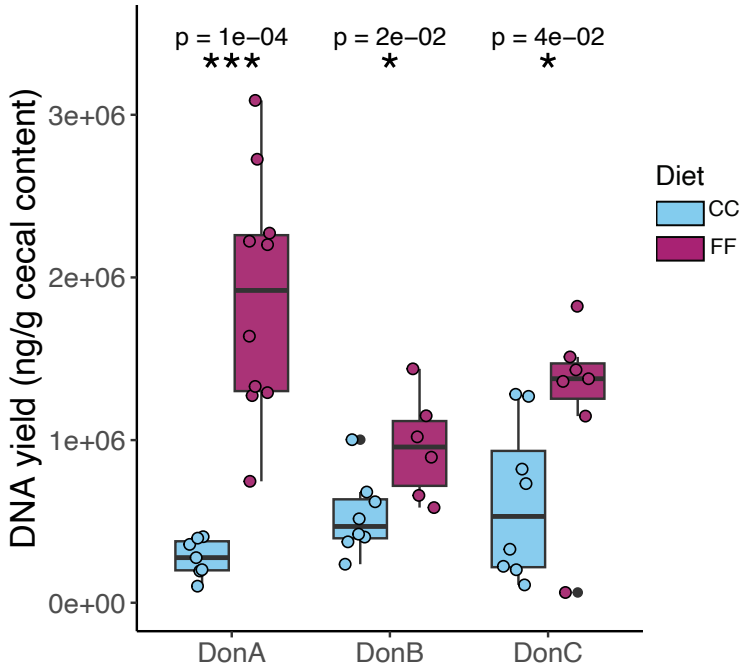


Supplementary Figure B.1 Schematic of the experimental design used in this study.

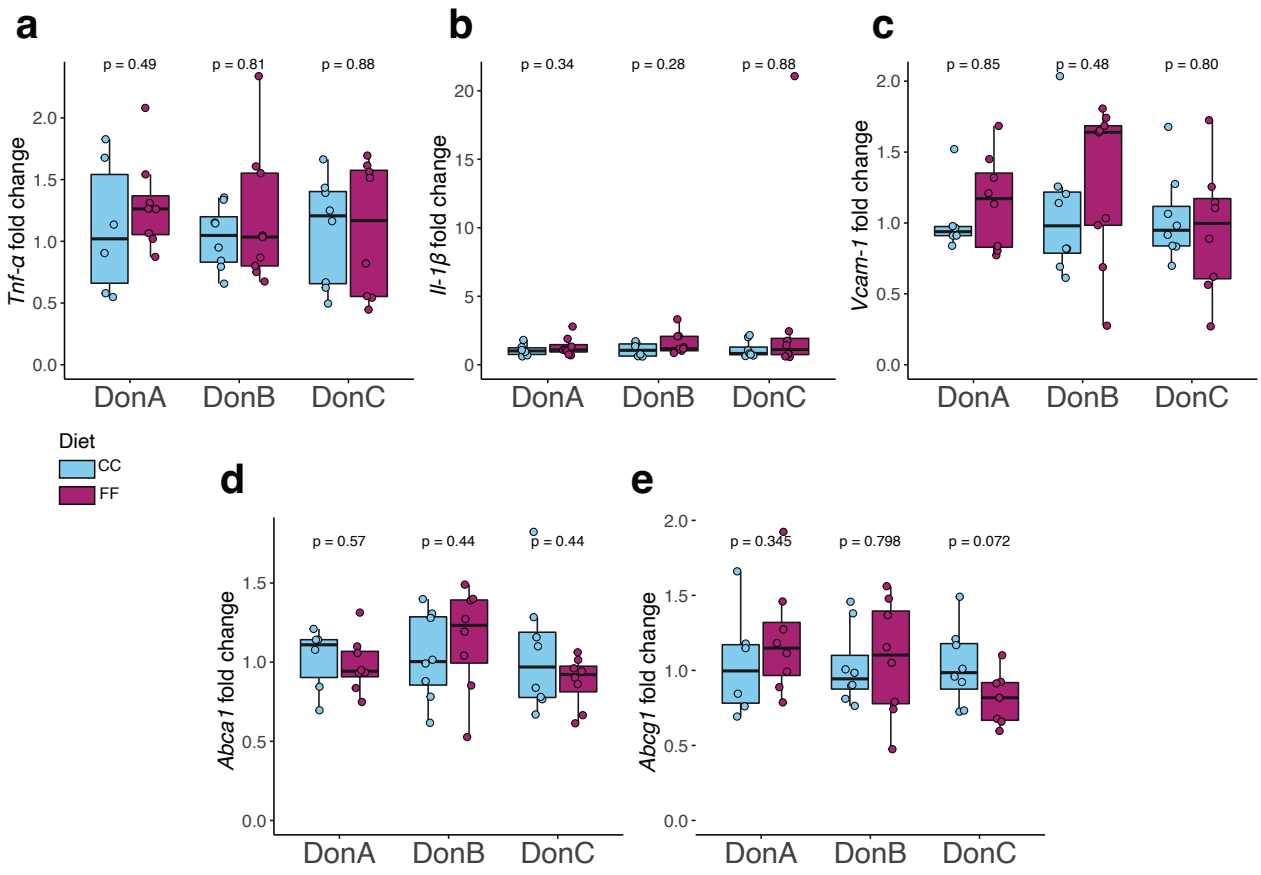


Supplementary Figure B.2 Engraftment of genera from each donor fecal sample into recipient mice prior to dietary treatment. Principal coordinate analysis (PCoA) of pre-treatment fecal samples using weighted (**a**) and unweighted (**b**) UniFrac distances. Colonization method is denoted by the inner color of each point (black = gavage, orange = cohouse) and assigned diet group is denoted by the outer color (FF- bound = magenta, CC-bound = blue). **c-e** Venn diagrams of the genera detected in the donor fecal sample and in at least one mouse belonging to the CC-bound group or the FF-bound group of mice two weeks after colonization and prior to beginning the dietary treatment phase. The genera listed below the Venn diagrams are denoted by their

