Genetic and environmental control of host-gut microbiota interactions

Elin Org1, Brian W. Parks1, Jong Wha J Joo2, Benjamin Emert1, William Schwartzman1, Eun Yong Kang3, Margarete Mehrabian1, Calvin Pan4, Rob Knight5, Robert Gunsalus6, Thomas A. Drake7, Eleazar Eskin3,4 and Aldons J. Lusis1,4*

1Department of Medicine/Division of Cardiology, David Geffen School of Medicine, University of California; Los Angeles, CA, 90095, USA
2Bioinformatics IDP, University of California
3Department of Computer Science, University of California
4Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, 90095, USA
5Departments of Pediatrics and Computer Science and Engineering, University of California, San Diego, CA, 92093, USA
6Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, CA, 90095, USA
7Department of Pathology and Laboratory Medicine, University of California, Los Angeles, CA, 90095, USA
8Department 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 inhabitant 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 Lachnospiraceae, *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).

**Akermansia 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 $1.44 \times 10^9$ 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, *Bacteroidetes* 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 x 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 (Lachnospiraceae, Ruminococcaceae 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=3.9x10⁻¹⁰) (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, Lrk2 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 x 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 x 10⁻¹²; males: p = 2.54 x 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 x 10⁻⁶) and 2 (rs27323290, p = 6.67 x 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 x 10⁻⁶) and gonadal fat (p = 7.44 x 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 x 10^{-9}) and total weight after 8 weeks on HF/HS diet (p= 5.8 x 10^{-9}) (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
from the fine tuning of a symbiotic relationship that has co-evolved for millions of years. Among the likely contributing factors are differences in immunoglobulin and antibacterial molecules secreted into the gut lumen (Wen et al. 2008; Vijay-Kumar et al. 2010; Shulzhenko et al. 2011), differences in the mucosal gut structure (Sommer et al. 2014; Wlodarska et al. 2014), and differences in bile acid metabolism (Ryan et al. 2014).

Previous studies have shown that certain large effect mutations affecting inflammatory signaling or metabolic traits can significantly affect microbiota composition and that, in some cases, these can be transmitted by transplantation of gut microbiota from such mice (Henao-Mejia et al. 2012; Peng et al. 2014). Here, we have examined whether common variations in gut microbiota are also causally involved in metabolic traits. To test this possibility, we chose two inbred strains, AxB19 and SWR, that differ strikingly in the response to a HF/HS diet for cross-fostering studies. When male SWR mice cross-fostered by AxB19 dams were subjected to the HF/HS diet, they gained ~8% in total body fat as compared to ~2% for SWR, while AxB19 gained about 25% (Fig. 2C). Thus, while the majority of the response was dictated by the host genetic background, the gut microbiota did contribute significantly. Human gut microbiota exhibit more diversity and quantitative variation than that we have observed among common mouse strains, suggesting that a significant fraction of variance in obesity and insulin resistance in human populations is due to microbiota composition. Conceivably, this could explain some of the “missing heritability” observed in GWAS. We used correlation analysis to identify candidate microbiota contributing to the response to the HF/HS diet. Several genera were found to be strongly associated with traits such as body fat, plasma lipids, and insulin resistance, including some, such as Roseburia spp. and A. muciniphila, that have previously been implicated in metabolic traits. We chose A. muciniphila, a mucin-degrading, gram negative anaerobe, to examine its potential response to our dietary challenge. We administered by oral gavage either live or 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
1. Sample collection and study design

