Title
Systems Genetic Analysis of Atherosclerosis and Gut Microbiota

Author
Kim, Myungsuk

Publication Date
2021

Peer reviewed|Thesis/dissertation

Systems Genetic Analysis of Atherosclerosis and Gut Microbiota

By

Myungsuk Kim

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In

Nutritional Biology

In the

OFFICE OF GRADUATE STUDIES

Of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Brian Bennett, Ph.D., Chair

John Newman, Ph.D.

Megan Dennis, Ph.D.

Angela Zivkovic, Ph.D.

Fawaz Haj, Ph.D.

Committee in Charge
2021

ABSTRACT

Systems Genetic Analysis of Atherosclerosis and Gut Microbiota

Atherosclerosis is a precipitating event in the development of cardiovascular disease (CVD). The progression of the disease is prevalent in developed countries and there are currently limited options for prevention and treatment interventions. Recent studies report that liver transcriptome and gut microbiota contributes to the pathogenesis of CVD, including metabolic syndrome. While host genetic variants are known factors that affect atherosclerosis development, liver transcriptome, and gut microbiota composition, the mechanisms underlying genetic variations are not yet clear. Here, we interrogated atherosclerosis regulatory networks in hyperlipidemic Diversity Outbred (DO) mice to reveal key insights into control of atherosclerosis using system genetic approaches of cardiometabolic traits, liver transcriptome, and microbiome. Global hepatic gene expression analysis showed that both atherogenic diet and host genetics have profound effects on the liver transcriptome in eight DO founder strains. These also include identifying sex-specific cardiometabolic traits, liver genetic pathways and networks, and abundance of fecal microbial taxa associated with atherosclerotic traits, defining the functionality of genes associated with the atherosclerotic traits and gut microbiota, and finding signatures of functional gene variants predicted to modulate those traits in the hyperlipidemic DO mice. Collectively, this study provides a rich resource for investigating the pathogenesis of atherosclerosis and suggests an opportunity to discover therapeutics and biomarkers in the setting of hyperlipidemia.

ACKNOWLEDGMENTS

First of all, I thank God for supporting me throughout my Ph.D. journey and for making me complete this journey with joy. Although I may not have been able to do the research I planned because of the sudden COVID19 pandemic, I believe that it is God who guides the path, even if a person plans the path with his heart.

Thank you to everyone who helped me in the Ph.D course and in completing this dissertation. First of all, I sincerely thank to Brian Bennett, my mentor and advisor, who has always supported me in completing my Ph.D. course. He's the role model I'm looking forward to be in the future, and he supported and inspired everything I need to do my project. It was the greatest luck to have the opportunity to learn from him. His philosophy and teachings really helped me grow as an independent researcher in the future. Thanks to John Newman, Megan Dennis, Angela Zivkovic, and Fawaz Haj, all members of the dissertation committee. Although COVID19 prevented me from meeting and getting advice in person, they were willing to offer suggestions and advice, and they encouraged me to grow as an independent researcher from the time I took the qualifying exam until the final dissertation was submitted.

Thanks also to all members of the Bennett lab. Bennetteers welcomed me, an international student from Korea, gave me suggestions and advice, helped me to focus on the biological meaning from a broader perspective, taught me the analytical skills, shared the code, and have been my advisors at every step. I will never forget everything I have learned from Bennett lab members.

In particular, I would like to thank Erik Gertz, Excel Que, Nazmul Huda, Kristen James, and Phoebe Yam for all the hard work and time spent together with the phenotypic and genetic

mapping of DO mice used in this study. I am grateful for their friendship, partnership, optimistic humor, and for elevating my mind and preventing loneliness. Thanks to Erik for patiently teaching me in phenotypic analysis and data QC. Thanks to Excel for teaching me basic statistical data analysis and genetic mapping research. Also, special thanks to Nazmul who taught me microbiota composition analysis and gave me valuable writing advice, and who worked with me in most of the research. What made me able to complete this study was their dedication and the numerous teachings that taught me about the analysis of the findings. And I sincerely thank Kristen, my GGNB colleague for sharing my Ph.D journey together. She is a lab mate that I admire, and with great leadership and wisdom, I can't believe she makes a big contribution to the advancement of the Bennett Lab. Especially thanks to my PhD senior, Phoebe. As a senior who went through course work and qualifying exams before me, she was of great help to me in achieving an important milestone in the doctoral program. Lastly, thanks to the UC Davis Genome Center for providing RNA-sequencing data.

I am deeply grateful to my dear parents who have always kept me amidst my most absurd things. Thank you for always asking my regards and trusting me and what I can achieve, even though we are thousands of kilometers away.

Lastly, I want to think of my beloved wife Soohyeon Lee. Even when I was busy with my research, you have always prayed and supported me, and helped me plan the next steps. She is the wonderful companion in my life, and none of my achievements would have been possible without her belief in my dreams.

Thanks to all of you and I look forward to the day when COVID19 ends, we will meet again in Davis.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER IV. Sexual Dimorphism of Atherosclerosis-associated Liver Transcriptome and Genetic Determinants in Hyperlipidemic Diversity Outbred-F1 Mice

CHAPTER V. Systems Genetic Analysis of Atherosclerosis and Gut Microbiota in a Hyperlipidemic Diversity Outbred F1 Mouse Population

LIST OF TABLES

CHAPTER IV. Sexual Dimorphism of Atherosclerosis-associated Liver Transcriptome and Genetic Determinants in Hyperlipidemic Diversity Outbred-F1 Mice

CHAPTER 1.

Introduction

1.1. Pathogenesis of Atherosclerosis

Cardiovascular disease (CVD) is a disease that affects the heart and blood vessels, and CVD is the leading cause of death annually than any other disease in the United States (Pagidipati and Gaziano, 2013). In the United States and EU, CVD accounts for 33-40% of mortality rates across all ages, with total economic losses in 2008 of $277.7 billion and €195 billion, respectively (Virani et al., 2020). In addition, CVD has been estimated to be the leading cause of morbidity worldwide although the standards for recording the cause of death vary from country to country (Turk-Adawi and Grace, 2015; Virani et al., 2020). Atherosclerosis is the dominant cause of CVD including myocardial infarction, heart failure, and stroke, and accounts for 25-30% of cardiovascular deaths (Frostegard, 2013; Gimbrone et al., 2000). Atherosclerosis can also occur asymptomatic in young individuals (Berenson et al., 1998; Tracy et al., 1995). Therefore, it is necessary to identify non-invasive diagnostic biomarkers to prevent the progression of atherosclerosis.

Mechanistically, atherosclerosis is triggered by dysfunction in the endothelial cells surrounding the blood vessel wall by stimulation of oxidized low-density lipoprotein (OxLDL) (Camejo et al., 1980). In the early stages of atherosclerosis, LDL penetrates the lining, binds to the proteoglycan matrix, and is oxidized. Ultimately, OxLDL deposited inside the vessel wall leads to macrophage recruitment and aggregation (Frostegård et al., 1990). When macrophages

swallow accumulated OxLDL, they increase in size and become foam cells filled with lipids. Accordingly, the inflammatory pathways of macrophages are further activated, increasing oxidative stress and cytokine/chemokine secretion, leading to more LDL, activation of endothelial cells, monocyte recruitment, and foam cell formation. This type of atherosclerotic plaque is called fatty streaks and macrophage chemoattractants stimulate the penetration and proliferation of smooth muscle cells. As the concentration of OxLDL continues to increase in the intima and more foam cells are produced in advanced lesions, migration of smooth muscle cells to the intima and the production of collagen are promoted, and fibrous caps that can block the lumen of blood vessels and limit blood flow cover the growing plaque (Lusis, 2000). The stress sheered by the bloodstream can rupture these stable fibrous plaques and completely limit blood flow by thrombus formation, resulting in heart failure or stroke (Virmani et al., 2002).

In addition to OxLDL, lysophosphatidylcholine, a major phospholipid in atherosclerotic lesions, have immune-stimulatory effects like OxLDL and contribute to the development of atherosclerosis (Huang et al., 1999). Other mechanisms that induce atherosclerosis include malondialdehyde produced during the oxidation of LDL, oxidized apolipoprotein B and cholesterol, and interaction with scavenger receptors and toll-like receptors (Greig et al., 2012; Miller et al., 2011).

1.2. Complexity of Atherosclerosis Risk Factors

1.2.1. Atherosclerosis and Genetic Factors

Atherosclerosis is a complex disease caused by the interaction of environmental and genetic factors. Early studies in humans using a forward genetics approach in families have reported hereditary dominant genotype for hyperlipidemia, the most common of which is hypercholesterolemia, which causes high cholesterol levels in serum and induce atherosclerosis

development. (Buja et al., 1979; Jensen and Blankenhorn, 1972). A follow-up large family-based study was conducted to investigate the genetic factors of hyperlipidemia associated with CVD, which helped successfully identify several genes that regulate plasma cholesterol and triglycerides, including CETP (cholesteryl ester transfer protein), LPL (lipoprotein lipase), and LDLR (low-density lipoprotein receptor) (Goldstein et al., 1973; Yang et al., 1995). Subsequent experimental analysis revealed that these genes are associated with impaired regulation of lipid levels, which have a major impact on the progression of atherosclerosis (Masucci-Magoulas et al., 1997). Collectively, hyperlipidemia has been found to contribute to the pathogenesis of atherosclerosis along with elevated total cholesterol and LDL, and these results have suggested specific guidelines for reducing the risk of CVD in individuals based on lipid level control. (Gotto and Moon, 2012; Grundy et al., 2004).

The Human Genome Project has uncovered millions of common single nucleotide polymorphisms (SNPs) where single SNP varies from individual to individual and occurs with >5% minor allele frequency (Collins et al., 2003). Since then, many human genome-wide association studies (GWAS) have been conducted to identify genes and genetic variants associated with disease susceptibility, including atherosclerosis (Lusis, 2012). The primary purpose of GWAS is to pinpoint genomic locations associated with increased risk of disease by associating common SNPs with disease states. In addition, GWAS is well suited for identifying novel genes associated with atherosclerosis, which can lead to new experimental hypotheses. However, SNPs located at the loci identified by human GWAS show a high correlation with each other, it is difficult to secure evidence that the discovered SNPs and genes are the causative ofdisease (Flint and Mott, 2001). For this reason, examples of therapeutic agents developed from GWAS are very limited, except for some treatments for type 2 diabetes (ZnT-8 antagonists),

autoimmune disease (targeting components of the IL-23 pathway), and schizophrenia (targeting dopamine and glutamate) (Visscher et al., 2017).

In order to overcome these limitations, integrative multi-omics approaches have been proposed that comprehensively analyzes data such as transcriptome, epigenome, metabolome, metagenome, as well as genomic information, which has been applied to the discovery of a number of disease-related biomarkers (Rau et al., 2020; Seldin et al., 2019). Many countries have recognized precision medicine centering on multi-omics information and medical information as a strategic field for the future of the 4th industrial revolution.

1.2.2. Atherosclerosis and Diets

Diet affects the development of atherosclerosis in humans and rodent models, and dietary components that increase or decrease the risk of atherosclerosis have been studied and documented for decades. After the institutionalization of the use of laboratory mice in the early 1900s, diet-induced atherosclerotic mouse models were developed. The first diet associations of diet with disease was reported for high-fat diet with cholic acid leading to aortic lesions in mice (Ishida et al., 1991). Mice fed normal diets seldom develop atherosclerotic lesions likely because they do not express the cholesteryl ester transfer protein (CETP) gene, a potential target for the prevention of atherosclerosis in humans. In addition, they always maintain a high level of high-density lipoprotein (HDL) in plasma, which may contribute to relative atheroresistance even in atherosusceptible mouse strains (Morrisett et al., 1982; Westerterp et al., 2006). The mouse atherosclerosis model therefore depend upon inducing non-HDL-based hypercholesterolemia, which is most readily achieved by genetic ablation of apoE or LDL receptors (LDLRs) (Getz and Reardon, 2012).

Similar to the inter-individual variation of atherosclerosis observed in humans, these initial studies confirmed differences in atherosclerosis development across inbred mouse strains (LeBoeuf et al., 1983; Lusis et al., 1983; Morrisett et al., 1982). These early studies of atherosclerosis in mice led to side effects that often-increased mortality and morbidity, including the development of fatty liver. A study by Paigen in 1985 developed an atherogenic diet with lower fat and cholesterol levels than the atherogenic diet from Roberts and Thompson (Roberts and Thompson, 1976), which induced aortic lesions in mice without increasing mortality (Paigen et al., 1985a). In addition, strain-specific differences for plasma total cholesterol and atherosclerosis in this diet were identified in more than 10 inbred strains, and some strains, including A/J, BALB/cJ, and C3H/J, did not show diet-induced aortic lesion development (Paigen et al., 1985b). Taken together, these results indicate that atherosclerosis susceptibility may vary depending on the genetic backgrounds of the inbred strain.

Dietary composition studies for the atherosclerosis mouse model have made a major contribution to optimizing the diet for the study of atherosclerosis. In 1990, composition of a low and high-fat refined synthetic diet was proposed by the Paigen group (Nishina et al., 1990). This synthetic diet ("Paigen diet") reduces the potential influence of variable components in the standard diet on lesion development. This diet was also found to induce aortic lesion development while simultaneously reducing pathological burdens of increased gall stones, liver damage, and fatty liver (Nishina et al., 1990).

