Genetic and environmental control of host-gut microbiota interactions

Elin Org, Brian W W Parks, Jong Wha J Joo, et al.

Genome Res. published online August 10, 2015
Access the most recent version at doi:10.1101/gr.194118.115

Genetic and environmental control of host-gut microbiota interactions

Elin Org[1], Brian W. Parks[1], Jong Wha J Joo[2], Benjamin Emert[1], William Schwartzman[1], Eun Yong Kang[3], Margarete Mehrabian[1], Calvin Pan[4], Rob Knight[5], Robert Gunsalus[6], Thomas A. Drake[7], Eleazar Eskin[3,4] and Aldons J. Lusis[1,4,8*]

[1]Department of Medicine/Division of Cardiology, David Geffen School of Medicine, University of California; Los Angeles, CA, 90095, USA

[2]Bioinformatics IDP, University of California

[3]Department of Computer Science, University of California

[4]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, 90095, USA

[5]Departments of Pediatrics and Computer Science and Engineering, University of California, San Diego, CA,92093, USA

[6]Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, CA, 90095, USA

[7]Department of Pathology and Laboratory Medicine, University of California, Los Angeles, CA, 90095, USA

[8]Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, CA, 90095, USA

*Correspondence:

Aldons J. Lusis

UCLA School of Medicine, A2-237 CHS, UCLA

Los Angeles, CA 90095-1679

Phone: (310) 825-1359; fax: (310) 794-7345

Email: jlusis@mednet.ucla.edu

Keywords: host-microbiota interactions, heritability, obesity, insulin resistance, systems genetics; genome-wide association study; *Akkermansia muciniphila*

Abstract

Genetics provides a potentially powerful approach to dissect host-gut microbiota interactions. Toward this end, we profiled gut microbiota using 16s rRNA gene sequencing in a panel of 110 diverse inbred strains of mice. This panel has previously been studied for a wide range of metabolic traits and can be used for high resolution association mapping. Using a SNP-based approach with a linear mixed model we estimated the heritability of microbiota composition. We conclude that in a controlled environment the genetic background accounts for a substantial fraction of abundance of most common microbiota. The mice were previously studied for response to a high fat, high sucrose diet, and we hypothesized that the dietary response was determined in part by gut microbiota composition. We tested this using a cross-fostering strategy in which a strain showing a modest response, SWR, was seeded with microbiota from a strain showing a strong response, AxB19. Consistent with a role of microbiota in dietary response, the cross-fostered SWR pups exhibited a significantly increased response in weight gain. To examine specific microbiota contributing to the response, we identified various genera whose abundance correlated with dietary response. Among these, we chose *Akkermansia muciniphila*, a common anaerobe previously associated with metabolic effects. When administered to strain AxB19 by gavage, the dietary response was significantly blunted for obesity, plasma lipids, and insulin resistance. In an effort to further understand host-microbiota interactions, we mapped loci controlling microbiota composition and prioritized candidate genes. Our publically available data provide a resource for future studies.

Introduction

Studies carried out over the last decade have revealed that gut microbiota contribute to a variety of common disorders, including obesity and diabetes (Musso et al. 2011), colitis (Devkota et al. 2012), atherosclerosis (Wang et al. 2011), rheumatoid arthritis (Vaahtovuo et al. 2008), and cancer (Yoshimoto et al. 2013). The evidence for metabolic interactions is particularly strong, as a large body of data now supports the conclusion that gut microbiota influence the energy harvest from dietary components, particularly complex carbohydrates, and that metabolites such as the short chain fatty acids produced by gut bacteria can perturb metabolic traits, including adiposity and insulin resistance (Turnbaugh et al. 2006; Backhed et al. 2007; Wen et al. 2008; Turnbaugh et al. 2009; Ridaura et al. 2013). Gut microbiota communities are assembled each generation, influenced by maternal seeding, environmental factors, host genetics and age, resulting in substantial variations in composition among individuals in human populations (Eckburg et al. 2005; Costello et al. 2009; Huttenhower and Consortium 2012; Goodrich et al. 2014). Most experimental studies of host-gut microbiota interactions have employed large perturbations, such as comparisons of germ-free versus conventional mice, and the significance of common variations in gut microbiota composition for disease susceptibility is still poorly understood. Furthermore, while studies with germ-free mice have clearly implicated microbiota in clinically relevant traits, it has proven difficult to identify the responsible taxa of bacteria.

We now report a population-based analysis of host-gut microbiota interactions in the mouse. One of the issues we explore is the role of host genetics. Although some evidence is consistent with significant heritability of gut microbiota composition, the extent to which the host controls microbiota composition under controlled environmental conditions is unclear. We also examined the role of common variations in gut microbiota in metabolic traits such as obesity and insulin resistance and mapped loci contributing to the abundance of certain microbiota. We performed our study using a resource termed the Hybrid Mouse

Diversity Panel (HMDP), consisting of about 100 inbred strains of mice that have been either sequenced or subjected to high density genotyping (Bennett et al. 2010). The resource has several advantages for genetic analysis as compared to traditional genetic crosses. First, it allows high resolution mapping by association rather than linkage analysis, and it has now been used for the identification of a number of novel genes underlying complex traits (Farber et al. 2011; Lavinsky et al. 2015; Parks et al. 2015; Rau et al. 2015). Second, since the strains are permanent the data from separate studies can be integrated, allowing the development of large, publically available databases of physiological and molecular traits relevant to a variety of clinical disorders (systems.genetics.ucla.edu and phenome.jax.org). Third, the panel is ideal for examining gene-by-environment interactions, since it is possible to examine individuals of a particular genotype under a variety of conditions (Orozco et al. 2012; Parks et al. 2013).

Results

Variation of gut microbiota in a large panel of mouse strains

We determined the composition and variability of gut microbiota in a total of 599 mice from 110 HMDP strains, of which 327 were male and 273 were female (average 3 mice per strain, Supplemental Table 1). All mice in the study were bred for 2 or more generations in the same facility at UCLA and each strain was maintained in separate cages (2 to 4 mice of each strain). The mice were maintained on a chow diet (6% kcal from fat) until 8 weeks of age, and then placed on a high fat, high sucrose (HF/HS) diet for additional 8 weeks. We performed multiplex 16S rRNA sequencing of the V4 amplicon using the Illumina MiSeq platform. On average 23,048 reads were obtained per sample (range from 6331 to 82,238). Reads were binned into individual samples based on barcode sequence and complementary taxon-based analysis methods were used to compare 16S rRNA sequences across the cecum microbial communities. The relative abundances of phylum, class, order, family and genus

were determined for the 599 mice. We focused on abundant microbes, defined as operational taxonomic units (OTUs) with at least 0.01% relative abundance across all samples (total 439 OTUs).

We previously showed that changing a chow to a HF/HS diet drastically changed microbiota composition across HMDP strains and that these shifts were strongly dependent on the genetic background of the mice (Parks et al. 2013). After HF/HS diet feeding we identified 49 genera, where the 17 most abundant genera were present in at least 75% of the samples (n=599). These seventeen most abundant genera accounted for 68% of reads and included members of the six phyla (Supplemental Table 2). Consistent with previous findings in both mice and humans, the most abundant phyla in the gut were Firmicutes (49.8% ± 10.9) and Bacteroidetes (41.8% ± 9.6). As compared to the chow diet, the HF/HS diet resulted in increased Firmicutes and decreased Bacteroidetes, consistent with previous studies (Wu et al. 2011; Carmody et al. 2015).

Microbiota composition varied greatly across the 110 strains of mice (Fig. 1A, Supplemental Table 2). For instance, the relative abundance of the Firmicutes across all the strains ranged between 20% and 82%. Even larger variations were observed at finer taxonomic levels; for example, a common mucus layer habitant in gut, *Akkermansia muciniphila* (*A. muciniphila*), varied in abundance from 0.005% to 40% across the strains. In contrast to human data, we were not able to detect any members of hydrogen-consuming methanogens, although not all methanogens, which are Archaea, would be expected to amplify with bacterial 16S RNA gene primers.