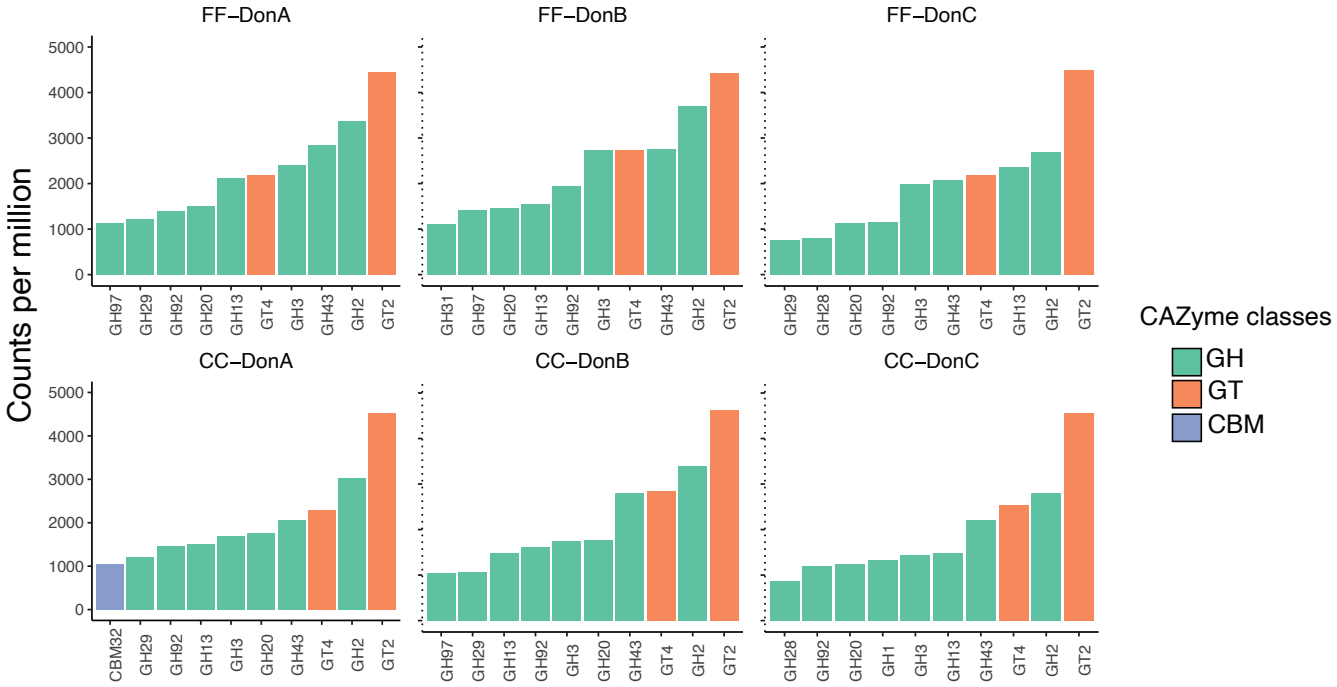
family (in brackets) and their group-average abundance in the cecal content of mice belong to the CC-fed group (blue text) or the FF-fed group (magenta text) at the end of the study after dietary treatment. Undet. = undetermined. **f** Engraftment efficiency for each mouse is expressed at the of genus level (teal) or at the ASVs level (salmon) two weeks after inoculation, but prior to dietary treatment.



Supplementary Figure B.3 Fecal DNA yields per gram of cecal content obtained from mice colonized with human fecal communities DonA, DonB, and DonC and fed either fermentable fiber diet (FF, magenta), or a cellulose control diet (CC, blue). Box and whisker plots denote the interquartile range, median, and spread of points within 1.5 times the interquartile range along with individual data points. Comparisons of means between diets within each donor group (*n* = 7-10/diet/donor group) were conducted using a Wilcoxon test, * *P* <0.05, *** *P* <0.001.



Supplementary Figure B.4 Quantitative RT-PCR measurements of inflammatory markers and cholesterol efflux genes in aortic samples. **a-c** Log2 fold-changes of mRNA abundance for inflammatory markers, and **d,e** genes encoding subunits of key cholesterol transporters between mice consuming FF and CC diets. All fold-changes are expressed relative to the donor-matched CC-fed group and were calculated using *Gapdh* as a reference gene. Box and whisker plots denote the interquartile range, median, and spread of points within 1.5 times the interquartile range along with individual data points. Comparisons of means between diets within each donor group were conducted using a Wilcoxon test.



Supplementary Figure B.5 Average (mean) CPM of the top 10 most abundant CAZyme families within each diet-donor group (FF top, CC bottom). CAZyme classes are indicated by color (Glycoside hydrolase = GH; glycosyltransferases = GT; carbohydrate binding modules = CBM). Values are expressed in counts per million (CPM).