In addition to the synthetic diet that induces atherosclerosis, dietary components that reduce CVD including atherosclerosis have been reported in many studies. Epidemiological, clinical, and experimental studies have demonstrated that diet and nutrition play a central role in the prevention of CVD (Torres et al., 2015). Based on these data the American Heart Association

has made formal dietary recommendations (https://www.heart.org/en/healthy-living/healthy-eating/eat-smart/nutrition-basics/aha-diet-and-lifestyle-recommendations). One of the most studied diets in atherosclerosis research is the Mediterranean diet. Epidemiological studies found that the Mediterranean diet has a cardio-protective effect compared to western diets (De Lorgeril et al., 1994; Estruch et al., 2013). In addition, large population-based observational studies (Bazzano et al., 2003) found that a diet containing high dietary fiber was associated with a reduced CVD risk, which was largely mediated via the reduction of LDL-cholesterol. Clinical trials found that ingestion of soluble fiber (2-10 g/day) was associated with a significant 7% LDL-cholesterol reduction (Brown et al., 1999) in a dose-dependent manner (Anderson et al., 2000). Growing evidence shows that prebiotics supplementation reduces CVD, specifically blood pressure and atherosclerosis, by manipulating the gut microbiota, which supports prebiotic interventions to prevent or treat CVD. Prebiotics are the non-digestible food ingredients that beneficially affect the host by selectively stimulating the growth and/or the activity of one or a limited number of bacterial species already resident in the colon (Gibson and Roberfroid, 1995). Inulin, a linear β-2,1 fructosyl-fructose polydisperse carbohydrate material, feeding decreased atherosclerosis in the aortic root of mice and normalized the altered microbial abundance of *Bifidobacterium, Lactobacillus, Akkermansia, Allobaculum,* and *Coprococcus* (Hoffman et al., 2020). β-glucan is a glucose polysaccharide that can sequester cholesterol, scavenges reactive oxygen species, and produces short-chain fatty acids when digested by gut microbiota (Nakashima et al., 2018). In addition to serving as an energy source for gut bacteria, β-glucans are immunostimulatory through activation of β-glucan receptors, such as dectin-1 or CR3 on the intestinal macrophages (Brown and Gordon, 2001; Chan et al., 2009). Oat β-glucan

supplementation increased HDL-cholesterol and decreased plasma triglyceride (TG) and atherosclerosis alone with enrichment of the genus *Akkermansia* in the gut (Ryan et al., 2017).

1.2.3. Atherosclerosis and Sex

Despite decades of advances in medical and cardiovascular care, CVD has been reported as the leading cause of death in men and women in most ethnic groups. A large number of studies have demonstrated sex differences in risk of atherosclerotic CVD. Sexual dimorphism in atherosclerosis susceptibility has been reported in humans and animals (Bennett et al., 2015; Bubb et al., 2012; Isensee et al., 2008; Karp et al., 2017; Mittelstrass et al., 2011; Yang et al., 2006). Men have a higher prevalence of stroke than women by age 70, but women have the highest prevalence of stroke in the elderly population (Virani et al., 2020). Women under the age of 55 have a significantly lower risk of hypertension than men in their age group, while women over 75 have a higher risk of hypertension than men in their age group (Fryar et al., 2017). In animal models, sexual dimorphism of atherosclerosis has been observed in $ApoE^{-/-}$ and $Ldlr^{-/-}$ mice, pigs, and rabbits (Fisher et al., 1967; Freeman et al., 2007; Matthan et al., 2018; Van Craeyveld et al., 2010). Previous studies have also found sexual dimorphism of cardiometabolic traits in F2 genetic crosses or >100 different inbred mouse strains (Bennett et al., 2015; Kayashima et al., 2014; Norheim et al., 2019; Su et al., 2006). These results suggest fundamental biological differences between sexes and underscore the need to include sex as an important component of the investigation in animal models and clinical trials.

Mechanistically, sex hormones play an important role in regulating gene expression related to metabolic disorders including CVD risk (AlSiraj et al., 2019; Kukurba et al., 2016; Mozhui et al., 2012; Yang et al., 2006). Besides sex hormones, sexual dimorphism also arises due to effects of sex chromosomes (epigenetic effects, genetic dosage, and different dosage

compensation) (Arnold and Lusis, 2012; Charlesworth, 1996; Disteche, 2012; Hager et al., 2008; McCarthy et al., 2009; Winham et al., 2015). For instance, several studies suggest that the X and Y chromosome play a crucial role in atherosclerosis (AlSiraj et al., 2019; Eales et al., 2019). Interestingly, this sexual dimorphism was also observed in the genetic mapping of the mouse model of atherosclerosis. In the quantitative trait loci (QTL) mapping performed in the F2 cross between the hyperlipidemic C57BL/6J strain and C3H/HeJ or FVB/N strain, most of the loci identified showed significant sex dependence (Teupser et al., 2006; Wang et al., 2007). Thus, these results imply that therapeutic effects to atherosclerosis may not work equally well in both sexes, and a personalized approach is needed that takes into account sex.

1.2.4. Atherosclerosis and Gut Microbiota

For over 50 years, the potential role of infectious microorganisms, including bacteria and viruses, as potential risk factors for atherosclerosis has been appreciated through a number of epidemiological studies (Cluff et al., 1968; Pankey, 1965; Saikku et al., 1988). Infection and the subsequent inflammatory processes are thought to induce the onset, progression, and rupture of atherosclerotic plaques (Libby et al., 2009). Mechanisms by which infection aggravates atherosclerosis include direct cell invasion that accelerates plaque growth through local effects, or indirect systemic production of inflammatory cytokines that promote the development of atherosclerosis (Pant et al., 2014). Representative infectious agents associated with atherosclerosis include *Chlamydia pneumonia*, *Helicobacter pylori*, *P. gingivalis*, and Influenza A virus (Pothineni et al., 2017).

Recent studies have revealed that the gut microbiota is an emerging contributor to human physiology and also affects the cardiovascular system (Brown and Hazen, 2018). Studies of human cohorts and animal models suggest that alteration of gut microbial diversity influences

the CVD development associated with risk factors such as atherosclerosis (Emoto et al., 2016), blood lipids (Fu et al., 2015), hypertension (Li et al., 2017), and heart failure (Luedde et al., 2017). In a number of case-control studies, distinct shifts of microbial composition have been identified in fecal samples from patients with atherosclerosis. It is not clear whether these microbial communities are causal to atherosclerosis or if they are simply affected by other environmental factors contributing to the development of atherosclerosis. Frequently, studies utilize a case-control design to identify the microbial differences among subjects with atherosclerosis compared to controls. Foremost, there have been several consistent results reported from a number of cohorts (Dinakaran et al., 2014; Emoto et al., 2016; Kelly et al., 2016). For example, *Strepcococcus* spp., exhibiting pathogenic expansion in intestinal dysbiosis (Taur and Pamer, 2013), showed increased abundance in stool samples from atherosclerosis patients (Feng et al., 2016; Jie et al., 2017; Sanchez-Alcoholado et al., 2017; Yan et al., 2017). In addition, the abundance of *Faecalibacterium* spp. and *Roseburia* spp., which are known to have anti-inflammatory effects and reduce atherosclerotic events in mice and humans (Kasahara et al., 2018; Sokol et al., 2008), were relatively depleted in atherosclerosis patients (Jie et al., 2017; Sanchez-Alcoholado et al., 2017). Finally, metagenome studies have shown depletion of butyrate-producing bacteria including *Faecalibacterium prausnitzii* and *Roseburia intestinalis*, and a high abundance of *Enterobacteriaceae* and *Streptococcus* in atherosclerotic patients (Jie et al., 2017). Studies have also demonstrated enrichment of genes encoding peptidoglycan biosynthesis in patients with atherosclerosis, while healthy cohorts are enriched in phytoene dehydrogenase genes (Karlsson et al., 2012). Thus, not only is the composition of the microbiota altered in atherosclerosis but the underlying functional capacity of the microbiota may be impacted as well.

Although the microbiota is complex, there are a number of approaches to establishing causality and infer mechanism(s) of microbiota and specific bacteria for atherosclerosis. In particular, studies utilizing microbial transplantation in a model organism can demonstrate the direct role of the gut microbial taxa in the risk of atherosclerosis. Studies utilizing germ-free animals have been critical to further our understanding of the functional role of the microbiota in disease susceptibility (Kennedy et al., 2018). Germ-free mice can also receive fecal microbiota transfers from samples collected in case/control studies. For example, studies have been performed using two strains of mice that differ in atherosclerosis susceptibility and that at least a portion of atherosclerosis can be attributed to the microbiota (Gregory et al., 2015). In addition to studies examining the transplantation of fecal samples, there have been attempts at understanding the role of specific bacteria in atherosclerosis. For example, colonization of germ-free ApoE-deficient ($ApoE^{-/-}$) mice with *Roseburia intestinalis* isolated from humans reduced the levels of inflammatory markers and atherosclerosis providing evidence of the causative role of *Roseburia intestinalis* in atherosclerosis (Kasahara et al., 2018). These results suggest that the role of the gut microbiota in mediating atherosclerosis should be considered in addition to the traditional host-focused approach in the development of therapeutic agents for atherosclerosis.

1.3. Advancement of Mouse Models for Genetic Mapping of Atherosclerosis

A forward genetics approach utilizing genetic mapping is to pinpoint genomic loci by associating SNPs with the trait of interest. This approach can identify the causative gene and genetic variants associated with the target trait by thoroughly examining the identified loci (Rao, 2001). It also has been particularly successful to identify genes that regulate rare Mendelian disorders. Based on these results, researchers have conducted GWAS to find loci associated with a variety of disease-related traits such as atherosclerosis, obesity, diabetes, and metabolic

syndrome using the genotyping data available for SNPs. However, these complex traits are inherently multi-genic and the contribution of environmental factors is also large (Belmont and Leal, 2005; Lander and Schork, 1994). In order to identify the significant QTL for complex trait such as atherosclerosis, researchers minimized environmental factors by using a mouse model with naturally occurring alleles and performed genetic mapping of quantitative phenotypes in a large number of mice with a high-density genotyping array (Smith, 2003).

Genetic mapping studies in mice have so far been successful in identifying at least 30 atherosclerotic loci (Chen et al., 2007; Lusis et al., 2016; Smallwood et al., 2014; Wang et al., 2007). For example, in early QTL mapping studies, a number of atherosclerotic QTLs were found in the F2 cross between atherosclerosis susceptible C57BL/6J strain and atherosclerosis resistant C3H/HeJ strain or CAST/EiJ strain (Mehrabian et al., 2001; Wang et al., 2007). These include seven F2 crosses of different inbred strains using $ApoE^{-/-}$ or $Ldlr^{-/-}$ mice to induce severe atherosclerosis. However, since most of these loci are large and contain hundreds of genes, it has been difficult to identify the gene that causes atherosclerosis.

To narrow in at high resolution locus mapping associated traits of interest in mouse models, studies have begun using strategies to generate more genetically heterogeneous mouse populations. A population of heterologous stock (HS) mice consisting of eight different inbred founder strains (A/J, AKR/J, BALBc/J, CBA/J, C3H/HeJ, C57BL/6J, DBA/2J and LP/J) was developed for fine genetic mapping of phenotypes (Solberg et al., 2006). These mice were used to identify hundreds of loci associated with disease-related traits including asthma, type 2 diabetes, obesity, and anxiety (Valdar et al., 2006).

In addition, the Hybrid Mouse Diversity Panel (HDMP), consisting of ~100 inbred strains, improves the resolution of genetic mapping of complex traits including atherosclerosis,

heart failure, obesity, diabetes, fatty liver, osteoporosis (Bennett et al., 2010; Ghazalpour et al., 2012; Lusis et al., 2016). The genetic mapping study of atherosclerosis using HMDP mice was performed by crossing HDMP mice with C57BL/6J mice carrying transgenes of human CETP and apolipoprotein E-Leiden, and 6 fine QTLs for aortic lesion area were identified on chromosomes 2, 5, and 9 (Bennett et al., 2015).

In 2004, an innovative plan to develop a population of recombinant inbred strains consisting of five laboratory inbred strains and three wild-derived strains was initiated to improve the resolution of genetic mapping in mice (Churchill et al., 2004). The Collaborative Cross (CC) mouse population is the construction of recombinant inbred lines using a funnel crossover strategy through backcross of mice with eight founder strains including A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, and NZO/HiLtJ with inclusion of wild-derived strains CAST/EiJ, PWK/PhJ, and WSB/EiJ. CC mice have increased genetic diversity due to the uniform distribution of genetic variation and a beneficial population for genome-wide association study (Aylor et al., 2011; Bogue et al., 2015).

1.4. Integrative Systems Genetic Analysis using the Diversity Outbred Mouse Model

Diversity outbred (DO) mice are a population derived by intercrossing CC mice that were intentionally outbred to maximize the number of recombination events per mouse that allow fine-resolution mapping with narrower chromosomal spacing (Churchill et al., 2012). Alleles present in DO mice are from five laboratory-derived inbred strains and three wild-derived inbred strains, representing ~90% of the known variations present in laboratory mice, and these genetic variations are distributed throughout the genome of DO mice (Churchill et al., 2012). The allelic contribution from each of the eight founder strains is about 12.5%, and these mice have a level of heterozygosity that is closer to that found in humans. DO mice breed randomly for 3-4

generations each year and are maintained genetically unique (Churchill et al., 2012). DO mice are superior to other mapping populations in that they have more than 40 million SNPs with a high average minor allele frequency and a fine recombination block structure (Churchill et al., 2012; Gatti et al., 2014). Therefore, since disease or trait-related genetic variants can be detected at a much higher resolution than human GWAS, which can only detect SNPs with > 5% a minor allele frequency, DO mice can be a valuable resource for the discovery of disease-related targets. Furthermore, researchers can generate disease-related traits from metabolic organs (liver, muscle, kidney, etc.) that are difficult to obtain from human samples, and increase the heritability of disease-related traits by carefully controlling environmental factors that act as confounding factors.

To date, genetic mapping using DO mice has been performed in various diseases such as atherosclerosis (Smallwood et al., 2014), diabetes (Keller et al., 2019), prostate cancer (Winter et al., 2017), pancreatic cancer (Yang et al., 2019), melanoma cancer (Ferguson et al., 2019), tuberculosis (Tavolara et al., 2020), respiratory depression (Bubier et al., 2020), kidney disease (Huda et al., 2020), and liver toxicity (Kurtz et al., 2020). In addition to diseases, DO mice have been used to investigate genetic factors for toxic chemicals (French et al., 2015; Recla et al., 2019), skeletal development (Katz et al., 2020), sleep-related phenotypes (Keenan et al., 2020), and gut microbiota (Carmody et al., 2015; Kemis et al., 2019). Furthermore, recent studies using DO mice have uncovered and validated disease-related biomarkers via genetic mapping using integrative multi-omics data such as transcriptome, proteome, metabolome, and microbiome with disease-related phenotypes (Chick et al., 2016; Kemis et al., 2019; Linke et al., 2020). This approach enables the development of an advanced system genetics analysis platform by

investigating whether SNPs, which are associated with the disease phenotype, are co-localized at the same loci with the SNPs associated with the multi-omics data.