Heritability estimation of gut microbiota composition

In examining individual mice housed in separate cages (to avoid cage effects since mice are coprophagic), we found that microbiota compositions were much more similar within strains than between strains ($p < 0.001$ for unweighted and weighted UniFrac) (Fig. 1B and Supplemental Fig. 1). However, because the members of an inbred strain share a

recent common ancestor, it is unclear to what extent the shared microbiota result from parent to offspring transfer of microbiota as compared to host genetic factors. The standard ways to estimate heritability, defined in an outbred population as the proportion of the phenotypic variance contributed by the genetic variance (Falconer and Mackay 1996; Lynch and Walsh 1998), are to examine pedigrees or compare monozygotic with dizygotic twins. In mice, this is traditionally done using intercrosses between strains differing in traits of interest. Because of the problem of maternal seeding, we estimated heritability by a different method, based on the proportion of phenotype variance accounted for by genetic relationships among the strains.

All of the HMDP strains have been either sequenced or densely genotyped (http://www.jax.org) allowing us to determine their genetic relatedness. Based on this SNP-based approach (rather than a family-based approach), we were able to estimate the heritability of the abundance of the major taxa of gut microbiota (Supplemental Table 3.). For the calculation, we utilized a linear mixed model and assumed additive effects (see Methods). The assumption behind the linear mixed model approach is that the covariance of the genetic component of the phenotypic data is proportional to the kinship matrix or genetic similarity matrix between the animals. In this model each individual mouse microbiome composition (relative abundance of each taxa) is affected by a genetic random effect, which is correlated across mice by virtue of sharing some of the genetic variants affecting microbiome, and an environmental random effect, which is uncorrelated across mice.

When maintained under controlled conditions, host genetic variation appears to explain a substantial amount of the variation in gut microbiota composition in the HMDP, up to 0.5 or more for many common taxa (Table 1; Supplemental Table 3). The range of heritabilities of microbiota was similar for phyla, families and genera, and for males and females (Supplemental Table 3) and approached the range of heritabilities we observed for measured clinical phenotypes (Supplemental Table 4).

Our approach for estimating the heritability of gut microbiota composition is potentially confounded by the complication of physical transmission (Ubeda et al. 2012; McCafferty et al. 2013). However, since most of the inbred strains have been separated for many decades, it seems unlikely that a particular composition would be maintained over so many generations if it was due largely to physical "seeding". A second possible caveat in this analysis is that multiple animals from the same strain were, in some cases, housed in the same cage and they may share similar microbiota compositions due to physical transfer rather than host genotype. We rule out this confounding by performing the same analysis using only one animal per strain (Supplemental Table 3). We expected lower heritability than in the complete cohort because of reduced total genetic relatedness and power. However, even in the reduced sample, while our estimates are lower than the complete dataset, host genotype still accounts for a substantial fraction of the variation.

Gut microbiota contribute to dietary responsiveness

Among the HMDP strains examined were striking differences in response to the HF/HS diet. Some strains showed as much as a 6-fold increase in body fat whereas others showed no significant change, and food consumption was only modestly associated with the gain in body weight (r^2 = 0.30) or the gain in fat mass (r^2 = 0.04) (Parks et al. 2013). Likewise, HOMA-IR, a measure of insulin resistance, showed over a 50-fold range among the strains (Parks et al. 2015). We hypothesized that the composition of the gut microbiota might contribute to this variation.

To test the hypothesis, we performed neonatal cross-fostering experiments between two strains, AxB19 and SWR, exhibiting diverse responses after HF/HS feeding. Figure 2A shows the response of the two strains to the diet in terms of body fat increase during 8 weeks HF/HS diet feeding. The AxB19 strain gained about 24g fat in response to the diet while strain SWR gained about 4g (Fig. 2A, Supplemental Fig. 2A). We cross-fostered newborn SWR mice with AxB19 dams and observed that the gut microbiota composition of

the cross-fostered pups resembled AxB19 mice rather than SWR mice, indicating efficient transfer (Fig. 2B). At 4 weeks of age, we placed the pups on the HF/HS diet and monitored fat gain. In a pilot study (Supplemental Fig. 2B), both cross-fostered male and female mice initially exhibited increased weight gain as compared to SWR mice. While the male mice continued to show an increased response up to 8 weeks on the diet, the female mice became similar to SWR mice after 8 weeks. We then repeated the study with a larger group of cross-fostered mice (n=8-11 per group). Again, the male cross-fostered mice showed significantly more weight gain and higher body fat composition ($p<0.01$) as compared to SWR control mice (Fig. 2C and 2D). In addition, cross-fostered male SWR mice also showed higher levels of plasma triglyceride compared to SWR mice (Fig. 2E). The female mice, on the other hand, did not exhibit a significant increase in body fat at 8 weeks of age (Supplemental Fig. 2C). After 8 weeks of HF/HS diet the microbiota composition of the cross-fostered SWR pups moved back towards to SWR microbiota composition, supporting the role of host genotype in microbiota community structure (Fig.2F). We conclude that common variations in the composition of the gut microbiome contribute in part to the response to a HF/HS diet.

Gut microbiota associations with metabolic and cardiovascular traits

To identify which bacteria contribute to obesity and metabolic phenotypes we sought to identify potential relationships between metabolic traits and the gut microbiota. Altogether, we identified many correlations, including some novel and some known relationships (Supplemental Table 5). Several of these appear to be significant, based on a false discovery rate of less than 0.01.

Two taxa from the family Lachnospiracea, *Roseburia spp.* and *Ruminococcus gnavus,* were positively associated with obesity and metabolic traits including body fat increase on a HF/HS diet, insulin levels and HOMA-IR ($p< 0.001$). Same traits were also associated with an unknown species of *Lactobacillus* (Supplemental Table 5). Our data are consistent with a recent study, showing that increased abundance of *Roseburia spp.* in obese subjects is

positively correlated with body mass index and inflammation (Tims et al. 2013; Verdam et al. 2013), and *Lactobacillus reuteri* has been previously linked to increased obesity in humans (Million et al. 2012). *A. muciniphila* was inversely correlated with body fat (r=-0.15; $p = 9.02 \times 10^{-4}$) and insulin levels (r=-0.20; $p = 4.57 \times 10^{-6}$) (Supplemental Table 5). *A. muciniphila* is a mucin-degrading, gram-negative anaerobe residing in intestinal mucus layers that has been associated with obesity and insulin resistance in humans and mice (Derrien et al. 2011; Everard et al. 2013).

***Akkermansia muciniphila* treatment improves obesity and metabolic parameters in mice fed a high-fat/high-sucrose diet.**

To test causality of the relationship, we administrated live or heat-killed *A. muciniphila* to obesity-prone AxB19 male mice (Supplemental Fig. 3A). Ten-week old male AxB19 mice were treated five times per week with *A. muciniphila* by oral gavage at a dose of 1.44×10^9 cfu/0.2 mL (HF/HS-Akk) while control mice were treated with an oral gavage of an equivalent volume of heat-killed *A. muciniphila* (HF/HS). After one week of gavage all mice were put on HF/HS diet for 4 additional weeks. After five weeks of gavage we observed that mice given *A. muciniphila* showed significantly improved metabolic parameters. Figure 3A shows that body weight and total body fat, including all fat depots examined, were significantly reduced in *A. muciniphila* treated mice. Plasma lipid levels showed substantial decreases in total cholesterol and triglycerides (Fig. 3B). Most striking were the effects on insulin resistance, with dramatically decreased levels of both glucose and insulin (Fig. 3B). Our data are consistent with the correlations observed among the HMDP as well as recent findings (Everard et al.).