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

2. Sample Preparation and sequencing of 16S rRNA genes

Microbial DNA was extracted using the PowerSoil DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA). Amplification and sequencing of V4 hypervariable region of the 16S rRNA gene was performed using the validated, region-specific bacterial primers 515F and 806R according to previously described methods (Caporaso et al. 2012) optimized for the Illumina MiSeq platform. The reverse amplification primer contained a 12-bp Golay error-correcting barcode sequence and amplicons were generated in triplicate using 5 Prime Hot MasterMix (Fischer Scientific, UK). The PCR conditions consisted of an initial denaturation step of 94°C for 3 min; 35 cycles of 94°C for 45 sec, 50°C for 30 sec, and 72°C for 90 sec, followed by 72°C for 5 min. Replicate amplicons were quantified with Quant-iT™PicoGreen® dsDNA Assay Kit (Life Technologies Corporation, Carlsbad, CA, USA), pooled (200ng from 96 samples) and purified using with the UltraClean® PCR Clean-up Kit (MOBIO Laboratories, Inc., Carlsbad, CA, USA). High-throughput sequencing analysis of 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 $\sigma^2_g$ and $\sigma^2_e$, 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
\[
\frac{\sigma_i^2}{\sigma_j^2 + \sigma_e^2}
\]

and is computed for each relative abundance. We note that the kinship matrix must be standardized for these estimates to be consistent with the classical definition of heritability (Kostem and Eskin 2013; Speed and Balding 2015). A standardized kinship matrix has a mean along the diagonal of 1 and a sum of 0.

4. Clinical Traits

Body composition, food intake, and blood and plasma assays were as previously described (Parks et al. 2013; Parks et al. 2015). Briefly, mice were measured for total body fat mass and lean mass using magnetic resonance imaging (NMR) using the BurkerMinispec with software from Eco Medical Systems (Houston, TX). Blood was collected from mice following fasting for 4-5 hours and plasma was isolated by centrifugation in Microtainer tubes with EDTA (Becton, Dickinson and Company, Franklin Lakes, NJ). Plasma glucose levels were measured using a Beckman Glucose Analyzer 2 (Beckman Instruments). Plasma total cholesterol, HDL cholesterol, free cholesterol, triglycerides, and free fatty acid concentrations were determined by enzymatic assays employing colorimetric endpoints as described previously (Hedrick et al. 1993). Insulin levels and HOMA-IR were determined as previously described (Castellani et al. 2008).

5. Cross-fostering Study

Within 24-hours of birth, the pups from SWR females were removed from birthing cages and placed with AxB19/PgnJ mothers. Pups were weaned on postnatal day 21 and at 8 weeks of age mice were placed on a high-fat/high-sucrose diet (Research Diets D12266B) for additional 8 weeks. Controls from both strains were fostered with different mothers from the same strain.

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 7.2x10^10 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/HIS 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 x 10^-8, 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
both genders are posted on our website (systems.genetics.ucla.edu) with the link called 
“Download high-fat microbiota vs clinical trait correlation table” and can be also find in 
Supplemental Data (Supplemental Table S8).

Acknowledgements

We thank Hannah Qi, Zhiqiang Zhou, Judy Wu, Tieyan Han, and Richard Davis for 
expert assistance with mouse experiments, and Jonathan T. Furuta for growing A. 
muciniphila. This work was supported by National Institutes of Health (NIH) grants 
HL028481, HL30568, and DK094311 to A.J.L and UCLA Iris Cantor Women's Health Center 
and UCLA CTSI to T.A.D. E.O. was supported by a MOBILiTAS Postdoctoral Research Grant 
(MJD252) and FP7 grant no 330381. B.W.P. was supported by NIH training grant T32- 
HD07228.

Author Contributions: A.J.L, E.E and T.A.D supervised the study. R.K provided 
collected microbial data. R.G. prepared A. muciniphila for gavage experiment. E.O performed 
the analysis with contribution from W.JJ, E.Y.Y., C.P and R.K. E.O. A.J.L and E.E. prepared the 
manuscript with comments from B.E, T.A.D. and R.K.