1.5. Significance of Findings in Dissertation

In Chapter 2, we examine the associations between gut microbiota and metabolic diseases such as CVD and type 2 diabetes. We discuss three potential mechanisms including gut permeability and endotoxemia, increased immune system activation, and microbial-derived metabolites. In addition to discussing these potential mechanisms, we highlight current studies manipulating the gut microbiota or microbial metabolites to move beyond sequenced based association studies. Finally, we discuss approaches to demonstrate causalities such as specific diet changes, inhibition of microbial pathways, and fecal microbiota transplantation.

In Chapter 3, we investigated the diet- and strain-dependent effects on metabolic traits in the eight Collaborative Cross founder strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HILtJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ). Liver transcriptomic analysis showed that both an atherogenic diet and host genetics have profound effects on the liver transcriptome, which may be related to differences in metabolic traits observed between strains. We found strain differences in circulating trimethylamine N-Oxide (TMAO) concentration and liver TG content, both of which are traits associated with metabolic diseases. Using a network approach, we identified a module of transcripts associated with TMAO and liver TG content which was enriched in functional pathways. Interrogation of the module related to metabolic traits identified NADPH oxidase 4 (*Nox4),* a gene for a key enzyme in the production of reactive oxygen species, which showed a strong association with plasma TMAO and liver triglyceride. Interestingly, *Nox4* was identified as the highest expressed in the C57BL/6J and NZO/HILtJ strains and the lowest expressed in the CAST/EiJ strain.

In Chapter 4, we studied sexual dimorphism in 20 cardiometabolic traits, liver gene expression, and genetic effect using eight hyperlipidemic inbred founder strains and Diversity Outbred-F1 mice to examine the role of sex and gene-by-sex interactions in atherosclerosis. Our results demonstrate tremendous effects of sex on cardiometabolic traits and hepatic gene expression regardless. We identified sex-specific liver genetic pathways and networks involved in cardiometabolic traits, suggesting that certain liver gene clusters may contribute to sex-specific atherosclerosis susceptibility. In support of this, genetic loci associated with the traits and transcripts frequently showed sex specificity. Genes highly correlated with the aortic lesion area were enriched in pathways relevant to atherosclerosis such as cholesterol homeostasis and cytokine-mediated signaling pathways in a sex-specific manner. Furthermore, the *Ptprk* gene in females and the *Pten* gene in males were shown to underlie a sex-specific locus for atherosclerosis.

In Chapter 5, we interrogated atherosclerosis regulatory networks in hyperlipidemic Diversity Outbred mice to reveal key insights into control of atherosclerosis using system genetic approaches of cardio-metabolic traits and microbiome. These include identifying abundance of fecal microbial taxa and functional predicted pathways associated with atherosclerotic traits and signatures of functional gene variants predicted to modulate those traits. Trans-omic analyses facilitated identification of *Lactococcus* as a previously unknown regulatory taxon for atherosclerotic traits.

1.6. References

Siraj, Y., Chen, X., Thatcher, S.E., Temel, R.E., Cai, L., Blalock, E., Katz, W., Ali, H.M., Petriello, M., Deng, P., *et al.* (2019). XX sex chromosome complement promotes atherosclerosis in mice. Nat Commun *10*, 2631.

Anderson, J.W., Allgood, L.D., Lawrence, A., Altringer, L.A., Jerdack, G.R., Hengehold, D.A., and Morel, J.G. (2000). Cholesterol-lowering effects of psyllium intake adjunctive to diet therapy in men and women with hypercholesterolemia: meta-analysis of 8 controlled trials. The American journal of clinical nutrition *71*, 472-479.

Arnold, A.P., and Lusis, A.J. (2012). Understanding the sexome: measuring and reporting sex differences in gene systems. Endocrinology *153*, 2551-2555.

Aylor, D.L., Valdar, W., Foulds-Mathes, W., Buus, R.J., Verdugo, R.A., Baric, R.S., Ferris, M.T., Frelinger, J.A., Heise, M., Frieman, M.B., *et al.* (2011). Genetic analysis of complex traits in the emerging Collaborative Cross. Genome Res *21*, 1213-1222.

Bazzano, L.A., He, J., Ogden, L.G., Loria, C.M., and Whelton, P.K. (2003). Dietary fiber intake and reduced risk of coronary heart disease in US men and women: the National Health and Nutrition Examination Survey I Epidemiologic Follow-up Study. Archives of internal medicine *163*, 1897-1904.

Belmont, J.W., and Leal, S.M. (2005). Complex phenotypes and complex genetics - an introduction to genetic studies of complex traits. Curr Atheroscler Rep *7*, 180-187.

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

Bennett, B.J., Farber, C.R., Orozco, L., Kang, H.M., Ghazalpour, A., Siemers, N., Neubauer, M., Neuhaus, I., Yordanova, R., Guan, B., *et al.* (2010). A high-resolution association mapping panel for the dissection of complex traits in mice. Genome Res *20*, 281-290.

Berenson, G.S., Srinivasan, S.R., and Nicklas, T.A. (1998). Atherosclerosis a nutritional disease of childhood. Am J Cardiol *82*, 22T-29T.

Bogue, M.A., Churchill, G.A., and Chesler, E.J. (2015). Collaborative Cross and Diversity Outbred data resources in the Mouse Phenome Database. Mamm Genome *26*, 511-520.

Brown, G.D., and Gordon, S. (2001). A new receptor for β-glucans. Nature *413*, 36-37.

Brown, J.M., and Hazen, S.L. (2018). Microbial modulation of cardiovascular disease. Nat Rev Microbiol *16*, 171-181.

Brown, L., Rosner, B., Willett, W.W., and Sacks, F.M. (1999). Cholesterol-lowering effects of dietary fiber: a meta-analysis. The American journal of clinical nutrition *69*, 30-42.

Bubb, K.J., Khambata, R.S., and Ahluwalia, A. (2012). Sexual dimorphism in rodent models of hypertension and atherosclerosis. Br J Pharmacol *167*, 298-312.

Bubier, J.A., He, H., Philip, V.M., Roy, T., Hernandez, C.M., Bernat, R., Donohue, K.D., O'Hara, B.F., and Chesler, E.J. (2020). Genetic variation regulates opioid-induced respiratory depression in mice. Sci Rep *10*, 14970.

Buja, L.M., Kovanen, P.T., and Bilheimer, D.W. (1979). Cellular pathology of homozygous familial hypercholesterolemia. Am J Pathol *97*, 327-357.

Camejo, G., Lalaguna, G., López, F., and Starosta, R. (1980). Characterization and properties of a lipoprotein-complexing proteoglycan from human aorta. Atherosclerosis *35*, 307-320.

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

Chan, G.C.-F., Chan, W.K., and Sze, D.M.-Y. (2009). The effects of β-glucan on human immune and cancer cells. Journal of hematology & oncology *2*, 1-11.

Charlesworth, B. (1996). The evolution of chromosomal sex determination and dosage compensation. Curr Biol *6*, 149-162.

Chen, Y., Rollins, J., Paigen, B., and Wang, X. (2007). Genetic and genomic insights into the molecular basis of atherosclerosis. Cell Metab *6*, 164-179.

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

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

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

Cluff, L.E., Reynolds, R.C., Page, D.L., and Breckenridge, J.L. (1968). Staphylococcal Bacteremia and Altered Host Resistance. Ann Intern Med *69*, 859-873.

Collins, F.S., Morgan, M., and Patrinos, A. (2003). The Human Genome Project - lessons from large-scale biology. Science *300*, 286-290.

De Lorgeril, M., Renaud, S., Salen, P., Monjaud, I., Mamelle, N., Martin, J., Guidollet, J., Touboul, P., and Delaye, J. (1994). Mediterranean alpha-linolenic acid-rich diet in secondary prevention of coronary heart disease. The Lancet *343*, 1454-1459.

Dinakaran, V., Rathinavel, A., Pushpanathan, M., Sivakumar, R., Gunasekaran, P., and Rajendhran, J. (2014). Elevated levels of circulating DNA in cardiovascular disease patients: metagenomic profiling of microbiome in the circulation. PLoS One *9*, e105221.

Disteche, C.M. (2012). Dosage Compensation of the Sex Chromosomes. Annual Review of Genetics *46*, 537-560.

Eales, J.M., Maan, A.A., Xu, X., Michoel, T., Hallast, P., Batini, C., Zadik, D., Prestes, P.R., Molina, E., Denniff, M., *et al.* (2019). Human Y Chromosome Exerts Pleiotropic Effects on Susceptibility to Atherosclerosis. Arterioscler Thromb Vasc Biol *39*, 2386-2401.

Emoto, T., Yamashita, T., Sasaki, N., Hirota, Y., Hayashi, T., So, A., Kasahara, K., Yodoi, K., Matsumoto, T., Mizoguchi, T., *et al.* (2016). Analysis of Gut Microbiota in Coronary Artery Disease Patients: a Possible Link between Gut Microbiota and Coronary Artery Disease. J Atheroscler Thromb *23*, 908-921.

Estruch, R., Ros, E., Salas-Salvadó, J., Covas, M.-I., Corella, D., Arós, F., Gómez-Gracia, E., Ruiz-Gutiérrez, V., Fiol, M., and Lapetra, J. (2013). Primary prevention of cardiovascular disease with a Mediterranean diet. New England Journal of Medicine *368*, 1279-1290.

Feng, Q., Liu, Z., Zhong, S., Li, R., Xia, H., Jie, Z., Wen, B., Chen, X., Yan, W., Fan, Y., *et al.* (2016). Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease. Sci Rep *6*, 22525.

Ferguson, B., Handoko, H.Y., Mukhopadhyay, P., Chitsazan, A., Balmer, L., Morahan, G., and Walker, G.J. (2019). Different genetic mechanisms mediate spontaneous versus UVR-induced malignant melanoma. Elife *8*, e42424.

Fisher, H., Griminger, P., and Siller, W.G. (1967). Effect of pectin on atherosclerosis in the cholesterol-fed rabbit. J Atheroscler Res *7*, 381-386.
Flint, J., and Mott, R. (2001). Finding the molecular basis of quantitative traits - successes and pitfalls. Nat Rev Genet *2*, 437-445.
Freeman, L., Amar, M.J., Shamburek, R., Paigen, B., Brewer, H.B., Jr., Santamarina-Fojo, S., and Gonzalez-Navarro, H. (2007). Lipolytic and ligand-binding functions of hepatic lipase protect against atherosclerosis in LDL receptor-deficient mice. J Lipid Res *48*, 104-113.
French, J.E., Gatti, D.M., Morgan, D.L., Kissling, G.E., Shockley, K.R., Knudsen, G.A., Shepard, K.G., Price, H.C., King, D., Witt, K.L., *et al.* (2015). Diversity Outbred Mice Identify Population-Based Exposure Thresholds and Genetic Factors that Influence Benzene-Induced Genotoxicity. Environ Health Perspect *123*, 237-245.
Frostegard, J. (2013). Immunity, atherosclerosis and cardiovascular disease. BMC Med *11*, 117.
Frostegård, J., Nilsson, J., Haegerstrand, A., Hamsten, A., Wigzell, H., and Gidlund, M. (1990). Oxidized low density lipoprotein induces differentiation and adhesion of human monocytes and the monocytic cell line U937. Proc Natl Acad Sci U S A *87*, 904-908.
Fryar, C.D., Ostchega, Y., Hales, C.M., Zhang, G., and Kruszon-Moran, D. (2017). Hypertension Prevalence and Control Among Adults - United States, 2015-2016. NCHS Data Brief *289*, 1-8.
Fu, J., Bonder, M.J., Cenit, M.C., Tigchelaar, E.F., Maatman, A., Dekens, J.A., Brandsma, E., Marczynska, J., Imhann, F., Weersma, R.K., *et al.* (2015). The Gut Microbiome Contributes to a Substantial Proportion of the Variation in Blood Lipids. Circ Res *117*, 817-824.
Getz, G.S., and Reardon, C.A. (2012). Animal models of atherosclerosis. Arterioscler Thromb Vasc Biol *32*, 1104-1115.
Ghazalpour, A., Rau, C.D., Farber, C.R., Bennett, B.J., Orozco, L.D., van Nas, A., Pan, C., Allayee, H., Beaven, S.W., Civelek, M., *et al.* (2012). Hybrid mouse diversity panel: a panel of inbred mouse strains suitable for analysis of complex genetic traits. Mamm Genome *23*, 680-692.
Gibson, G.R., and Roberfroid, M.B. (1995). Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. The Journal of nutrition *125*, 1401-1412.
Gimbrone, M.A., Topper, J.N., Nagel, T., Anderson, K.R., and Garcia-Cardeña, G. (2000). Endothelial dysfunction, hemodynamic forces, and atherogenesis. Ann N Y Acad Sci *902*, 230-239.
Goldstein, J.L., Schrott, H.G., Hazzard, W.R., Bierman, E.L., and Motulsky, A.G. (1973). Hyperlipidemia in coronary heart disease. II. Genetic analysis of lipid levels in 176 families and delineation of a new inherited disorder, combined hyperlipidemia. J Clin Invest *52*, 1544-1568.
Gotto, A.M., Jr., and Moon, J.E. (2012). Management of cardiovascular risk: the importance of meeting lipid targets. Am J Cardiol *110*, 3A-14A.
Gregory, J.C., Buffa, J.A., Org, E., Wang, Z., Levison, B.S., Zhu, W., Wagner, M.A., Bennett, B.J., Li, L., DiDonato, J.A., *et al.* (2015). Transmission of atherosclerosis susceptibility with gut microbial transplantation. J Biol Chem *290*, 5647-5660.

Greig, F.H., Kennedy, S., and Spickett, C.M. (2012). Physiological effects of oxidized phospholipids and their cellular signaling mechanisms in inflammation. Free Radic Biol Med *52*, 266-280.

Grundy, S.M., Cleeman, J.I., Merz, C.N., Brewer, H.B., Jr., Clark, L.T., Hunninghake, D.B., Pasternak, R.C., Smith, S.C., Jr., Stone, N.J., and Coordinating Committee of the National Cholesterol Education, P. (2004). Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III Guidelines. J Am Coll Cardiol *44*, 720-732.

Hager, R., Cheverud, J.M., Leamy, L.J., and Wolf, J.B. (2008). Sex dependent imprinting effects on complex traits in mice. BMC Evol Biol *8*, 303.