In addition to metabolic changes, the administration of *A. muciniphila* altered the gut microbiota composition (Fig. 3C). Thus, using both chow and HF/HS diets *A. muciniphila* treated mice clustered separately from mice that received heat-inactivated bacteria (Supplemental Figure 3A and 3B). This difference was evident at phylum level, showing

significant shifts between two dominant phyla, *Bacteoridetes* and *Firmicutes* (Supplemental Figure 3C). Surprisingly, we did not observe significant differences in the total abundance of *A. muciniphila*, perhaps because DNA from heat-killed bacteria was also present.

Genome-wide association (GWAS) analysis of loci controlling gut microbiota in mice

Next we aimed to obtain evidence for specific interactions between gut microbiota and host genetics. Rather than using linkage analysis, as is traditional in mouse genetics, we employed association across the HMDP strains, since the resolution of mapping is one or two orders improved (Bennett et al. 2010). Such association analysis has now been used to identify novel genes which were subsequently validated in a number of cases (Farber et al. 2011; Lavinsky et al. 2015; Parks et al. 2015; Rau et al. 2015), but given the structure of the inbred mouse population, there is some potential for long-range linkage disequilibrium. The proportion of each common taxon was treated as an individual trait and association analyses were performed with 198,431 informative SNPs spaced throughout the mouse genome using a mixed-model algorithm that corrects for population structure (Kang et al. 2008). The threshold for genome-wide significance was based on simulation and permutations as previously described (Farber et al. 2011). Altogether, seven genome-wide significant loci ($P < 4 \times 10^{-6}$) were found to be associated with common genera (Fig. 4; Supplemental Table 6). Loci ranged from 800 kb to 3 Mb in size and in most cases contained several genes within a linkage disequilibrium block. The majority of these significant associations were detected with members of the classes Clostridia (Lachospiraceae, Ruminococcacea and Bacilli) and most exhibited similar associations in both sexes (Supplemental Table 6). In order to test whether the GWAS results were inflated by the effect that multiple animals from the same strain were housed in the same cage, we performed GWAS using only one animal per strain. Even with the reduced sample size we were able to detect GWAS associations in the same regions (Supplemental Table 6), albeit with reduced significance. We have carried out expression profiling of adipose and liver of

the HMDP strains when maintained on HF/HS diets and used the data to identify *cis* expression quantitative trait loci (eQTL). These provide a useful means of prioritizing candidate genes at the relevant loci since they provide evidence of functional variation (Civelek and Lusis 2014). The significant cis eQTLs at each of the GWAS loci are shown in Supplemental Figure 5 and Supplemental Table 7. Here we focus on those genera that show strong correlation with clinical traits, as discussed above; additional loci are described in detail in Supplemental Data.

For *Roseburia spp.* we identified significant associations spanning 2.6 Mb on chromosome 15 (Fig. 4A, Supplemental Table 6). The same region showed a significant association with subcutaneous fat mass on a HF/HS diet ($p<10^{-7}$) (Supplemental Figure 4A), a clinical trait that is also positively correlated with the abundance of *Roseburia spp.* ($r=0.25$, $p=3.9\times10^{-10}$) (Table 2). Global gene expression in epididymal adipose tissue and liver showed a significant *cis*-eQTL between the peak SNP (rs31843241) and transcript levels of the *Kif21a*, *Lrrk2* and *Irak4* genes (Table 2, Fig. 5A; Supplemental Figure 5, Supplemental Table 7). The expression of *Irak4*, a gene involved in the initiation of the innate immune response, was correlated with the abundance of *Roseburia spp.* and HOMA-IR, suggesting a causal relationship (Fig. 5B and 5C).

Ruminococcus gnavus exhibited genome-wide significant association to a locus on chromosome 19 (Fig. 4B, Supplemental Table 6). The peak SNP (rs30796836, $p=8.37 \times 10^{-7}$) has a significant cis-eQTL with the transcript levels of the *Osbp* (oxysterol binding protein) gene in adipose tissue of HF/HS fed mice (females: $p=1.51 \times 10^{-12}$; males: $p = 2.54 \times 10^{-10}$), and the abundance of *Ruminococcus gnavus* negatively correlated with expression of *Osbp* ($r=-0.36$; $p = 0.00014$) (Table 2, Supplemental Figure 5, Supplemental Table 7).

Finally, we detected significant associations for the abundance of *A. muciniphila* on chromosomes 7 (rs33129247; $p = 2.59 \times 10^{-6}$) and 2 (rs27323290, $p = 6.67 \times 10^{-6}$) (Figs. 4C and 6A, Supplemental Table 6). The peak SNP on chromosome 7 (rs33129247) was also associated with triglyceride levels ($p = 6.47 \times 10^{-9}$) and gonadal fat ($p = 7.44 \times 10^{-7}$) (Table 2,

Fig. 6B, Supplemental Figure 4B). Strong candidates for this locus are *Igf1r* and *Nr2f2* genes, since both have been shown to play a role in glucose and insulin regulation (Ueki et al. 2006; Garg et al. 2011). The chromosome 2 locus contains *Ctnnbl1,* a gene implicated in obesity (Liu et al. 2008; Tan et al. 2014) and two interesting candidates, bactericidal/permeability-increasing protein (*Bpi*) and lipopolysaccharide binding protein (*Lpb*). *Ctnnbl1* also showed a significant association with food intake ($p=1.17 \times 10^{-9}$) and total weight after 8 weeks on HF/HS diet ($p= 5.8 \times 10^{-8}$) (Table 2). The *Lbp* and *Ctnnbl1* genes both have significant *cis*-eQTLs in adipose and liver (Fig. 6C, Table 2, Supplemental Figure 5, Supplemental Table 7) and are associated with body fat percentage increase and insulin levels (Fig. 6D and 6E).

Discussion

We previously showed that inbred strains of mice differ dramatically in their response to a high fat, high sucrose diet (Parks et al. 2013). Based on the large body of evidence indicating that gut microbiota can influence metabolic traits (Backhed et al. 2004; Turnbaugh et al. 2006; Backhed et al. 2007; Turnbaugh et al. 2009; Ridaura et al. 2013), we hypothesized that the dietary response was dictated in part by differences in gut microbiota. We showed that different inbred strains differ strikingly in the composition of gut microbiota and provided evidence that the variation is determined in part by the host genetic background. Consistent with our hypothesis, we showed that cross-fostering between two strains of mice affected dietary response to the high fat, high sucrose diet. By correlating microbiota composition with dietary response among the HMDP inbred strains, we were able to identify several candidate microbiota influencing dietary response. We chose one of these, *A, muciniphila*, to examine using gavage with the cultured microbe, and observed striking effects on weight gain, adiposity, plasma lipids, and insulin resistance. Finally, to help identify novel host-microbiota interactions, we have mapped loci controlling certain microbiota taxa. We discuss each of these findings in turn below.

Experimental studies have shown that the host genetic background can influence gut microbiota composition. For example, mice with mutations affecting inflammatory signaling or diabetes differ in microbiota composition from their wild-type littermates (Henao-Mejia et al. 2012; Peng et al. 2014). But the importance of common genetic variations in contributing to the composition of the gut microbiota is unclear. While twin pairs and related individuals share gut microbiota to a greater extent than unrelated individuals, early studies did not find a statistically significant difference in gut microbiota sharing in monozygotic (MZ) and dizygotic twins. However a recent study with 416 twin pairs demonstrated that MZ twins have greater overall microbial community similarities than DZ twin pairs and identified several microbial taxa with relative abundances that differ depending on host genetics (Goodrich et al. 2014). Studies using genetic crosses of mice, where the environment can be controlled, also suggest that host genetics can significantly alter gut microbiota composition (Benson et al. 2010; McKnite et al. 2012; Srinivas et al. 2013).