Disclosure statement: The authors declare no conflict of interest


**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 x 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 *Ira4* gene in HMDP mice fed a HF/HS diet (B, C). Correlations of *Ira4* 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 x 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 mid correlation; $p$, p value.
Table 1. Heritability estimates for gut microbiota in HMDP strains

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

See also Table S3.
<table>
<thead>
<tr>
<th>Taxa</th>
<th>Chr</th>
<th>Clinical/ Metabolism QTL</th>
<th>Peak SNP</th>
<th>P value</th>
<th>cis eQTL with peak SNP in adipose and liver</th>
<th>Peak SNP</th>
<th>P value</th>
<th>Microbe-Transcript Correlations</th>
<th>Correlation</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oscillospira spp.</td>
<td>4</td>
<td>Food intake</td>
<td>rs28116779</td>
<td>4.29E-07</td>
<td>Caap1 Lft74</td>
<td>rs28133761</td>
<td>7.05E-07</td>
<td>1.83E-07</td>
<td>-0.29</td>
<td>2.51E-03</td>
</tr>
<tr>
<td>Ruminococcus Gnavus</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Caap1</td>
<td>-0.36</td>
<td>1.43E-04</td>
</tr>
<tr>
<td>Roseburia spp.</td>
<td>15</td>
<td>SubQ Fat</td>
<td>rs31730982</td>
<td>6.18E-07</td>
<td>2.03E-08 Kij21 a Lrrk2 Itrak4</td>
<td>rs31843241</td>
<td>9.47E-13</td>
<td>7.06E-09 &amp; 5.77E-10</td>
<td>0.32</td>
<td>1.93E-03</td>
</tr>
<tr>
<td>Turicibacter spp.</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td>Cccdc85a Efemp1</td>
<td>rs29413813</td>
<td>2.24E-17</td>
<td>2.58E-22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turicibacter spp.</td>
<td>9</td>
<td>Malonate/ N-acetyl- methionine</td>
<td>rs51650764 rs30331879</td>
<td>6.18E-07</td>
<td>2.03E-08 Rbm5</td>
<td>rs51650764</td>
<td>1.98E-15*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akkermansia muciniphila</td>
<td>7</td>
<td>TG Gonadal Fat</td>
<td>rs33129247</td>
<td>6.47E-09</td>
<td>7.44E-07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akkermansia muciniphila</td>
<td>2</td>
<td>Body weight 8 weeks</td>
<td>rs27307435</td>
<td>5.32E-07</td>
<td>Ctnnb1 Rprd1b Lbp</td>
<td>rs27323290 rs27323290 rs27334738</td>
<td>2.70E-07</td>
<td>6.31E-45 &amp; 7.62E-08</td>
<td>0.17</td>
<td>0.0743</td>
</tr>
</tbody>
</table>

* 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; Kij21a, kinesin family member 21A; Lrrk2, Leucine-rich repeat kinase 2; Itrak4, interleukin-1 receptor-associated kinase 4; Cccdc85a, coiled-coil domain containing 85A; Efemp1, epidermal growth factor-containing fibulin-like extracellular matrix protein 1; Rbm5, RNA Binding Motif Protein, Ctnnb1, 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. Relative Abundance of different bacterial classes across strains 1-113.

B. Box plot showing the distribution of unweighted UniFrac distances within and between strains, with significantly more different than similar.
B. Roseburia spp. eQTL with IRAK4 in liver

C. Roseburia spp. GWAS with IRAK4 expression

Chromosome 15

IRAK4 expression

HOMA-IR
**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

<table>
<thead>
<tr>
<th>Supplemental Material</th>
<th><a href="http://genome.cshlp.org/content/suppl/2015/08/17/gr.194118.115.DC1">http://genome.cshlp.org/content/suppl/2015/08/17/gr.194118.115.DC1</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Accepted Manuscript</td>
<td>Published online August 10, 2015 in advance of the print journal.</td>
</tr>
<tr>
<td>Open Access</td>
<td>Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.</td>
</tr>
<tr>
<td>Creative Commons License</td>
<td>Freely available online through the <em>Genome Research</em> Open Access option.</td>
</tr>
</tbody>
</table>

This manuscript is Open Access. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International license), as described at [http://creativecommons.org/licenses/by/4.0/](http://creativecommons.org/licenses/by/4.0/).

**Email Alerting Service**
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](http://genome.cshlp.org/subscriptions).

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to *Genome Research* go to: [http://genome.cshlp.org/subscriptions](http://genome.cshlp.org/subscriptions)