Hoffman, J.B., Petriello, M.C., Morris, A.J., Mottaleb, M.A., Sui, Y., Zhou, C., Deng, P., Wang, C., and Hennig, B. (2020). Prebiotic inulin consumption reduces dioxin-like PCB 126-mediated hepatotoxicity and gut dysbiosis in hyperlipidemic Ldlr deficient mice. Environ Pollut *261*, 114183.

Huang, Y.H., Schäfer-Elinder, L., Wu, R., Claesson, H.E., and Frostegård, J. (1999). Lysophosphatidylcholine (LPC) induces proinflammatory cytokines by a platelet-activating factor (PAF) receptor-dependent mechanism.

Huda, M.N., VerHague, M., Albright, J., Smallwood, T., Bell, T.A., Que, E., Miller, D.R., Roshanravan, B., Allayee, H., Manuel de Villena, F.P., *et al.* (2020). Dissecting the Genetic Architecture of Cystatin C in Diversity Outbred Mice. G3 (Bethesda) *10*, 2529-2541.

Isensee, J., Witt, H., Pregla, R., Hetzer, R., Regitz-Zagrosek, V., and Noppinger, P.R. (2008). Sexually dimorphic gene expression in the heart of mice and men. J Mol Med (Berl) *86*, 61-74.

Ishida, B.Y., Blanche, P.J., Nichols, A.V., Yashar, M., and Paigen, B. (1991). Effects of atherogenic diet consumption on lipoproteins in mouse strains C57BL/6 and C3H. Journal of Lipid Research *32*, 559-568.

Jensen, J., and Blankenhorn, D.H. (1972). The inheritance of familial hypercholesterolemia. Am J Med *52*, 499-516.

Jie, Z., Xia, H., Zhong, S.L., Feng, Q., Li, S., Liang, S., Zhong, H., Liu, Z., Gao, Y., Zhao, H., *et al.* (2017). The gut microbiome in atherosclerotic cardiovascular disease. Nat Commun *8*, 845.

Karlsson, F.H., Fak, F., Nookaew, I., Tremaroli, V., Fagerberg, B., Petranovic, D., Backhed, F., and Nielsen, J. (2012). Symptomatic atherosclerosis is associated with an altered gut metagenome. Nat Commun *3*, 1245.

Karp, N.A., Mason, J., Beaudet, A.L., Benjamini, Y., Bower, L., Braun, R.E., Brown, S.D.M., Chesler, E.J., Dickinson, M.E., Flenniken, A.M., *et al.* (2017). Prevalence of sexual dimorphism in mammalian phenotypic traits. Nat Commun *8*, 15475.

Kasahara, K., Krautkramer, K.A., Org, E., Romano, K.A., Kerby, R.L., Vivas, E.I., Mehrabian, M., Denu, J.M., Backhed, F., Lusis, A.J., *et al.* (2018). Interactions between Roseburia intestinalis and diet modulate atherogenesis in a murine model. Nat Microbiol *3*, 1461-1471.

Katz, D.C., Aponte, J.D., Liu, W., Green, R.M., Mayeux, J.M., Pollard, K.M., Pomp, D., Munger, S.C., Murray, S.A., Roseman, C.C., *et al.* (2020). Facial shape and allometry quantitative trait locus intervals in the Diversity Outbred mouse are enriched for known skeletal and facial development genes. PLoS One *15*, e0233377.

Kayashima, Y., Tomita, H., Zhilicheva, S., Kim, S., Kim, H.S., Bennett, B.J., and Maeda, N. (2014). Quantitative trait loci affecting atherosclerosis at the aortic root identified in an intercross between DBA2J and 129S6 apolipoprotein E-null mice. PLoS One *9*, e88274.

Keenan, B.T., Galante, R.J., Lian, J., Simecek, P., Gatti, D.M., Zhang, L., Lim, D.C., Svenson, K.L., Churchill, G.A., and Pack, A.I. (2020). High-throughput sleep phenotyping produces robust and heritable traits in Diversity Outbred mice and their founder strains. Sleep *43*.

Keller, M.P., Rabaglia, M.E., Schueler, K.L., Stapleton, D.S., Gatti, D.M., Vincent, M., Mitok, K.A., Wang, Z., Ishimura, T., Simonett, S.P., *et al.* (2019). Gene loci associated with insulin secretion in islets from nondiabetic mice. Journal of Clinical Investigation *129*, 4419-4432.

Kelly, T.N., Bazzano, L.A., Ajami, N.J., He, H., Zhao, J., Petrosino, J.F., Correa, A., and He, J. (2016). Gut Microbiome Associates With Lifetime Cardiovascular Disease Risk Profile Among Bogalusa Heart Study Participants. Circ Res *119*, 956-964.

Kemis, J.H., Linke, V., Barrett, K.L., Boehm, F.J., Traeger, L.L., Keller, M.P., Rabaglia, M.E., Schueler, K.L., Stapleton, D.S., Gatti, D.M., *et al.* (2019). Genetic determinants of gut microbiota composition and bile acid profiles in mice. PLoS Genet *15*, e1008073.

Kennedy, E.A., King, K.Y., and Baldridge, M.T. (2018). Mouse Microbiota Models: Comparing Germ-Free Mice and Antibiotics Treatment as Tools for Modifying Gut Bacteria. Front Physiol *9*, 1534.

Kukurba, K.R., Parsana, P., Balliu, B., Smith, K.S., Zappala, Z., Knowles, D.A., Fave, M.J., Davis, J.R., Li, X., Zhu, X., *et al.* (2016). Impact of the X Chromosome and sex on regulatory variation. Genome Res *26*, 768-777.

Kurtz, S.L., Rossi, A.P., Beamer, G.L., Gatti, D.M., Kramnik, I., and Elkins, K.L. (2020). The Diversity Outbred Mouse Population Is an Improved Animal Model of Vaccination against Tuberculosis That Reflects Heterogeneity of Protection. mSphere *5*.

Lander, E.S., and Schork, N.J. (1994). Genetic dissection of complex traits. Science *265*, 2037-2048.

LeBoeuf, R.C., Puppione, D.L., Schumaker, V.N., and Lusis, A.J. (1983). Genetic control of lipid transport in mice. I. Structural properties and polymorphisms of plasma lipoproteins. Journal of Biological Chemistry *258*, 5063-5070.

Li, J., Zhao, F., Wang, Y., Chen, J., Tao, J., Tian, G., Wu, S., Liu, W., Cui, Q., Geng, B., *et al.* (2017). Gut microbiota dysbiosis contributes to the development of hypertension. Microbiome *5*, 14.

Libby, P., Ridker, P.M., Hansson, G.K., and Leducq Transatlantic Network on, A. (2009). Inflammation in atherosclerosis: from pathophysiology to practice. J Am Coll Cardiol *54*, 2129-2138.

Linke, V., Overmyer, K.A., Miller, I.J., Brademan, D.R., Hutchins, P.D., Trujillo, E.A., Reddy, T.R., Russell, J.D., Cushing, E.M., Schueler, K.L., *et al.* (2020). A large-scale genome-lipid association map guides lipid identification. Nat Metab *2*, 1149-1162.

Luedde, M., Winkler, T., Heinsen, F.A., Ruhlemann, M.C., Spehlmann, M.E., Bajrovic, A., Lieb, W., Franke, A., Ott, S.J., and Frey, N. (2017). Heart failure is associated with depletion of core intestinal microbiota. ESC Heart Fail *4*, 282-290.

Lusis, A. (2000). Atherosclerosis. Nature *407*, 233-241.

Lusis, A.J. (2012). Genetics of atherosclerosis. Trends Genet *28*, 267-275.

Lusis, A.J., Seldin, M.M., Allayee, H., Bennett, B.J., Civelek, M., Davis, R.C., Eskin, E., Farber, C.R., Hui, S., Mehrabian, M., *et al.* (2016). The Hybrid Mouse Diversity Panel: a resource for systems genetics analyses of metabolic and cardiovascular traits. J Lipid Res *57*, 925-942.

Lusis, A.J., Taylor, B.A., Wangenstein, R.W., and LeBoeuf, R.C. (1983). Genetic control of lipid transport in mice. II. Genes controlling structure of high density lipoproteins. Journal of Biological Chemistry *258*, 5071-5078.

Masucci-Magoulas, L., Goldberg, I.J., Bisgaier, C.L., Serajuddin, H., Francone, O.L., Breslow, J.L., and Tall, A.R. (1997). A mouse model with features of familial combined hyperlipidemia. Science *275*, 391-394.

Matthan, N.R., Solano-Aguilar, G., Meng, H., Lamon-Fava, S., Goldbaum, A., Walker, M.E., Jang, S., Lakshman, S., Molokin, A., Xie, Y., *et al.* (2018). The Ossabaw Pig Is a Suitable Translational Model to Evaluate Dietary Patterns and Coronary Artery Disease Risk. J Nutr *148*, 542-551.

McCarthy, M.M., Auger, A.P., Bale, T.L., De Vries, G.J., Dunn, G.A., Forger, N.G., Murray, E.K., Nugent, B.M., Schwarz, J.M., and Wilson, M.E. (2009). The epigenetics of sex differences in the brain. J Neurosci *29*, 12815-12823.

Mehrabian, M., Wong, J., Wang, X., Jiang, Z., Shi, W., Fogelman, A.M., and Lusis, A.J. (2001). Genetic locus in mice that blocks development of atherosclerosis despite extreme hyperlipidemia. Circ Res *89*, 125-130.

Miller, Y.I., Choi, S.H., Wiesner, P., Fang, L., Harkewicz, R., Hartvigsen, K., Boullier, A., Gonen, A., Diehl, C.J., Que, X., *et al.* (2011). Oxidation-specific epitopes are danger-associated molecular patterns recognized by pattern recognition receptors of innate immunity. Circ Res *108*, 235-248.

Mittelstrass, K., Ried, J.S., Yu, Z., Krumsiek, J., Gieger, C., Prehn, C., Roemisch-Margl, W., Polonikov, A., Peters, A., Theis, F.J., *et al.* (2011). Discovery of sexual dimorphisms in metabolic and genetic biomarkers. PLoS Genet *7*, e1002215.

Morrisett, J.D., Kim, H.S., Patsch, J.R., Datta, S.K., and Trentin, J.J. (1982). Genetic susceptibility and resistance to diet-induced atherosclerosis and hyperlipoproteinemia. Atherosclerosis *2*, 312-324.

Mozhui, K., Lu, L., Armstrong, W.E., and Williams, R.W. (2012). Sex-specific modulation of gene expression networks in murine hypothalamus. Front Neurosci *6*, 63.

Nakashima, A., Yamada, K., Iwata, O., Sugimoto, R., Atsuji, K., Ogawa, T., Ishibashi-Ohgo, N., and Suzuki, K. (2018). beta-Glucan in Foods and Its Physiological Functions. J Nutr Sci Vitaminol (Tokyo) *64*, 8-17.

Nishina, P.M., Verstuyft, J., and Paigen, B. (1990). Synthetic low and high fat diets for the study of atherosclerosis in the mouse. Journal of Lipid Research *31*, 859-869.

Norheim, F., Hasin-Brumshtein, Y., Vergnes, L., Chella Krishnan, K., Pan, C., Seldin, M.M., Hui, S.T., Mehrabian, M., Zhou, Z., Gupta, S., *et al.* (2019). Gene-by-Sex Interactions in Mitochondrial Functions and Cardio-Metabolic Traits. Cell Metabolism *29*, 932-949.e934.

Pagidipati, N.J., and Gaziano, T.A. (2013). Estimating deaths from cardiovascular disease: a review of global methodologies of mortality measurement. Circulation *127*, 749-756.

Paigen, B., Havens, M.B., and Morrow, A. (1985a). Effect of 3-methylcholanthrene on the development of aortic lesions in mice. Cancer Res *45*, 3850-3855.

Paigen, B., Morrow, A., Brandon, C., Mitchell, D., and Holmes, P. (1985b). Variation in susceptibility to atherosclerosis among inbred strains of mice. Atherosclerosis *57*, 65-73.
Pankey, G.A. (1965). EFFECT OF VIRUSES ON THE CARDIOVASCULAR SYSTEM. Am J Med Sci.
Pant, S., Deshmukh, A., Gurumurthy, G.S., Pothineni, N.V., Watts, T.E., Romeo, F., and Mehta, J.L. (2014). Inflammation and atherosclerosis--revisited. J Cardiovasc Pharmacol Ther *19*, 170-178.
Pothineni, N.V.K., Subramany, S., Kuriakose, K., Shirazi, L.F., Romeo, F., Shah, P.K., and Mehta, J.L. (2017). Infections, atherosclerosis, and coronary heart disease. Eur Heart J *38*, 3195-3201.
Rao, D.C. (2001). Genetic Dissection of Complex Traits. Adv Genet *42*, 13-34.
Rau, C.D., Lusis, A.J., and Wang, Y. (2020). Systems Genetics for Mechanistic Discovery in Heart Diseases. Circ Res *126*, 1795-1815.
Recla, J.M., Bubier, J.A., Gatti, D.M., Ryan, J.L., Long, K.H., Robledo, R.F., Glidden, N.C., Hou, G., Churchill, G.A., Maser, R.S., *et al.* (2019). Genetic mapping in Diversity Outbred mice identifies a Trpa1 variant influencing late-phase formalin response. Pain *160*, 1740-1753.
Roberts, A., and Thompson, J.S. (1976). Inbred mice and their hypbrids as an animal model for atherosclerosis research. Adv Exp Med Biol *67*, 313-327.
Ryan, P.M., London, L.E., Bjorndahl, T.C., Mandal, R., Murphy, K., Fitzgerald, G.F., Shanahan, F., Ross, R.P., Wishart, D.S., Caplice, N.M., *et al.* (2017). Microbiome and metabolome modifying effects of several cardiovascular disease interventions in apo-E(-/-) mice. Microbiome *5*, 30.
Saikku, P., Leinonen, M., Mattila, K., Ekman, M.R., Nieminen, M.S., Mäkelä, P.H., Huttunen, J.K., and Valtonen, V. (1988). SEROLOGICAL EVIDENCE OF AN ASSOCIATION OF A NOVEL CHLAMYDIA, TWAR, WITH CHRONIC CORONARY HEART DISEASE AND ACUTE MYOCARDIAL INFARCTION. Lancet *2*, 983-986.
Sanchez-Alcoholado, L., Castellano-Castillo, D., Jordan-Martinez, L., Moreno-Indias, I., Cardila-Cruz, P., Elena, D., Munoz-Garcia, A.J., Queipo-Ortuno, M.I., and Jimenez-Navarro, M. (2017). Role of Gut Microbiota on Cardio-Metabolic Parameters and Immunity in Coronary Artery Disease Patients with and without Type-2 Diabetes Mellitus. Front Microbiol *8*, 1936.
Seldin, M., Yang, X., and Lusis, A.J. (2019). Systems genetics applications in metabolism research. Nat Metab *1*, 1038-1050.
Smallwood, T.L., Gatti, D.M., Quizon, P., Weinstock, G.M., Jung, K.C., Zhao, L., Hua, K., Pomp, D., and Bennett, B.J. (2014). High-resolution genetic mapping in the diversity outbred mouse population identifies Apobec1 as a candidate gene for atherosclerosis. G3 (Bethesda) *4*, 2353-2363.
Smith, J. (2003). Quantitative trait locus mapping for atherosclerosis susceptibility. Curr Opin Lipidol *14*, 499-504.
Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermúdez-Humarán, L.G., Gratadoux, J.J., Blugeon, S., Bridonneau, C., Furet, J.P., Corthier, G., *et al.* (2008). Faecalibacterium prausnitzii is an anti-inflammatory ... erium identified by gut microbiota analysis of Crohn disease patients. Proc Natl Acad Sci U S A *105*, 16731-16736.
Solberg, L.C., Valdar, W., Gauguier, D., Nunez, G., Taylor, A., Burnett, S., Arboledas-Hita, C., Hernandez-Pliego, P., Davidson, S., Burns, P., *et al.* (2006). A protocol for high-