Heritability represents the fraction of variation that is attributable to genetic variation and is a relative value that depends on the environment and the degree to which the population varies. Traditionally, heritability has been estimated using pedigrees in outbred populations or by comparing monozygotic versus dizygotic twins. In mice, heritability is generally estimated by analyzing genetic crosses, but for studies of gut microbiota this is confounded by the fact that there is physical transmission of "seed" microbiota from generation to generation (Ubeda et al. 2012). To circumvent this problem, we used a SNP-based approach to determine relatedness (Yang et al. 2010) rather than a family-based approach. All of the inbred strains constituting the HMDP are separated from one another by many generations (Silver 1995) and thus are unlikely to share microbiota as a result of physical transmission. Our results indicate a high degree of heritability of the major groups of microbiota in mice, ranging from about 0.3 to more than 0.5, although we note certain caveats in our approach (see Results). This high heritability presumably results

352 from the fine tuning of a symbiotic relationship that has co-evolved for millions of years.
353 Among the likely contributing factors are differences in immunoglobulin and antibacterial
354 molecules secreted into the gut lumen (Wen et al. 2008; Vijay-Kumar et al. 2010;
355 Shulzhenko et al. 2011), differences in the mucosal gut structure (Sommer et al. 2014;
356 Wlodarska et al. 2014), and differences in bile acid metabolism (Ryan et al. 2014).
357
358 Previous studies have shown that certain large effect mutations affecting
359 inflammatory signaling or metabolic traits can significantly affect microbiota composition
360 and that, in some cases, these can be transmitted by transplantation of gut microbiota from
361 such mice (Henao-Mejia et al. 2012; Peng et al. 2014). Here, we have examined whether
362 common variations in gut microbiota are also causally involved in metabolic traits. To test
363 this possibility, we chose two inbred strains, AxB19 and SWR, that differ strikingly in the
364 response to a HF/HS diet for cross-fostering studies. When male SWR mice cross-fostered
365 by AxB19 dams were subjected to the HF/HS diet, they gained ~8% in total body fat as
366 compared to ~2% for SWR, while AxB19 gained about 25% (Fig. 2C). Thus, while the
367 majority of the response was dictated by the host genetic background, the gut microbiota
368 did contribute significantly. Human gut microbiota exhibit more diversity and quantitative
369 variation than that we have observed among common mouse strains, suggesting that a
370 significant fraction of variance in obesity and insulin resistance in human populations is due
371 to microbiota composition. Conceivably, this could explain some of the "missing heritability"
372 observed in GWAS. We used correlation analysis to identify candidate microbiota
373 contributing to the response to the HF/HS diet. Several genera were found to be strongly
374 associated with traits such as body fat, plasma lipids, and insulin resistance, including some,
375 such as *Roseburia spp.* and *A. muciniphila*, that have previously been implicated in metabolic
376 traits. We chose *A. muciniphila*, a mucin-degrading, gram negative anaerobe, to examine its
377 potential response to our dietary challenge. We administered by oral gavage either live or
378 heat killed bacteria to strain AxB19 male mice for one week and then began the HF/HS

challenge, continuing to administer the live or heat killed bacteria for a total of 5 weeks. Significant differences were observed between the groups in body fat gain, plasma lipid levels, and insulin resistance. We noted significant correlations between the abundances of certain taxa across the panel of strains (data not shown), providing information about microbiota community interactions.

Finally, we used the HMDP to perform high resolution mapping of loci contributing to microbiota abundance. Using association analysis, we identified seven significant loci for 5 out of 17 common genus level taxa. Most of the loci were observed in both males and females, supporting the conclusion that they are true positives. The loci contain a number of strong candidate genes based on the literature, functional variations, and correlations with clinical and molecular traits (see Supplemental Data). The chromosome 15 locus for *Roseburia* spp. contains the *Irak4* gene, which is involved in signaling innate immune responses from Toll-like receptors (Flannery and Bowie 2010; Liu et al. 2011). Mice deficient in *Irak4* expression are more susceptible to viral and bacterial infectious (Suzuki et al. 2002) and *Irak4* has previously been associated with gut microbiota composition in a subset of BxD RI strains (McKnite et al. 2012). The *A. muciniphila* locus in chromosome 2 contains two closely linked genes, *Bpi* and *Lbp*. Both bind to bacterial lipopolysaccharide (LPS) and elicit immune responses by presenting LPS to CD14 and TLR4 and signaling the acute-phase immunological response (Muta and Takeshige 2001). *Bpi* acts as an endogenous antibiotic protein with potent killing activity against Gram-negative bacteria (Wittmann et al. 2008).

Our data constitute a resource for the further dissection of mechanistic host-gut microbiota interactions. We have identified a number of highly significant associations between gut microbiota and clinical traits and the loci reported here provide a means of identifying novel host factors controlling gut microbiota abundances.

Methods

406

407 **1. Sample collection and study design**

408 All mice were obtained from The Jackson Laboratory and were bred at UCLA for 2 or
409 more generations for use in this study. Briefly, until 8 weeks of age mice were maintained
410 on a chow diet (Ralson Purina Company) and then placed on a high-fat, high-sucrose diet
411 (Research Diets D12266B) for an additional 8 weeks (Parks et al. 2013). Samples were
412 obtained from the cecum of 599 mice from 113 strains, with an average of 6 mice per strain
413 (327 males and 297 females) (Supplemental Table 1). Mice from different strains and
414 genders were housed in separate cages, but in the same room throughout the study. Cecum
415 and fecal samples were snap frozen with liquid nitrogen and stored at -80°C. The animal
416 protocol for the study was approved by the Institutional Care and Use Committee (IACUC)
417 at University of California, Los Angeles.

418

419 **2. Sample Preparation and sequencing of 16S rRNA genes**

420 Microbial DNA was extracted using the PowerSoil DNA Isolation Kit (MO BIO
421 Laboratories, Carlsbad, CA, USA). Amplification and sequencing of V4 hypervariable region
422 of the 16S rRNA gene was performed using the validated, region-specific bacterial primers
423 515F and 806R according to previously described methods (Caporaso et al. 2012)
424 optimized for the Illumina MiSeq platform. The reverse amplification primer contained a
425 12-bp Golay error-correcting barcode sequence and amplicons were generated in triplicate
426 using 5 Prime Hot MasterMix (Fischer Scientific, UK). The PCR conditions consisted of an
427 initial denaturation step of 94°C for 3 min; 35 cycles of 94°C for 45 sec, 50°C for 30 sec, and
428 72°C for 90sec, followed by 72°C for 5 min. Replicate amplicons were quantified with
429 Quant-iT™PicoGreen®dsDNA Assay Kit (Life Technologies Corporation, Carlsbad, CA, USA),
430 pooled (200ng from 96 samples) and purified using with the UltraClean® PCR Clean-up Kit
431 (MOBIO Laboratories, Inc., Carlsbad, CA, USA). High-throughput sequencing analysis of
432 bacterial rRNA genes was performed on the purified, pooled sample using the Illumina

MisSeq platform (Illumina, Inc., San Diego, CA, USA).

De-multiplexing 16S rRNA gene sequences, quality control and operational taxonomic unit (OTU) binning were performed using the open source pipeline Quantitative Insights Into Microbial Ecology (QIIME) version 1.7.0 (Caporaso et al. 2010; Bokulich et al. 2013). The total number of sequencing reads was 13,805,813 (an average of 23,048 reads per sample) with an average length of 153 base pair reads. Sequences were binned into OTUs based on 97% identity using UCLUST (Edgar 2010) against the Greengenes reference database (McDonald et al. 2012). Each sample's sequences were rarefied to 7,000 reads to reduce the effect of sequencing depth. Seven samples were omitted from further analysis due to insufficient sequence coverage, yielding 592 samples.

Microbial composition at each taxonomic level was defined using the summarize_taxa function in QIIME. Prior to genome-wide association analysis, taxa at any taxon present in fewer than 75% of samples was discarded, yielding in total of 43 common taxa from different taxonomic levels (3 phyla, 5 classes, 6 order, 12 family and 17 genus level taxa). The relative abundance of each taxon was calculated by dividing the sequences pertaining to a specific taxon by the total number of bacterial sequences for that sample.