throughput phenotyping, suitable for quantitative trait analysis in mice. Mamm Genome *17*, 129-146.
Su, Z., Li, Y., James, J.C., McDuffie, M., Matsumoto, A.H., Helm, G.A., Weber, J.L., Lusis, A.J., and Shi, W. (2006). Quantitative trait locus analysis of atherosclerosis in an intercross between C57BL/6 and C3H mice carrying the mutant apolipoprotein E gene. Genetics *172*, 1799-1807.
Taur, Y., and Pamer, E.G. (2013). The intestinal microbiota and susceptibility to infection in immunocompromised patients. Curr Opin Infect Dis *26*, 332-337.
Tavolara, T.E., Niazi, M.K.K., Ginese, M., Piedra-Mora, C., Gatti, D.M., Beamer, G., and Gurcan, M.N. (2020). Automatic discovery of clinically interpretable imaging biomarkers for Mycobacterium tuberculosis supersusceptibility using deep learning. EBioMedicine *62*, 103094.
Teupser, D., Tan, M., Persky, A.D., and Breslow, J.L. (2006). Atherosclerosis quantitative trait loci are sex-an ... rcross of C57BL - 6 and FVB - N low-density lipoprotein receptor-mice. Proc Natl Acad Sci U S A *103*, 123-128.
Torres, N., Guevara-Cruz, M., Velazquez-Villegas, L.A., and Tovar, A.R. (2015). Nutrition and Atherosclerosis. Arch Med Res *46*, 408-426.
Tracy, R.E., Newman, W.P., Wattigney, W.A., Srinivasan, S.R., Strong, J.P., and Berenson, G.S. (1995). Histologic features of atherosclerosis and hypertension from autopsies of young individuals in a defined geographic population the Bogalusa Heart Study. Atherosclerosis *116*, 163-179.
Turk-Adawi, K.I., and Grace, S.L. (2015). Narrative review comparing the benefits of and participation in cardiac rehabilitation in high-, middle- and low-income countries. Heart Lung Circ *24*, 510-520.
Valdar, W., Solberg, L.C., Gauguier, D., Burnett, S., Klenerman, P., Cookson, W.O., Taylor, M.S., Rawlins, J.N., Mott, R., and Flint, J. (2006). Genome-wide genetic association of complex traits in heterogeneous stock mice. Nat Genet *38*, 879-887.
Van Craeyveld, E., Jacobs, F., Feng, Y., Thomassen, L.C., Martens, J.A., Lievens, J., Snoeys, J., and De Geest, B. (2010). The relative atherogenicity of VLDL and LDL is dependent on the topographic site. J Lipid Res *51*, 1478-1485.
Virani, S.S., Alonso, A., Benjamin, E.J., Bittencourt, M.S., Callaway, C.W., Carson, A.P., Chamberlain, A.M., Chang, A.R., Cheng, S., Delling, F.N., *et al.* (2020). Heart Disease and Stroke Statistics-2020 Update: A Report From the American Heart Association. Circulation *141*, e139-e596.
Virmani, R., Burke, A.P., Kolodgie, F.D., and Farb, A. (2002). Vulnerable plaque the pathology of unstable coronary lesions. J Interv Cardiol *15*, 439-446.
Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet *101*, 5-22.
Wang, S.S., Schadt, E.E., Wang, H., Wang, X., Ingram-Drake, L., Shi, W., Drake, T.A., and Lusis, A.J. (2007). Identification of pathways for atherosclerosis in mice: integration of quantitative trait locus analysis and global gene expression data. Circ Res *101*, e11-30.
Westerterp, M., van der Hoogt, C.C., de Haan, W., Offerman, E.H., Dallinga-Thie, G.M., Jukema, J.W., Havekes, L.M., and Rensen, P.C. (2006). Cholesteryl ester transfer protein decreases high-density lipoprotein and severely aggravates atherosclerosis in APOE*3-Leiden mice. Arterioscler Thromb Vasc Biol *26*, 2552-2559.

Winham, S.J., de Andrade, M., and Miller, V.M. (2015). Genetics of cardiovascular disease: Importance of sex and ethnicity. Atherosclerosis *241*, 219-228.

Winter, J.M., Gildea, D.E., Andreas, J.P., Gatti, D.M., Williams, K.A., Lee, M., Hu, Y., Zhang, S., Program, N.C.S., Mullikin, J.C., *et al.* (2017). Mapping Complex Traits in a Diversity Outbred F1 Mouse Population Identifies Germline Modifiers of Metastasis in Human Prostate Cancer. Cell Syst *4*, 31-45 e36.

Yan, Q., Gu, Y., Li, X., Yang, W., Jia, L., Chen, C., Han, X., Huang, Y., Zhao, L., Li, P., *et al.* (2017). Alterations of the Gut Microbiome in Hypertension. Front Cell Infect Microbiol *7*, 381.

Yang, C., Wang, Y., Xu, W., Liu, Z., Zhou, S., Zhang, M., and Cui, D. (2019). Genome-wide association study using diversity outcross mice identified candidate genes of pancreatic cancer. Genomics *111*, 1882-1888.

Yang, W.S., Nevin, D.N., Peng, R., Brunzell, J.D., and Deeb, S.S. (1995). A mutation in the promoter of the lipoprotein lipase ... a patient with familial combined hyperlipidemia and low LPL activity. Proc Natl Acad Sci U S A *92*, 4462-4466.

Yang, X., Schadt, E.E., Wang, S., Wang, H., Arnold, A.P., Ingram-Drake, L., Drake, T.A., and Lusis, A.J. (2006). Tissue-specific expression and regulation of sexually dimorphic genes in mice. Genome Res *16*, 995-1004.

CHAPTER 2.

Literature review —Microbiota, Cardiovascular Disease, and Type 2 Diabetes

Myungsuk Kim[1,2], M. Nazmul Huda[1], Brian J. Bennett[1,2*]

2.1. Author Contributions

This manuscript was published in the Cardiovascular Research journal on January 29th, 2021. B.J.B conceived and initiated this review study. He supervised all portion of the review process and mentored manuscript writing. M.K. and M.N.H. conducted the literature search, extracting the information and drafting the manuscript. M.K. and M.N.H. also addressed co-authors comments and concerns. B.J.B., M.K., and M.N.H. critically revised the manuscript. B.J.B had primary responsibility for final content. All authors read and approved the final manuscript.

2.2. Abstract

The discovery that gut-microbiota plays a profound role in human health has opened a new avenues of basic and clinical research. Application of ecological approaches where the Bacterial 16S rRNA gene is queried has provided a number of candidate Bacteria associated with coronary artery disease and hypertension. We examine the associations between gut microbiota and a variety of CVD including atherosclerosis, coronary artery disease and blood pressure. These approaches are associative in nature and there is now increasing interest in identifying the mechanisms underlying these associations. We discuss three potential mechanisms including: gut permeability and endotoxemia, increased immune system activation, and microbial derived metabolites. In addition to discussing these potential mechanisms, we highlight current studies manipulating the gut microbiota or microbial metabolites to move beyond sequenced based association studies. The goal of these mechanistic studies is to determine the mode of action by which the gut microbiota may affect disease susceptibility and severity. Importantly, the gut microbiota appears to have a significant effect on host metabolism and CVD by producing metabolites entering the host circulatory system such as short chain fatty acids (SCFAs) and trimethylamine N-Oxide (TMAO). Therefore, the intersection of metabolomics and microbiota research may yield novel targets to reduce disease susceptibility. Finally, we discuss approaches to demonstrate causality such as specific diet changes, inhibition of microbial pathways and fecal microbiota transplant.

2.3. Introduction

We are beginning to appreciate the role of commensal microbiota in CVD risk (Qin et al., 2010; Velmurugan et al., 2017) and these microbiota are located in a variety of niches within the body including the skin, oral, and gut, and are composed of bacteria, viruses, and fungi. There

has been a particular focus on the gut as it is heavily colonized with microbes. It is estimated that more than 70% of all the microbes in the human body are present in the colon alone (Ley et al., 2006a) and trillions of commensal microbes including bacteria, bacteriophage, fungi, archaea, and unicellular eukaryotes live in the large intestine that forms a highly diverse dynamic interdependent complex ecological community known as the intestinal microbiota (Lozupone et al., 2012). The healthy gut microbiota depends on the host for their energy need and provides a range of health benefits by providing nutrients, improved barrier function, shaping host immune system, and preventing diseases including CVD (Sekirov et al., 2010).

The gut microbiota is not the only commensal communities studied in relation to CVD risk. There is also evidence that the oral microbiota, which contains more than 10,000 bacterial species from 22 phyla (Keijser et al., 2008) is also associated with CVD. Data from large epidemiological studies associated poor oral hygiene with increased risk of CVD (de Oliveira et al., 2010), and several oral bacteria including *Porphyromonas gingivalis* have been detected in atherosclerotic plaques (Hayashi et al., 2011; Zhang et al., 2010b). What isn't clear is if the association between oral microbiota and CVD indicates a specific mechanism or metabolic profile leading to increased disease or is a marker of other important factors to consider such as access to healthcare. Additional work remains to determine the mechanism(s) underlying the oral microbe-CVD associations and thus the majority of the review focuses on the gut microbiota and CVD.

The combined complexity of the microbiota is only beginning to be appreciated. Maternal transmission (Van Daele et al., 2019), diet (Wu et al., 2011), host genetics (Goodrich et al., 2014), aging (Takagi et al., 2019), and sex (Sinha et al., 2019) all impact the composition of the microbiota. Recent large-scale sequencing efforts have identified tremendous interindividual

variation in the microbiota which totals in excess of 45 million genes in the oral and gut niches. Moreover, nearly 50% of all genes were "unique to a single metagenomic sample (Tierney et al., 2019) and only ~550,000, less than 2%, overlapped between the oral and gut niches. This heterogeneity combined with the possibility of copy number variants within individual bacterial species highlights just how much of the microbiota remains to be characterized.

In this review, we focus on the association between microbiota and CVD, specifically coronary artery disease and hypertension. We examine the steps to establishing causality and then discuss the underline mechanisms by which gut microbiota may affect CVD. Finally, we examine potential therapeutics for CVD using gut microbiota and microbial metabolites.

2.4. Microorganisms and CVD Risk

For over 50 years, the potential role of infectious microorganisms, including bacteria and viruses, as potential risk factors for CVD has been appreciated through a number of epidemiological studies (Cluff et al., 1968; Pankey, 1965; Saikku et al., 1988). Infection and the subsequent inflammatory processes are thought to induce the onset, progression, and rupture of atherosclerotic plaques (Libby et al., 2009). The mechanisms by which infection aggravates atherosclerosis is direct cell invasion that accelerates plaque growth through local effects, or indirect systemic production of inflammatory cytokines that promote the development of atherosclerosis (Pant et al., 2014). Representative infectious agents associated with atherosclerosis include *Chlamydia pneumonia*, *Helicobacter pylori*, *P. gingivalis*, and Influenza A virus (Pothineni et al., 2017).

Although there are a number of pathogenic bacteria associated with CVD, *C. pneumonia*, a Gram-negative and intracellular bacteria, is an exemplar. *C. pneumonia* was the first proposed bacteria of CVD etiology responsible for the induction of inflammation in the vascular wall of

CVD patients (Kuo et al., 1995). In addition, antibiotics targeting *C. pneumonia* are thought to have anti-inflammatory effects, and may contribute to atherosclerotic plaque stability (Pothineni et al., 2017). For example, *C. pneumoniae* infection in rabbits accelerated the thickness of the tunica intimal wall and atherosclerosis, and treatment with antibiotics reduced the extent of atherosclerosis (Muhlestein et al., 1998). This effect is not universal as studies in mice have not shown a similar reduction in lesion size by antibiotics after infection with *C. pneumoniae* (Blessing et al., 2005). Nevertheless, there are examples of pathogenic bacteria potentially having direct effects on CVD. We now discuss how commensal bacteria of the microbiota may also affect CVD risk.

2.5. Microbiota and Cardiovascular Disease (CVD) Associations

CVD includes a number of pathologies and diseases such as heart failure, stroke, peripheral artery disease, aortic valve disease, atherosclerosis, and hypertension but we limit our discussion to coronary artery disease (CAD) and hypertension. In a number of case-control studies, distinct shifts of microbial composition have been identified in fecal samples from patients with CVD (**Table 2.1**). It is not entirely clear whether these microbial communities are causal to CVD or if they are simply affected by other environmental factors contributing to the development of CVD. In the following section, we provide a summary of a number of recently reported associations between gut microbiota and coronary artery disease or hypertension.