3. Heritability Calculations

Heritability was estimated using a linear mixed with the EMMAX software model (Kang et al. 2010). In this approach the phenotypes (in this case the relative abundances) are assumed to be generated by genetic and environmental components. The assumption behind the linear mixed model approach is that the covariance of the genetic component of the phenotypic data is proportional to the kinship or genetic similarity matrix between the animals. The analysis provides estimates of σ_g^2 and σ_e^2, the variances corresponding to the genetic and environmental component respectively. The heritability is then the fraction of the variance accounted for by the genetics or

459 $$\frac{\sigma_g^2}{\sigma_g^2 + \sigma_\varepsilon^2}$$

460 and is computed for each relative abundance. We note that the kinship matrix must be
461 standardized for these estimates to be consistent with the classical definition of heritability
462 (Kostem and Eskin 2013; Speed and Balding 2015). A standardized kinship matrix has a
463 mean along the diagonal of 1 and a sum of 0.
464
465 **4. Clinical Traits**
466 Body composition, food intake, and blood and plasma assays were as previously
467 described (Parks et al. 2013; Parks et al. 2015). Briefly, mice were measured for total body
468 fat mass and lean mass using magnetic resonance imaging (NMR) using the BurkerMinispec
469 with software from Eco Medical Systems (Houston, TX). Blood was collected from mice
470 following fasting for 4-5 hours and plasma was isolated by centrifugation in Microtainer
471 tubes with EDTA (Becton, Dickinson and Company, Franklin Lakes, NJ). Plasma glucose
472 levels were measured using a Beckman Glucose Analyzer 2 (Beckman Instruments). Plasma
473 total cholesterol, HDL cholesterol, free cholesterol, triglycerides, and free fatty acid
474 concentrations were determined by enzymatic assays employing colorimetric endpoints as
475 described previously (Hedrick et al. 1993). Insulin levels and HOMA-IR were determined as
476 previously described (Castellani et al. 2008).
477 **5. Cross-fostering Study**
478 Within 24-hours of birth, the pups from SWR females were removed from birthing
479 cages and placed with AxB19/PgnJ mothers. Pups were weaned on postnatal day 21 and at
480 8 weeks of age mice were placed on a high-fat/high-sucrose diet (Research Diets D12266B)
481 for additional 8 weeks. Controls from both strains were fostered with different mothers
482 from the same strain.
483
484 **6. *Akkermansia muciniphila* gavage**

A. muciniphila (ATCC BAA-835) was grown in a Columbia broth medium supplemented with 0.05 % hog gastric Mucin type III (Sigma) under anaerobic conditions (Ganesh et al. 2013). Cells were harvested in late logarithmic phase by centrifugation at 6.000 rpm in a table top Fisher at room temperature and re-suspended in 0.05 volume sterile anaerobic PBS containing 25% glycerol to a concentration of $7.2x10e^{10}$ per ml prior to storage at -80 C. For gavage, anaerobic cell suspensions were diluted ten-fold in PBS. The sterile anaerobic PBS (pH 7) was supplemented with 0.05% cysteine HCl, degassed with N_2, sealed in serum bottles with butyl rubber stopper under anaerobic conditions provided by a gas phase of 1.8 atm N_2/CO_2 (80:20, vol/vol).

A. muciniphila was administrated to ten-week old AxB19 male mice (n= 5 per group), housed in groups of 2-3 mice per cage. by oral gavage at a dose $1.44x10^9$ cfu/0.2 mL suspended in sterile anaerobic PBS (HF-Akk). Treatment was five days per week 5 weeks. Control groups were treated with an equivalent volume of heat inactivated *A. muciniphila*. After the first week of *A. muciniphila* treatment all mice were placed on the HF/HS diet for an additional 4 weeks.

7. Association Analyses

Association analyses of taxa were performed using the Factored Spectrally Transformed Linear Mixed Models (FaST-LMM) algorithm adjusting for population structure and using gender as a covariate (Kang et al. 2008; Lippert et al. 2011). To achieve a normal distribution, the sequence counts for each taxonomic bin were Arcsine transformed. A total of 198,431 informative SNPs (minor allele frequency >5%; missing genotype rate <10%) spaced throughout the genome were used. For genome-wide significance we used a p value threshold $< 4 \times 10^{-6}$, based on permutation and simulation and which roughly corresponds to a Bonferroni correction (Bennett et al. 2010; Farber et al. 2011). In some cases multiple animals from the same strain were housed together, and in order to rule out the possibility that GWAS results were inflated because strains shared

similar microbiota compositions due to physical transfer, we also performed GWAS choosing a single sample from each strain at random (n=113). Linkage disequilibrium (http://pngu.mgh.harvard.edu/~purcell/plink/) boundaries were determined by calculating SNP correlations and visualizing $r^2 > 0.8$ in Haploview.

8. Expression QTL analysis

To help identify candidate genes at loci associated with taxa abundances, we carried out global expression analysis of epididymal adipose and liver tissue in male and female mice (16 weeks old) fed a HF/HS diet as described (Parks et al. 2015). Isolated RNA (2 mice per strain) was analyzed for global gene expression using Affymetrix HT_MG-430A arrays and filtered as described (Bennett et al. 2010). Microarray data have been submitted to the Genome Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE64770. The loci controlling transcript levels were mapped with FaST-LMM and are referred to as expression quantitative trait loci (eQTL). Loci are defined as *cis* if the peak SNP mapped within 1 Mb of gene position (*p* value threshold$< 1.4 \times 10^{-3}$).

9. Statistics

All correlations involving bacterial relative abundance were performed using biweight midcorrelation, which is robust to outliers (Wilcox 2005). The statistical cutoff of $P = 0.1$ after False Discovery Rate (FDR) correction for multiple comparisons was used to define statistical significance for correlations. Statistical analyses were performed using GraphPad Prism. Data are expressed as mean ± SEM, and significance was set at a two-tailed p value < 0.05.

Data access

16S rRNA sequencing data generated for this study have been submitted to the Sequence Read Archive (SRA) under accession number SRP059760. The summary tables for

539 both genders are posted on our website (systems.genetics.ucla.edu) with the link called
540 "Download high-fat microbiota vs clinical trait correlation table" and can be also find in
541 Supplemental Data (Supplemental Table S8).
542
543 **Acknowledgements**
544 We thank Hannah Qi, Zhiqiang Zhou, Judy Wu, Tieyan Han, and Richard Davis for
545 expert assistance with mouse experiments, and Jonathan T. Furuta for growing *A.*
546 *muciniphila*. This work was supported by National Institutes of Health (NIH) grants
547 HL028481, HL30568, and DK094311 to A.J.L. and UCLA Iris Cantor Women's Health Center
548 and UCLA CTSI to T.A.D. E.O. was supported by a MOBILITAS Postdoctoral Research Grant
549 (MJD252) and FP7 grant no 330381. B.W.P. was supported by NIH training grant T32-
550 HD07228.
551 *Author Contributions*: A.J.L, E.E and T.A.D supervised the study. R.K provided
552 comments and discussion. E.O and B.W.P oversaw collection of samples. E.O, W.S. and B.E
553 collected microbial data. R.G. prepared *A. muciniphila* for gavage experiment. E.O performed
554 the analysis with contribution from W.J.J, E.Y.Y., C.P and R.K. E.O. A.J.L and E.E. prepared the
555 manuscript with comments from B.E, T.A.D. and R.K.
556
557 **Disclosure statement:** The authors declare no conflict of interest
558

559 **References**
560
561
562 Backhed F, Ding H, Wang T, Hooper LV, Koh GY, Nagy A, Semenkovich CF, Gordon JI.
563 2004. The gut microbiota as an environmental factor that regulates fat
564 storage. *Proc Natl Acad Sci U S A* **101**(44): 15718-15723.
565 Backhed F, Manchester JK, Semenkovich CF, Gordon JI. 2007. Mechanisms
566 underlying the resistance to diet-induced obesity in germ-free mice. *Proc*
567 *Natl Acad Sci U S A* **104**(3): 979-984.
568 Bennett BJ, Farber CR, Orozco L, Kang HM, Ghazalpour A, Siemers N, Neubauer M,
569 Neuhaus I, Yordanova R, Guan B et al. 2010. A high-resolution association
570 mapping panel for the dissection of complex traits in mice. *Genome Res*
571 **20**(2): 281-290.
572 Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J, Zhang M, Oh PL, Nehrenberg D,
573 Hua K et al. 2010. Individuality in gut microbiota composition is a complex
574 polygenic trait shaped by multiple environmental and host genetic factors.
575 *Proc Natl Acad Sci U S A* **107**(44): 18933-18938.
576 Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, Mills DA,
577 Caporaso JG. 2013. Quality-filtering vastly improves diversity estimates from
578 Illumina amplicon sequencing. *Nat Methods* **10**(1): 57-59.
579 Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer
580 N, Pena AG, Goodrich JK, Gordon JI et al. 2010. QIIME allows analysis of high-
581 throughput community sequencing data. *Nat Methods* **7**(5): 335-336.
582 Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM,
583 Betley J, Fraser L, Bauer M et al. 2012. Ultra-high-throughput microbial
584 community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* **6**(8):
585 1621-1624.
586 Carmody RN, Gerber GK, Luevano JM, Jr., Gatti DM, Somes L, Svenson KL, Turnbaugh
587 PJ. 2015. Diet dominates host genotype in shaping the murine gut microbiota.
588 *Cell Host Microbe* **17**(1): 72-84.
589 Castellani LW, Nguyen CN, Charugundla S, Weinstein MM, Doan CX, Blaner WS,
590 Wongsiriroj N, Lusis AJ. 2008. Apolipoprotein AII is a regulator of very low
591 density lipoprotein metabolism and insulin resistance. *J Biol Chem* **283**(17):
592 11633-11644.
593 Civelek M, Lusis AJ. 2014. Systems genetics approaches to understand complex
594 traits. *Nat Rev Genet* **15**(1): 34-48.
595 Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. 2009. Bacterial
596 community variation in human body habitats across space and time. *Science*
597 **326**(5960): 1694-1697.
598 Derrien M, Van Baarlen P, Hooiveld G, Norin E, Muller M, de Vos WM. 2011.
599 Modulation of Mucosal Immune Response, Tolerance, and Proliferation in
600 Mice Colonized by the Mucin-Degrader Akkermansia muciniphila. *Frontiers*
601 *in microbiology* **2**: 166.
602 Devkota S, Wang Y, Musch MW, Leone V, Fehlner-Peach H, Nadimpalli A,
603 Antonopoulos DA, Jabri B, Chang EB. 2012. Dietary-fat-induced taurocholic
604 acid promotes pathobiont expansion and colitis in Il10-/- mice. *Nature*
605 **487**(7405): 104-108.

606 Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR,
607 Nelson KE, Relman DA. 2005. Diversity of the human intestinal microbial
608 flora. *Science* **308**(5728): 1635-1638.
609 Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST.
610 *Bioinformatics* **26**(19): 2460-2461.
611 Everard A, Belzer C, Geurts L, Ouwerkerk JP, Druart C, Bindels LB, Guiot Y, Derrien
612 M, Muccioli GG, Delzenne NM et al. Cross-talk between Akkermansia
613 muciniphila and intestinal epithelium controls diet-induced obesity. *Proc*
614 *Natl Acad Sci U S A* **110**(22): 9066-9071.
615 -. 2013. Cross-talk between Akkermansia muciniphila and intestinal epithelium
616 controls diet-induced obesity. *Proc Natl Acad Sci U S A* **110**(22): 9066-9071.
617 Falconer DS, Mackay TFC. 1996. *Introduction to quantitative genetics, 4th edition*.
618 Longman, New York.
619 Farber CR, Bennett BJ, Orozco L, Zou W, Lira A, Kostem E, Kang HM, Furlotte N,
620 Berberyan A, Ghazalpour A et al. 2011. Mouse genome-wide association and
621 systems genetics identify Asxl2 as a regulator of bone mineral density and
622 osteoclastogenesis. *PLoS Genet* **7**(4): e1002038.
623 Flannery S, Bowie AG. 2010. The interleukin-1 receptor-associated kinases: critical
624 regulators of innate immune signalling. *Biochem Pharmacol* **80**(12): 1981-
625 1991.
626 Ganesh BP, Klopfleisch R, Loh G, Blaut M. 2013. Commensal Akkermansia
627 muciniphila exacerbates gut inflammation in Salmonella Typhimurium-
628 infected gnotobiotic mice. *PLoS One* **8**(9): e74963.
629 Garg N, Thakur S, McMahan CA, Adamo ML. 2011. High fat diet induced insulin
630 resistance and glucose intolerance are gender-specific in IGF-1R
631 heterozygous mice. *Biochem Biophys Res Commun* **413**(3): 476-480.
632 Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, Beaumont M, Van
633 Treuren W, Knight R, Bell JT et al. 2014. Human genetics shape the gut
634 microbiome. *Cell* **159**(4): 789-799.
635 Hedrick CC, Castellani LW, Warden CH, Puppione DL, Lusis AJ. 1993. Influence of
636 mouse apolipoprotein A-II on plasma lipoproteins in transgenic mice. *J Biol*
637 *Chem* **268**(27): 20676-20682.
638 Henao-Mejia J, Elinav E, Jin C, Hao L, Mehal WZ, Strowig T, Thaiss CA, Kau AL,
639 Eisenbarth SC, Jurczak MJ et al. 2012. Inflammasome-mediated dysbiosis
640 regulates progression of NAFLD and obesity. *Nature* **482**(7384): 179-185.
641 Huttenhower C, Consortium THMP. 2012. Structure, function and diversity of the
642 healthy human microbiome. *Nature* **486**(7402): 207-214.
643 Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E.
644 2010. Variance component model to account for sample structure in genome-
645 wide association studies. *Nat Genet* **42**(4): 348-354.
646 Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E. 2008.
647 Efficient control of population structure in model organism association
648 mapping. *Genetics* **178**(3): 1709-1723.
649 Kostem E, Eskin E. 2013. Improving the accuracy and efficiency of partitioning
650 heritability into the contributions of genomic regions. *American journal of*
651 *human genetics* **92**(4): 558-564.
652 Lavinsky J, Crow AL, Pan C, Wang J, Aaron KA, Ho MK, Li Q, Salehide P, Myint A,
653 Monges-Hernadez M et al. 2015. Genome-wide association study identifies

nox3 as a critical gene for susceptibility to noise-induced hearing loss. *PLoS Genet* **11**(4): e1005094.

Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. 2011. FaST linear mixed models for genome-wide association studies. *Nat Methods* **8**(10): 833-835.

Liu YJ, Liu XG, Wang L, Dina C, Yan H, Liu JF, Levy S, Papasian CJ, Drees BM, Hamilton JJ et al. 2008. Genome-wide association scans identified CTNNBL1 as a novel gene for obesity. *Hum Mol Genet* **17**(12): 1803-1813.

Liu Z, Lee J, Krummey S, Lu W, Cai H, Lenardo MJ. 2011. The kinase LRRK2 is a regulator of the transcription factor NFAT that modulates the severity of inflammatory bowel disease. *Nat Immunol* **12**(11): 1063-1070.

Lynch M, Walsh JB. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.

McCafferty J, Muhlbauer M, Gharaibeh RZ, Arthur JC, Perez-Chanona E, Sha W, Jobin C, Fodor AA. 2013. Stochastic changes over time and not founder effects drive cage effects in microbial community assembly in a mouse model. *ISME J* **7**(11): 2116-2125.

McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* **6**(3): 610-618.