2.5.1. Gut Microbiota and Coronary Artery Disease

Dysbiosis refers to an imbalance in the microbial community in the human body. Alterations in the composition of the gut microbiota associated with the risk of CVD have been the focus of the majority of microbiota studies. Indeed, identifying patterns of specific composition of microbiota associated with CVD susceptibility contribute greatly to

understanding the pathogenesis of diseases. The composition of the microbiota associated with aortic lesion size is distinct from the oral and gut microbiota (Koren et al., 2011). Frequently, studies utilize a case-control design to identify the microbial differences among subjects with CVD compared to controls. Foremost, there have been several consistent results reported from a number of cohorts (**Table 2.1)** (Dinakaran et al., 2014; Emoto et al., 2016; Kelly et al., 2016). For example, *Strepcococcus* spp., exhibiting pathogenic expansion in intestinal dysbiosis (Taur and Pamer, 2013), showed increased abundance in stool samples from various CVD patients (Feng et al., 2016; Jie et al., 2017; Sanchez-Alcoholado et al., 2017; Yan et al., 2017). In addition, the abundance of *Faecalibacterium* spp. and *Roseburia* spp., which are known to have anti-inflammatory effects and reduce atherosclerotic events in mice and humans (Kasahara et al., 2018; Sokol et al., 2008), were relatively depleted in CAD patients (Jie et al., 2017; Sanchez-Alcoholado et al., 2017). Finally, metagenome studies have shown depletion of butyrate-producing bacteria including *Faecalibacterium prausnitzii* and *Roseburia intestinalis,* and a high abundance of *Enterobacteriaceae* and *Streptococcus* in atherosclerotic patients (Jie et al., 2017). Studies has also identified alterations in the metagenomic profile of the microbiome in CVD patients. These data demonstrate enrichment of genes encoding peptidoglycan biosynthesis, while healthy cohorts are enriched in phytoene dehydrogenase genes (Karlsson et al., 2012). Thus, not only are the composition of the microbiota altered in CVD but the underlying functional capacity of the microbiota may be altered.

Table 2.1. Reported alterations in the gut microbiota in various CVD cohorts.

Condition	Ethnicity	Technique	Associated microbiota changes	References
Myocardial infarction	European (Swedish)	Metagenomic sequencing	Increased: *Collinsella* Decreased: *Eubacterium, Roseburia*	(Karlsson et al., 2012)
CAD / T2D	European (Spanish)	16S rRNA V2-V3 region	Increased: *Enterobacteriaceae, Streptococcus,* and *Desulfovibrio* Decreased: *Faecalibacterium prausnitzii* and *Bacteroides fragilis*	(Sanchez-Alcoholado et al., 2017)
CVD	European (American)	16S rRNA V4 region	Increased: *Prevotella* and *Tyzzerella* Decreased: *Alloprevotella*, *Catenibacterium*	(Kelly et al., 2016)
CVD	Asian (Indian)	16S rRNA & Metagenomic sequencing	Increased: *Proteobacteria, Actinobacteria, Propionibacterium phages, Pseudomonas phages, Rhizobium phages, Lymphocystis virus,* and *Torque Teno viruses*	(Dinakaran et al., 2014)
CAD	Asian (Japanese)	16S rRNA V3-V4 region	Increased: *Firmicutes/bateriodetes ratio* and *Lactobacillales* Decreased: *Bacteroides* and *Prevotella*	(Emoto et al., 2016)
CAD	Asian (Chinese)	Metagenomic sequencing	Increased: *Streptococcus sp. M334 and M143,* and *Clostridium sp. HGF2*	(Feng et al., 2016)
CVD	Asian (Chinese)	Metagenomic sequencing	Increased: *Escherichia coli, Klebsiella spp, Enterobacter aerogenes, Streptococcus sp, Lactobacillus salivarius, Solobacterium moorei, Atropobium parvulum, Ruminococcus gnavus,* and *Eggerthella lenta* Decreased: *Roseburia intestinalis, Faecalibacterium cf. prausnitzii, Bacteriodes spp, Prevotella copri,* and *Alistipes shahii*	(Jie et al., 2017)
Hypertension	Asian (Chinese)	Metagenomic sequencing	Increased: *Prevotella, Klebsiella, Porphyromonas,* and *Actinomyces* Decreased: *Faecalibacterium, Oscillibacter, Roseburia, Bifidobacterium, Coprococcus,* and *Butyrivibrio*	(Li et al., 2017)
Hypertension	Asian (Chinese)	Metagenomic sequencing	Increased: *Klebsiella, Clostridium, Streptococcus, Parabacteroides, Eggerthella,* and *Salmonella* Decreased: *Faecalibacterium, Roseburia,* and *Synergistetes*	(Yan et al., 2017)
Hypertension	European (Finns)	Metagenomic sequencing	Increased: *Blautia, Cellulomonas, Collinsella,* and *Desulfovibrio, Dielma, Eisenbergiella, Holdemania, Megasphaera, Phascolarctobacterium, Ruthenibacterium, Sutterella,* and *Turicibacter* Decreased: *Lactobacillusa,* and *Citrobacter, Coprobacillus*	(Palmu et al., 2020)
Hypertension	European/ African (American)	16S rRNA V3-V4 region	Increased: *Anaerovorax, Clostridium IV, Oscillibacter,* and *Sporobacter* Decreased: *Akkermansia, Ruminococcus, Anaerovorax, Sporobacter,* and *Asaccharobacter*	(Sun et al., 2019)

2.5.2. Gut Microbiota and Hypertension

Similar to CAD there is increasing interest in how gut microbiota may affect blood pressure and the development of hypertension. Studies of angiotensin II infused mice, spontaneous hypertensive rats, and hypertensive patients have identified a distinct microbial composition of gut microbiota compared to controls including decreased microbial abundance, diversity, and low intestinal epithelial integrity (Adnan et al., 2017; Sata et al., 2020; Yang et al., 2015). Broad-based antibiotic treatment in mice reduced intestinal dysbiosis and attenuated hypertension in angiotensin II-infused mice (Pevsner-Fischer et al., 2017). Furthermore, angiotensin II-induced hypertension is alleviated in germ-free mice compared to conventionally raised mice, suggesting that the gut microbiota is important for the development of hypertension (Karbach et al., 2016). Studies in humans have clearly demonstrated that there is an inverse association between measures of α-diversity and hypertension (Sata et al., 2020). In addition to reduced diversity, there is an increased abundance of opportunistic pathogenic taxa such as *Klebsiella*, and *Streptococcus* in subjects with hypertension (Li et al., 2017; Sun et al., 2019; Yan et al., 2017), while several taxa are associated with high blood pressure (see **Table 2.1**), and may affect intestinal cell inflammation (Kim et al., 2018a). In a longitudinal cohort, CARDIA, containing both Caucasians and African American subjects, identified a slightly different profile of bacteria positively associated with hypertension: *Anaerovorax*, *Clostridium* IV, *Oscillibacter*, and *Sporobacter* (Sun et al., 2019). Conversely, a number of studies have identified *Bacteroides thetaiotaomicron* as associated with reduced blood pressure in both hypertensive and healthy subjects (Kim et al., 2018a; Palmu et al., 2020; Sun et al., 2019; Verhaar et al., 2020). Although these differences between hypertensive subjects and healthy cohorts do not indicate causality, they do suggest that the composition and diversity of the gut microbiota are associated with

clinical features of hypertension. Importantly, having the microbiota assessed in longitudinal cohorts such as CARDIA allows for epidemiologists to assess the risk of developing hypertension (or CAD) based on the current microbiota composition and thereby allowing for the assessment of the clinical utility of these data for predicting disease.

2.6. Limitations of Association-Based Microbiota Studies and CVD Risk

A majority of microbiome-based studies utilize analysis of the bacterial 16S rRNA gene. For details on the methods and analytical techniques needed for 16S analysis, we direct the reader to several excellent reviews (Boers et al., 2019; Golebiewski and Tretyn, 2020; Kim et al., 2017; Mandal et al., 2015). 16S rRNA studies have provided tremendous insight into the specific bacteria associated with CVD and undeniability have transformed our understanding of disease risk but there are limitations. Commonly used 16S rRNA gene sequencing techniques often do not reach species or sub-species level resolution, and the analysis often excludes less abundant microbial taxa. Thus, studies focused primarily on abundance neglect the functional pathways of microbes that contribute to disease risk, although the disease can be driven by microbes that can represent a small fraction of the microbial community. Additionally, many well-described cohorts designed to longitudinally assess CVD risk have not routinely collected fecal samples. Thus, many of the current reports are cross-sectional in nature and therefore precludes assessment of the relative risk of developing CVD on the current composition of the microbiota, which is critical to developing therapeutic strategies.

We now appreciate some of the biases and intrinsic issues associated with 16S analysis. These include a number of items and several have recently been discussed in the context of hypertension (Marques et al., 2019). First, the quality of DNA obtained from samples can be varied by differences in sample collection method (sample type, collection time, or processing

method), storage and processing techniques (DNA extraction, library construction, sequencing depth, or quality filters). Second, from a technical point of view, amplification bias, improper or no internal sequencing control, contamination, or non-standard databases for mapping can lead to alteration in microbial composition independent of actual microbial changes. Finally, the results of these analyzes are usually expressed as a proportion rather than an absolute number of the microbes per gram of sample, and the proportion of specific microbial taxa in a sample may not be related to the risk of disease. Numerous large-scale efforts are underway to standardize microbiome methods and protocols in larger cohorts in an attempt to address some of these issues (Amos et al., 2020; Brumfield et al., 2020; Shkoporov et al., 2018). More recently, there is interest in utilizing whole-genome sequencing of microbiota samples termed "metagenomics", to provide much better taxonomic resolution down to species or strain level. As metagenomics results in sequences of genes contained in the sample, it provides an opportunity for functional profiling of the metabolic pathways present in community (Thomas et al., 2012) while also providing the taxonomic details available in 16S studies. While metagenomics provides a complete view of the microbial communities, it is expensive and the analytical approaches are computationally complex and time-consuming.

2.7. Towards Function and Mechanistic Understanding of the Microbiota in CVD

Although the microbiota is complex, there are a number of approaches to establishing causality and infer mechanism(s) of microbiota and specific bacteria for CVD (**Figure 2.1**). In particular, studies utilizing microbial transplantation in model organism can demonstrate the direct role of the gut microbial taxa in the risk of CVD. Specifically, the microbiota can be ablated from model organisms such as mice using antibiotic cocktails or germ-free animals. This

provides a naïve state by which individual or complex communities can be added and then examined for hypertension or CVD.

Studies utilizing germ-free animals have been critical to further our understanding of the functional role of the microbiota in disease susceptibility (Kennedy et al., 2018). Germ-free mice can also receive fecal microbiota transfers from samples collected in case/control studies. For example, blood pressure of germ-free mice increased when they received a transplant of stool from hypertensive patients as compared to mice receiving samples from healthy donors. Blood pressure elevations after FMT have also been observed in conventional mouse model (Kim et al., 2018b). Germ-free mice administered stool samples from hypertensive patients or hypertensive rats have an elevation of blood pressure (Li et al., 2017; Toral et al., 2019). These studies provide evidence that gut microbiota can affect the development of hypertension. Studies have been performed using 2 strains of mice that differ in atherosclerosis susceptibility and that at least a portion of atherosclerosis can be attributed to the microbiota (Gregory et al., 2015). It is important to note that the complexity of the microbiota can produce unexpected results such as increased blood pressure in Dahl salt-sensitive rats receiving fecal transplants from salt-resistant normotensive rats (Mell et al., 2015). The latter indicates that the relationship between microbiota and host biology is more complex than simply classifying microbiota as “good” or “bad”.

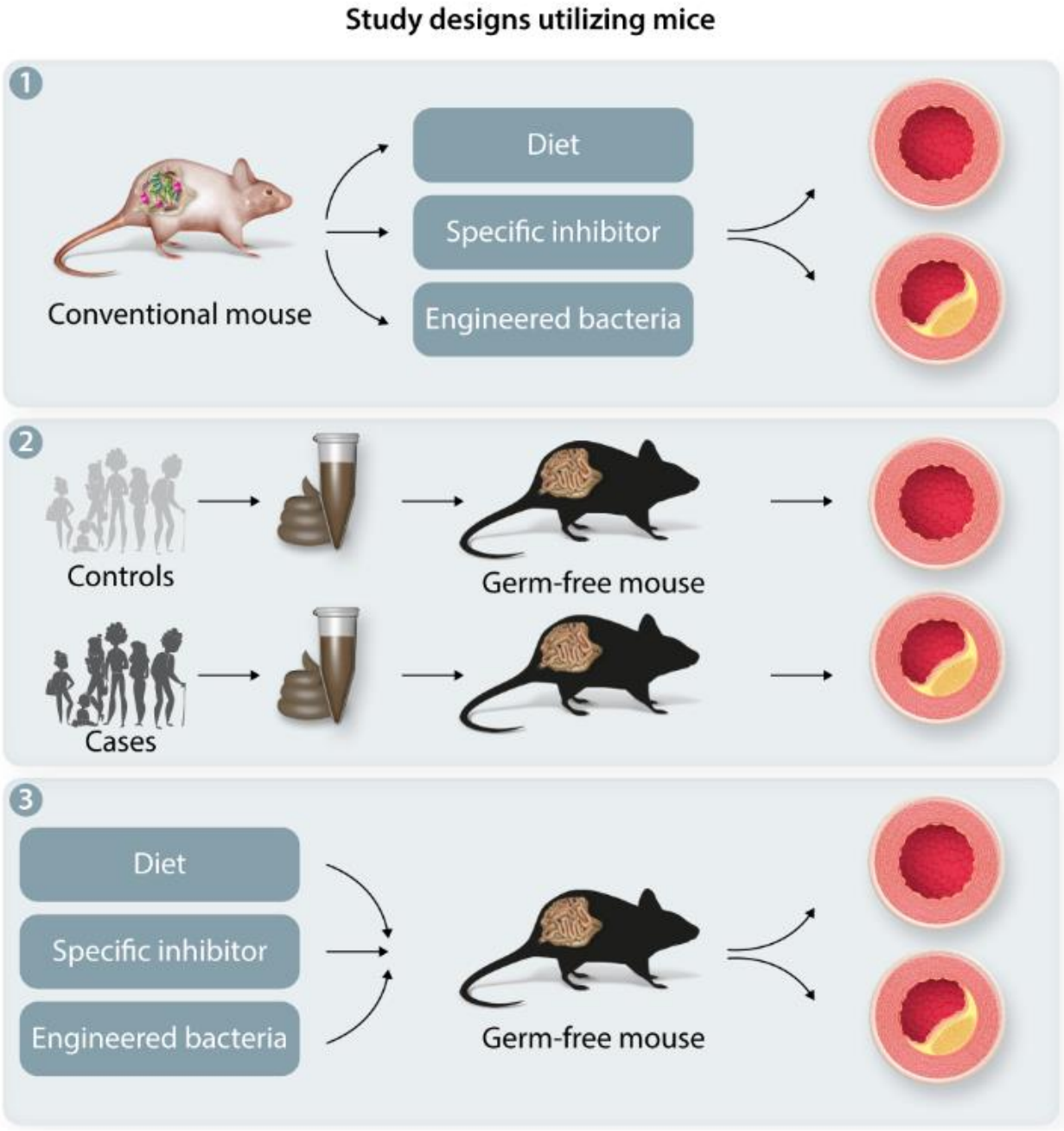


Figure 2.1. Towards function and mechanistic understanding of the microbiota in CVD. Examples of ways to move from association to causation. (A) Examples of microbiome approaches include where conventional mice are provided an agent to perturb the microbiota (either expand specific taxa or inhibit a pathway). (B) Examples of forward microbiome approaches include a comparison of the gut microbiome between controls and CVD subjects from human cohorts. These fecal microbiota samples are then transplanted into germ-free or antibiotic-treated animals to determine the gut microbiota community effect on BP. (C) Diet, specific bacterial enzyme inhibitor, or engineered microbiota can be introduced to animal model to establish a causal relation.