McKnite AM, Perez-Munoz ME, Lu L, Williams EG, Brewer S, Andreux PA, Bastiaansen JW, Wang X, Kachman SD, Auwerx J et al. 2012. Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. *PLoS One* **7**(6): e39191.

Million M, Angelakis E, Paul M, Armougom F, Leibovici L, Raoult D. 2012. Comparative meta-analysis of the effect of Lactobacillus species on weight gain in humans and animals. *Microbial pathogenesis* **53**(2): 100-108.

Musso G, Gambino R, Cassader M. 2011. Interactions between gut microbiota and host metabolism predisposing to obesity and diabetes. *Annu Rev Med* **62**: 361-380.

Muta T, Takeshige K. 2001. Essential roles of CD14 and lipopolysaccharide-binding protein for activation of toll-like receptor (TLR)2 as well as TLR4 Reconstitution of TLR2- and TLR4-activation by distinguishable ligands in LPS preparations. *Eur J Biochem* **268**(16): 4580-4589.

Orozco LD, Bennett BJ, Farber CR, Ghazalpour A, Pan C, Che N, Wen P, Qi HX, Mutukulu A, Siemers N et al. 2012. Unraveling inflammatory responses using systems genetics and gene-environment interactions in macrophages. *Cell* **151**(3): 658-670.

Parks BW, Nam E, Org E, Kostem E, Norheim F, Hui ST, Pan C, Civelek M, Rau CD, Bennett BJ et al. 2013. Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell Metab* **17**(1): 141-152.

Parks BW, Sallam T, Mehrabian M, Psychogios N, Hui ST, Norheim F, Castellani LW, Rau CD, Pan C, Phun J et al. 2015. Genetic architecture of insulin resistance in the mouse. *Cell Metab* **21**(2): 334-346.

Peng J, Narasimhan S, Marchesi JR, Benson A, Wong FS, Wen L. 2014. Long term effect of gut microbiota transfer on diabetes development. *J Autoimmun* **53**: 85-94.

Rau CD, Wang J, Avetisyan R, Romay MC, Martin L, Ren S, Wang Y, Lusis AJ. 2015. Mapping genetic contributions to cardiac pathology induced by Beta-adrenergic stimulation in mice. *Circulation Cardiovascular genetics* **8**(1): 40-49.

Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V, Henrissat B, Bain JR et al. 2013. Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341**(6150): 1241214.

Ryan KK, Tremaroli V, Clemmensen C, Kovatcheva-Datchary P, Myronovych A, Karns R, Wilson-Perez HE, Sandoval DA, Kohli R, Backhed F et al. 2014. FXR is a molecular target for the effects of vertical sleeve gastrectomy. *Nature* **509**(7499): 183-188.

Shulzhenko N, Morgun A, Hsiao W, Battle M, Yao M, Gavrilova O, Orandle M, Mayer L, Macpherson AJ, McCoy KD et al. 2011. Crosstalk between B lymphocytes, microbiota and the intestinal epithelium governs immunity versus metabolism in the gut. *Nat Med* **17**(12): 1585-1593.

Silver LM. 1995. *Mouse Genetics: Concepts and Applications*. Oxford University Press.

Sommer F, Adam N, Johansson ME, Xia L, Hansson GC, Backhed F. 2014. Altered mucus glycosylation in core 1 O-glycan-deficient mice affects microbiota composition and intestinal architecture. *PLoS One* **9**(1): e85254.

Speed D, Balding DJ. 2015. Relatedness in the post-genomic era: is it still useful? *Nat Rev Genet* **16**(1): 33-44.

Srinivas G, Moller S, Wang J, Kunzel S, Zillikens D, Baines JF, Ibrahim SM. 2013. Genome-wide mapping of gene-microbiota interactions in susceptibility to autoimmune skin blistering. *Nat Commun* **4**: 2462.

Suzuki N, Suzuki S, Duncan GS, Millar DG, Wada T, Mirtsos C, Takada H, Wakeham A, Itie A, Li S et al. 2002. Severe impairment of interleukin-1 and Toll-like receptor signalling in mice lacking IRAK-4. *Nature* **416**(6882): 750-756.

Tan LJ, Zhu H, He H, Wu KH, Li J, Chen XD, Zhang JG, Shen H, Tian Q, Krousel-Wood M et al. 2014. Replication of 6 obesity genes in a meta-analysis of genome-wide association studies from diverse ancestries. *PLoS One* **9**(5): e96149.

Tims S, Derom C, Jonkers DM, Vlietinck R, Saris WH, Kleerebezem M, de Vos WM, Zoetendal EG. 2013. Microbiota conservation and BMI signatures in adult monozygotic twins. *ISME J* **7**(4): 707-717.

Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP et al. 2009. A core gut microbiome in obese and lean twins. *Nature* **457**(7228): 480-484.

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**(7122): 1027-1031.

Ubeda C, Lipuma L, Gobourne A, Viale A, Leiner I, Equinda M, Khanin R, Pamer EG. 2012. Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. *J Exp Med* **209**(8): 1445-1456.

Ueki K, Okada T, Hu J, Liew CW, Assmann A, Dahlgren GM, Peters JL, Shackman JG, Zhang M, Artner I et al. 2006. Total insulin and IGF-I resistance in pancreatic beta cells causes overt diabetes. *Nat Genet* **38**(5): 583-588.

Vaahtovuo J, Munukka E, Korkeamaki M, Luukkainen R, Toivanen P. 2008. Fecal microbiota in early rheumatoid arthritis. *The Journal of rheumatology* **35**(8): 1500-1505.

Verdam FJ, Fuentes S, de Jonge C, Zoetendal EG, Erbil R, Greve JW, Buurman WA, de Vos WM, Rensen SS. 2013. Human intestinal microbiota composition is associated with local and systemic inflammation in obesity. *Obesity* **21**(12): E607-615.

Vijay-Kumar M, Aitken JD, Carvalho FA, Cullender TC, Mwangi S, Srinivasan S, Sitaraman SV, Knight R, Ley RE, Gewirtz AT. 2010. Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**(5975): 228-231.

Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, Dugar B, Feldstein AE, Britt EB, Fu X, Chung YM et al. 2011. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**(7341): 57-63.

Wen L, Ley RE, Volchkov PY, Stranges PB, Avanesyan L, Stonebraker AC, Hu C, Wong FS, Szot GL, Bluestone JA et al. 2008. Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature* **455**(7216): 1109-1113.

Wilcox RR. 2005. *Introduction to Robust Estimation and Hypothesis Testing*. Elsevier/Academy Press, Amsterdam, The Netherlands.

Wittmann I, Schonefeld M, Aichele D, Groer G, Gessner A, Schnare M. 2008. Murine bactericidal/permeability-increasing protein inhibits the endotoxic activity of lipopolysaccharide and gram-negative bacteria. *J Immunol* **180**(11): 7546-7552.

Wlodarska M, Thaiss CA, Nowarski R, Henao-Mejia J, Zhang JP, Brown EM, Frankel G, Levy M, Katz MN, Philbrick WM et al. 2014. NLRP6 inflammasome orchestrates the colonic host-microbial interface by regulating goblet cell mucus secretion. *Cell* **156**(5): 1045-1059.

Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R et al. 2011. Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**(6052): 105-108.

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW et al. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**(7): 565-569.

Yoshimoto S, Loo TM, Atarashi K, Kanda H, Sato S, Oyadomari S, Iwakura Y, Oshima K, Morita H, Hattori M et al. 2013. Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. *Nature* **499**(7456): 97-101.

Figure legends

Figure 1

Phylum-level variability of gut microbiota composition across 113 inbred strains of mice. (A) Columns represent the relative abundance of microbial phyla in 113 strains (327 male and 297 female). (B) Boxplot of β diversity distances between microbial communities

obtained when comparing mice within and between strains . The specific distance metric used is indicated on the axes. *** $p < 0.001$ for Student's t test with 1,000 Monte Carlo simulations. See also Supplemental Table 2.

Figure 2.

Cross-fostering influences dietary responsiveness.

(A) Body fat increase in SWR and AxB19 strains during 8 weeks of HF/HS diet (B) Principal Coordinate Analysis (PCoA) of unweighted UniFrac distances for fecal samples after cross-fostering newborn SWR and AxB19 pups between parents (blue= AxB19 mothers, red= SWR mother and green= cross-fostered SWR pups), $p < 0.05$ for unweighted UniFrac using Student's t test with 1,000 Monte Carlo simulation. (C) Body fat changes after cross-fostering SWR pups with AxB19 mother (CF-SWR) compared to SWR and AxB19 controls. (D) Total body fat percentage after 8 weeks of HF/HS feeding. (E) Plasma triglycerides (TG) levels. (F) PCoA of unweighted UniFrac distances for cecum samples after 8 weeks of HF/HS diet (blue = AxB19, red= SWR, green= cross-fostered SWR) * Indicates significant differences (* $p<0.05$, *** $p<0.001$) according with unpaired two-tailed Student T test. See also supplemental Figure 2.

Figure 3.

***Akkermansia muciniphila* treatment reduces obesity and metabolic syndrome traits in mice fed a HF/HS diet.**

(A) Total body weight, body fat, mesenteric, retroperitoneal, gonadal, and subcutaneous fat depot weights (g per 100 g body weight) in mice treated by oral gavage with live or heat inactivated *A. muciniphila* and fed a HF/HS diet (n=5). (B) Glucose, insulin, HOMA-IR, unesterified cholesterol (UC), total cholesterol (TC) and triglycerides (TG) levels. (C) Relative abundance of bacterial genera between different treatment groups. Data are shown

as means ±SD. * Indicates significant differences (* p<0.05, *** p<0.001) with unpaired two-tailed Student T test.

Figure 4.

Genome-wide association mapping of gut microbiota genera in the HMDP. Association was performed using the FaST-LMM algorithm (Lippert et al. 2011) following correction for population structure using about 200,000 filtered SNPs genotyped in all strains. The black line indicates the threshold for genome-wide significance ($p< 4 \times 10^{-6}$). See also Supplemental Table 6.

Figure 5.

Chromosome 15 locus for abundance of genus *Roseburia spp.* (A) Overlapping genome-wide significant associations with the abundance of *Roseburia spp.* and liver and adipose eQTLs of the *Irak4* gene in HMDP mice fed a HF/HS diet (B, C). Correlations of *Irak4* adipose gene expression with the relative abundance of *Roseburia spp* and HOMA-IR in the HMDP mice. The black line indicates the threshold for genome-wide significance ($p< 4 \times 10^{-6}$). *r*, biweight midcorrelation; *p*, p value. See also Supplemental Tables S6 and S7.

Figure 6.

Chromosome 2 and 7 loci for abundance of A*. muciniphila*

(A) Locus plot for genome-wide significant association of *A. muciniphila* abundance to a chromosome 7 locus, indicating the LD block (shaded in grey) and peak SNP rs33129247. Locations of candidate genes are indicated. (B) Locus plot for association with TG (triglyceride) at the chromosome 7 locus. See also Supplemental Tables S6 and S7. (C) Chromosome 2 locus showing overlapping associations with the abundance of *A. muciniphila* and *cis*-eQTLs of the *Ctnnbl1 and Lbp* genes adipose. (D) Correlation of epididymal adipose gene expression of the *Ctnnbl1* with body fat and insulin levels after 8

weeks of HF/HS diet. (E) Correlation of epididymal adipose gene expression of the *Lbp* with body fat response and insulin levels after 8 weeks of HF/HS diet. BF, body fat; adip, adipose tissue; *r*, biweight midcorrelation; *p*, p value.

Table 1. Heritability estimates for gut microbiota in HMDP strains

Taxa	Heritability %
Rikenellaceae;Unknown	54
S24-7; Unknown	60
Lactococcus spp.	31
Turicibacter spp.	54
Clostridiaceae;Unknown	61
Lachnospiraceae;Unknown	56
Coprococcus spp	28
Roseburia spp.	33
Ruminococcus gnavus	48
Peptostreptococcaceae;Unknown	49
Ruminococcaceae;Unknown	39
Oscillospira spp.	53
Ruminococcus spp.	35
Mogibacteriaceae;Unknown	26
Erysipelotrichaceae;Unknown	65
Akkermansia muciniphila	54

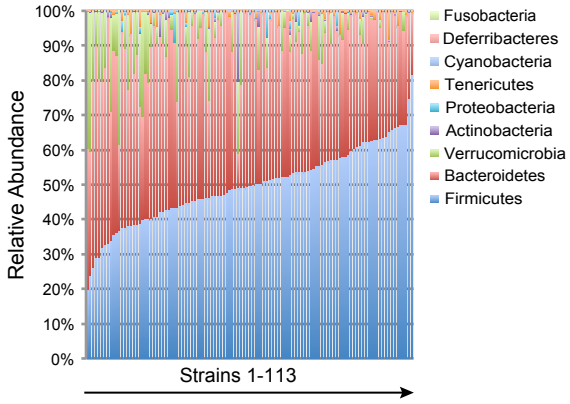
See also Table S3.

Table 2. Co-mapping of microbiota and clinical trait loci

Taxa	Chr	Clinical/ Metabolom QTL	Peak SNP	*P* value	cis eQTL with peak SNP in adipose and liver	*Peak SNP*	*P* value	Microbe-Transcript Correlations	Correlation	*P* value
Oscillospira spp.	4	Food intake	rs28116779	4.29E-07	*Caap1* *Ift74*	rs28133761	7.05E-07 1.83E-07	*Caap1*	-0.29	2.51E-03
Ruminococcus Gnavus	19				*Osbp*	rs30796836	2.54E-10	*Osbp*	-0.36	1.43E-04
Roseburia spp.	15	SubQ Fat	rs31730982	6.18E-07 2.03E-08	*Kif21a* *Lrrk2* *Irak4*	rs31843241	9.47E-13 7.06E-09[&] 5.77E-10	*Irak4*	0.32	1.93E-03
Turicibacter spp.	11				*Ccdc85a* *Efemp1*	rs29413813	2.24E-17 2.58E-22[#]			
Turicibacter spp.	9	Malonate/ N-acetyl-methionine	rs51650764 rs30331879	6.18E-07 2.03E-08	*Rbm5*	rs51650764	1.98E-15[#]			
Akkermansia muciniphila	7	TG Gonadal Fat	rs33129247	6.47E-09 7.44E-07						
Akkermansia muciniphila	2	Body weight 8 weeks	rs27307435	5.32E-07	*Ctnnbl1* *Rprd1b* *Lbp*	rs27323290 rs27323290 rs27334738	2.70E-07 6.31E-45 7.62E-08	*Ctnnbl1*	0.17	0.0743

[$] cis eQTL with global gene expression in epididymal adipose tissue in high fat/high sucrose diet (detailed data see Supplementary Table);[&] global gene expression in muscle,[#] global gene expression in liver. SubQ, subcutaneous fat; TG, triglycerides; *Caap1*, caspase activity and apoptosis inhibitor 1; *Ift74*, intraflagellar transport 74; *Osbp*, oxysterol binding protein; *Kif21a,* kinesin family member 21A; *Lrrk2*, Leucine-rich repeat kinase 2; I*rak4*, interleukin-1 receptor-associated kinase 4; *Ccdc85a*, coiled-coil domain containing 85A; *Efemp1*, epidermal growth factor-containing fibulin-like extracellular matrix protein 1; *Rbm5*, RNA Binding Motif Protein, *Ctnnbl1*, catenin, beta like 1; *Rprd1b,* regulation of nuclear pre-mRNA domain containing 1B; *Lbp,* lipopolysaccharide binding protein. See also Tables S6 and S7 and Figures S4 and S5.

A.



B.