In addition to studies examining the transplantation of fecal samples, there have been attempts at understanding the role of specific bacteria in CVD. For example, colonization of germ-free ApoE-deficient ($ApoE^{-/-}$) mice with *Roseburia intestinalis* isolated from humans reduced the levels of inflammatory markers and atherosclerosis providing evidence of the causative role of *Roseburia intestinalis* in CVD (Kasahara et al., 2018). As the literature expands with similar studies, novel pathways and therapeutic targets may be identified. In the following section, we provide an overview of studies utilizing these functional approaches to provide potential mechanisms for how the microbiota affects CVD risk. In particular, we focus on host:microbe effects on the immune system, gut permeability and microbial derived metabolites. It is important to note that by its nature the gut microbiota is a diverse community and thus it is likely that there are multiple mechanisms for its effects on CVD.

2.7.1. Effect of Gut Microbiota on Cardiovascular Health via Modulation of Immune Function

Gut microbiota is a strong modulator of host immunity and host immune response plays a key role in a wide range of pathology including CVD. For example, atherosclerosis which underlies many forms of CAD is considered a chronic inflammatory disease with the involvement of both innate and adaptive immunity (Ross, 1999). Several gut bacteria have been reported to influence distinct immune cells and there is an indication that T-cells are important in these processes (Atarashi et al., 2011; Furusawa et al., 2013; Gil-Cruz et al., 2019; Round and Mazmanian, 2010). How these systemic alterations in adaptive immunity impact due to specific microbiota CAD or hypertension remain to be fully elucidated.

In mice, there is growing evidence suggesting that the gut microbiota can affect atherosclerosis and inflammatory pathways. For example, microbiota transfer from mice prone to inflammatory dysregulation (Caspase $1^{-/-}$) into atherosclerosis prone low-density lipoprotein receptor deficient ($LDLR^{-/-}$) mice does in fact increase atherosclerotic plaque formation compared to those who received microbiota from $LDLR^{-/-}$ mice (Brandsma et al., 2019). One possible mechanism of this effect is that gut microbiota elevated systemic inflammatory cytokines interleukin (IL)-1β, IL-2, and interferon (IFN)-γ. On the other hand, transplants with *Bacteroides vulgatus* and *Bacteroides dorei* attenuated atherosclerotic lesions and decreased plasma tumor necrosis factor-α (TNFα) level (Yoshida et al., 2018). Similarly, *Lactobacillus plantarum* ATCC 14917 supplementation inhibited atherosclerotic lesion formation by decreasing serum oxidized LDL (OxLDL), TNFα, and IL-1β production in the aorta (Hassan et al., 2020). Higher abundance of the *Roseburia* and *Blautia* among others was associated with a decreased atherosclerotic lesion in mice and reduced plasma total cholesterol, TNFα, and IL-1β concentration (Wu et al., 2020b).

Similarly, hypertension is associated with inflammation and a number of alterations in immune system [reviewed in (Norlander et al., 2018)] In hypertensive humans, decreased relative abundances of butyrate-producing *Roseburia* and *Faecalibacterium* was observed along with increased TNFα:IFN-γ ratio, TNFα and IL-6 production in the isolated peripheral blood mononuclear cells compared to normotensive people. Additionally, high-salt diet-induced hypertension is associated with depleted gut *Lactobacillus* abundance, increased IFN-γ^{+} CD4 T cells and serum IFN-γ level, and decreased TGF-$\beta1^{+}$ CD4 T-cells and serum TGF-β1 concentration (Liu et al., 2019a). These findings do suggest that therapies that modulate the gut microbiota may simultaneously affect inflammatory processes and subsequently CVD.

2.7.2. Gut Permeability and Intestinal Barrier Dysfunction

The cells of the intestine serve as a critical barrier to the bacteria in the gut. This function is maintained by tight junctions between epithelial cells, mucus production, and mucosal immunity. When the intestinal barrier is compromised, lipopolysaccharides (LPS) derived from Gram-negative bacteria can enter the host circulation resulting in endotoxemia. Endotoxemia constitutes a strong risk factor of early atherogenesis (Wiedermann et al., 1999) and circulating LPS can bind to the host's Toll-like receptor (TLR) resulting in an inflammatory response in the host (Hug et al., 2018). Patients with CVD have higher levels of endotoxin in the blood compared to normal individuals (Hsu et al., 2017; Ibrahim et al., 2018). Translocation of LPS from the intestine is supported by a higher concentration of endotoxin in the hepatic vein compared to blood drawn directly from the ventricle (Peschel et al., 2003). There is evidence that specific Bacteria may alter the gut permeability and endotoxemia. Administration of live *A. muciniphila* reduced intestinal permeability, circulating endotoxin and aortic atherosclerosis in $ApoE^{-/-}$ mice (Li et al., 2016a). A recent proof-of-concept study in humans found that administration of pasteurized *A. muciniphila* for 3 months reduced circulating LPS in obese patients (Depommier et al., 2019).

2.7.3. Microbial Metabolites and CVD

Seminal studies by Hazen and colleagues (Zhu et al., 2016) have demonstrated that gut microbiota may in fact generate metabolites that affect overall health or CVD pathogenesis. We are beginning to appreciate that many gut-derived metabolites can act on organs such as the liver via portal circulation and be metabolized by host enzymes. Microbial metabolites also help us understand the underlying mechanism by which gut bacterial taxa may influence host biology and thus health and disease (**Table 2.2; Figure 2.2**). We will briefly discuss the association

between trimethylamine N-Oxide (TMAO) and CVD, and focus on other microbial metabolites such as short-chain fatty acids (SCFAs), tryptophan metabolites, and bile acid metabolites in this review.

2.7.3.1. Trimethylamine-oxide (TMAO)

Many studies have been conducted to investigate metabolites associated with the pathogenesis of CVD, and one of the most studied and reviewed metabolites to date is TMAO. We briefly report the association between TMAO and CVD since there are several well written published reviews which provide detailed reviews of TMAO (Witkowski et al., 2020). Dietary choline, and L-carnitine are known to be converted to trimethylamine (TMA) by microbial enzymes (TMA lyase) contained in the genomes of specific gut microbiota. TMA is then absorbed in the intestine, delivered to the liver via portal vein, and then converted to TMAO by hepatic flavin monooxygenase 3 (FMO3) (Bennett et al., 2013b). In subsequent studies combining data from more than 4,000 subjects who underwent coronary angiography, TMAO elevation is associated with death, MI, and stroke over a 3-year period (Tang et al., 2013; Wang et al., 2014b). These prognostic effects also have been evaluated in patients with a history of diabetes (Tang et al., 2017), chronic kidney disease, heart failure, MI, and peripheral arterial disease (Witkowski et al., 2020).

Table 2.2. Mechanism of microbiota-related metabolites on CVD

Metabolites	Foods rich in metabolites	Bacteria producing metabolites	Mechanism	Study group	References
TMAO	Red meat, eggs, fish, poultry	*Anaerococcus*, *Clostridium*, *Desulfovibrio*, *Edwardsiella*, *Proteus*, *Providencia*, and others(Qi et al., 2018)	Cholesterol accumulation	ApoE$^{-/-}$ mice	(Koeth et al., 2013b)
			Foam cell formation	ApoE$^{-/-}$ mice	(Koeth et al., 2013b)
			Platelet hyper-reactivity	Human and germ-free mice	(Zhu et al., 2016)
			Vascular endothelial dysfunction	Human and germ-free mice	(Zhu et al., 2016)
			Fibrosis and remodeling	Human and germ-free mice	(Zhu et al., 2016)
			Activation of PERK-FoxO1 signaling	ob/ob mice, In vitro studies (primary rat hepatocytes and HEK239T cells)	(Chen et al., 2019)
SCFAs	Fermented foods (cheese, butter, yoghurt)	*Anaerostipes*, *Blautia*, *Coprococcus*, *Eubacterium*, *Faecalibacterium*, *Marvinbryantia*, *Megasphaera*, *Roseburia*, *Ruminococcus*, and others(Morrison and Preston, 2016; Tan et al., 2014)	Lowering blood pressure by binding SCFA binding G protein-coupled receptor (GPR41, GPR43, GPR109A)	GPR41$^{-/-}$ mice, GPR43/GPR109A$^{-/-}$ mice, and germ-free mice	(Kaye et al., 2020; Natarajan et al., 2016)
Tryptophan metabolites					
Indole, ILA, IPA, IAA	Red meat, eggs, fish, poultry	*Bacteroides*, *Bifidobacterium*, *Clostridium*, *Lactobacillus*, *Peptostreptococcus*, *Ruminococcus*, *Ruminiclostridium*, and others(Dodd et al., 2017; Russell et al., 2013)	Act on AHR found in intestinal immune cells and thereby alter innate and adaptive immune responses	GF and conventional mouse	(Krishnan et al., 2018)
Indoxyl sulfate			Induces expression of proinflamatory cytokine IL-6	Hypertensive rats	(Adelibieke et al., 2014)
			Induces expression of monocyte chemo-attractant protein-1	Cell culture	(Watanabe et al., 2013)
			Caused endothelial cell senescence,	Cell culture	(Koizumi et al., 2014)
			Proliferation and migration of vascular smooth muscle cells	Cell culture	(Shimizu et al., 2009)

TMAO = Trimethylamine N-oxide, SCFA = short chain fatty acid, IPA = indolepropionic acid, ILA = indolelactic acid, IAA = indoleacetic acid,

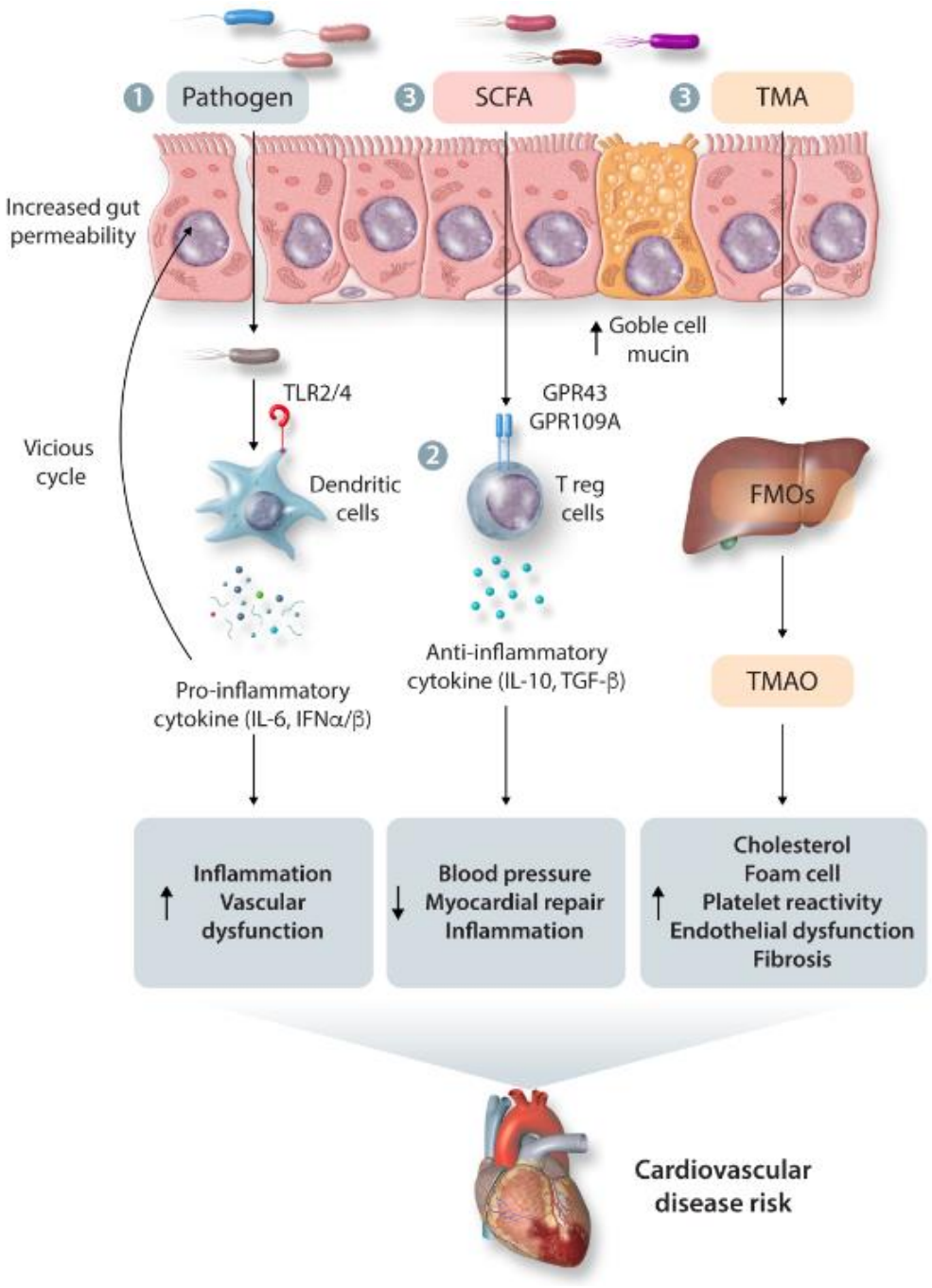


Figure 2.2. Overview of potential mechanisms of microbiota-host interactions and CVD. 1-Pathogenic bacteria or bacterial components can enter the host circulation resulting in endotoxemia and inflammatory response. 2-local and systemic inflammatory processes can affect CVD development 3- Metabolites such as SCFAs and TMA can either bind host receptors or be further metabolized to pro- or anti- CVD molecules. SCFAs, short-chain fatty acids; GPR, G protein-coupled receptor; TLR, toll-like receptor; TMA, trimethylamine; TMAO, trimethylamine N-oxide; FMOs, flavin-containing monooxygenases; IL, interleukin; IFN, interferon; TGF-B, transforming growth factor beta.

Although the association between TMAO levels and various CAD events has been reproduced in many studies, there are studies where TMAO is not associated with CAD. For example, a recent study showed that TMAO levels were not associated with atherosclerosis in the Framingham Heart Study Offspring cohort (1215 individuals) or supporting animal studies (Koay et al., 2020), or CARDIA (Meyer et al., 2016). One small clinical trial examined FMT of low TMAO producing vegan donors into subjects at risk of CVD but failed to observe a significant reduction in TMAO levels (Smits et al., 2018). Understanding what role, if any differences in the microbiota or diet, play in these disparate results remains to be determined.

Evidence supporting the role of TMAO in the development of hypertension is not yet clear. Preclinical studies have shown that experimental hypertensive rats have higher intestinal permeability and portal blood TMA level in the colon tissue (Jaworska et al., 2017), and TMAO treatment increased plasma aquaporin-2 concentration which elicits greater water reabsorption, and eventually leads to hypertension (Liu et al., 2019b). Apart from these animal studies, evidence has been established in humans through a systematic review that high TMAO plasma levels are associated with high blood pressure risk in 8 studies with 11,750 individuals and 6176 hypertensive cases (Ge et al., 2020). However, the studies included in this systematic review recruited most of the participants were from the United States. Further large-scale prospective cohorts are expected to characterize the association, especially the causality in the general population.

2.7.3.2. Short-Chain Fatty Acids

The main products of microbial enzyme reaction of dietary fiber are SCFAs, such as acetate, butyrate, and propionate. The main butyrate producers in the human colon are *Firmicutes* (phylum), with *Lachnospiraceae* (family), and the *Ruminococcaceae* (family) the two

most abundant groups (Louis and Flint, 2009, 2017). An additional pathway in lactic acid-utilizing bacteria exists where lactate and acetate are converted to butyrate (Louis and Flint, 2009). A number of other phyla produce butyrate and the efficiency of this production may reflect the expression of specific genes such as butyryl-CoA transferase, butyryl-CoA dehydrogenase, and butyrate kinase (Vital et al., 2014). Two species are well characterized with regard to SCFA production: *Bifidobacterium* species produce acetate and lactate (Riviere et al., 2016) and *Akkermansia muciniphila* produces acetate and propionate (Derrien et al., 2004; Louis and Flint, 2017).

It is known that the most direct route through which SFCA modulates the risk of CVD is the regulation of blood pressure. Initial clinical intervention study showed that fiber intake reduced blood pressure and that SCFAs are involved in blood pressure control (Streppel et al., 2005). SCFA metabolites lower blood pressure by modulating the SCFA binding G protein-coupled receptor (GPR) 41 or olfactory receptor 78 *in vitro* and *in vivo* (Kim et al., 2018a; Natarajan et al., 2016; Pluznick et al., 2013). Supplementing the diet with SCFAs protects against the development of hypertension and involves the SCFA receptor GPR43/GPR109A, which also regulates the abundance of Treg cells in mice (Bartolomaeus et al., 2019; Kaye et al., 2020; Marques et al., 2017). Thus, SCFAs may regulate cardiovascular homeostasis by activating receptors in the cells of the cardiovascular system. Supporting these mechanistic studies are data from a controlled trial showed that butyrate (600 mg/day) significantly reduced diastolic blood pressure that measured after a 10-min rest period in 15 patients with type 2 diabetes (Roshanravan et al., 2017). Although SCFAs are among the most frequently published intestinal microbial-derived metabolites, more clinical trials and mechanistic studies are needed to validate the effects of SCFAs on CVD and its risk factors.

2.7.3.3. Other Microbial Metabolites

Identification of microbial derived metabolites that affect host physiology is an active area of investigation. There is particular interest in small molecules derived from dietary tryptophan (Hayashi et al., 2018; Yu et al., 2017). In the gut, tryptophan is catabolized by a variety of pathways associated with several classes of gut bacteria including *Lactobacillus, Bacteroides, Bifidobacterium,* and *Clostridium* to produce a number of metabolites such as tryptamine, indole, indolelactic acid (ILA), indolepropionic acid (IPA), indoleacetic acid (IAA), indolealdehyde (IAld), and metabolite 3-methylindole (skatole) (Cervantes-Barragan et al., 2017; Dodd et al., 2017; Russell et al., 2013; Zelante et al., 2013). These tryptophan metabolites have diverse biological effects including enhancing the intestinal epithelial barrier, stimulate gastrointestinal motility, help the secretion of gut hormones, exert anti-inflammatory and anti-oxidative effects in the systemic circulation, modulate gut microbial composition (Roager and Licht, 2018). Some of these effects may be through the aryl hydrocarbon receptor (AHR) and thereby alter innate and adaptive immune responses in a ligand-specific fashion (Krishnan et al., 2018) which could modulate CVD risk. Not all tryptophan metabolites are associated with beneficial effects on gut health or CVD risk. In particular, indoxyl sulfate is associated with aortic calcification (Barreto et al., 2009), increased carotid intima-media thickness (Sato et al., 2012), increases expression of cytokines in vascular cells (Adelibieke et al., 2014). Understanding how compositional differences in the microbiota or if specific bacteria module metabolism of tryptophan metabolites may further shed light on the underlying mechanisms by which taxa affect CVD risk.

2.8. Therapeutic Potential for Modulating Microbiota on CVD Risk

The association between altered composition of intestinal microbiota in CVD patients, production of microbial metabolites, and CVD risk mentioned above suggest that gut microbiota may be a significant modulator of CVD, and their relationship has become a potential target for new therapeutics (**Figure 2.3**). We review a number of potential strategies to module the microbiota and CVD risk including diet, inhibition of microbial pathways, and fecal microbiota transplants.

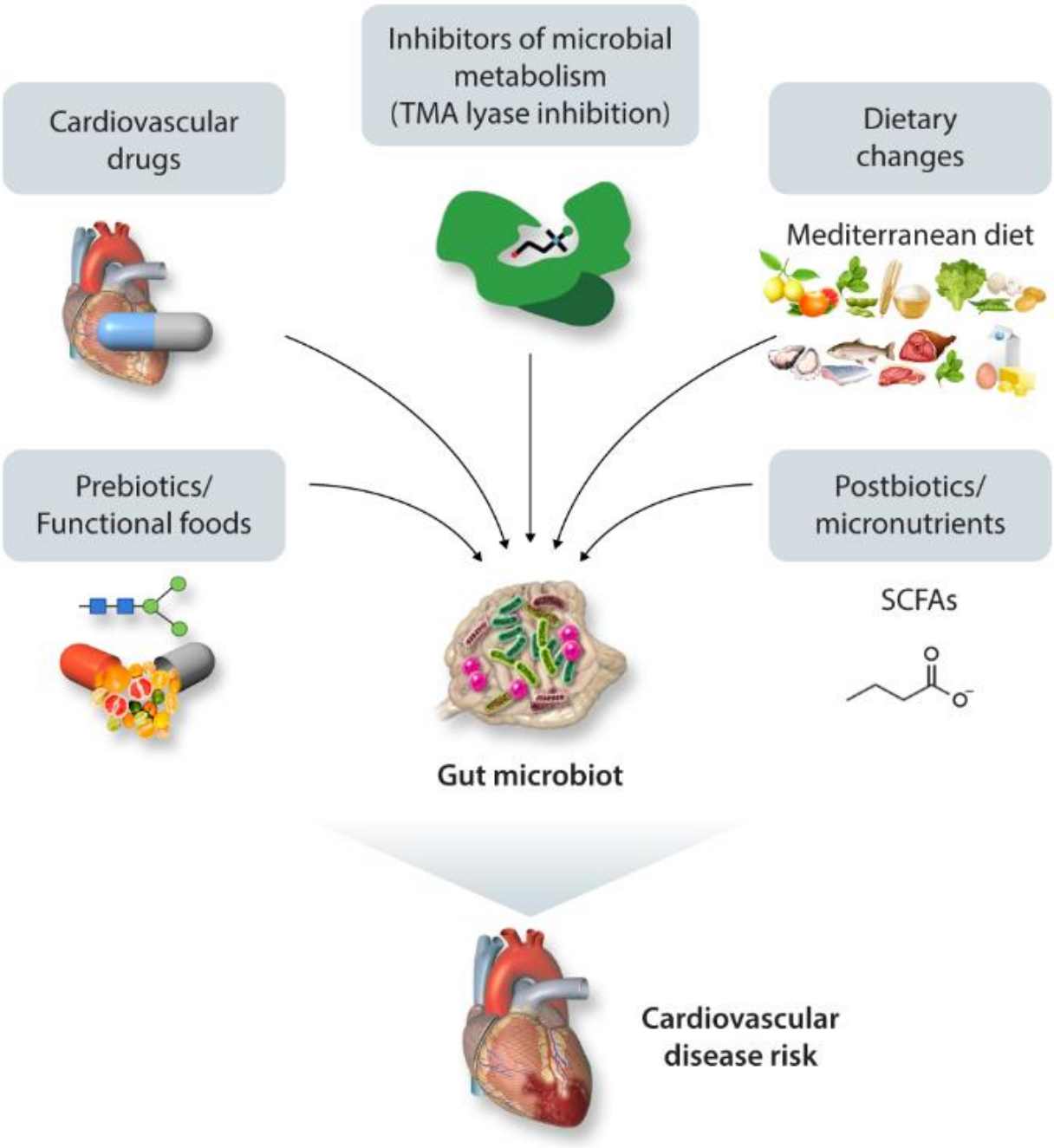


Figure 2.3. Therapeutic interventions for improving CVD. Current strategies for improving cardiovascular disease by manipulating intestinal microbiota, including bacterial TMA lyase enzyme inhibitors, fecal microbial transplantation, prebiotics, probiotics, and dietary interventions.

2.8.1. Diet and Prebiotics in the CVD-Microbiota Axis

Diet has been considered to be the most direct driving factor for gut microbiota composition (David et al., 2014; Duncan et al., 2007; Louis et al., 2007; Zhang et al., 2010a). In many cases, diet is a factor modulating both CAD and hypertension (Jama et al., 2019) and gut microbiota. Epidemiological, clinical, and experimental studies have demonstrated that diet and nutrition play a central role in the prevention of CVD (Torres et al., 2015). Based on these data the American Heart Association has made formal dietary recommendations (https://www.heart.org/en/healthy-living/healthy-eating/eat-smart/nutrition-basics/aha-diet-and-lifestyle-recommendations). Some of these specific recommendations are now known to affect the microbiota.

An example of these are recommendations to reduce salt intake as certain humans and model organisms are sensitive to a high-salt diet (Ha, 2014; Mell et al., 2015). We now appreciate that high salt intake influences gut microbial diversity in rodents and humans (Bier et al., 2018; Wang et al., 2017; Wilck et al., 2017). High salt intake altered microbiota diversity and increased the abundance of *Erwinia* genus and Corynebacteriaceae family in mice (Bier et al., 2018) and depleted the abundance of *Lactobacillus murinus* in human. Mechanistically, a high salt diet promotes local and systemic tissue inflammation via increases of pro-inflammatory cytokines and increased gut permeability in both human and animal studies (Kleinewietfeld et al., 2013; Wilck et al., 2017; Yi et al., 2015). Some of the effects of dietary salt may be attributed to specific bacteria such as *Bacteroides fragilis*, which through a variety of intermediate metabolic effects ultimately activates the mineralocorticoid receptor and increases blood pressure (Yan et al., 2020).

In addition to specific dietary components that affect CVD risk and the microbiota, there is evidence that dietary patterns also play a role. For example, one of the most studied diets in CVD research is the Mediterranean diet (MeD). Epidemiological studies found that MeD has a cardio-protective effect compared to western diets (De Lorgeril et al., 1994; Estruch et al., 2013). Several studies (De Filippis et al., 2016; Garcia-Mantrana et al., 2018) have demonstrated that MeD elicits beneficial microbiota profiles and microbial metabolite production. For example, one study (De Filippis et al., 2016) found that higher consumption of MeD is associated with increased levels of *Prevotella*, fiber degrading *Firmicutes*, fecal SCFAs, and lower urinary TMAO compared to a western diet. Similarly, another study (Garcia-Mantrana et al., 2018) reported that closer adherence to the Mediterranean dietary pattern and greater consumption of plant-based nutrients such as vegetable proteins and polysaccharides were associated with a lower ratio of Firmicutes: Bacteroidetes and *Streptococcus*; higher *Catenibacterium, Bifidobacterium*, and fecal SCFAs. One recent study (Pellegrini et al., 2020) found that MeD with probiotics containing *Bifidobacterium longum* and *Lactobacillus rhamnosus* increased gut microbial diversity and decreased Bacteroidetes-to-Firmicutes ratio along with some other health benefits including improved BMI, fasting glucose, and homeostasis, which was not fully observed in MeD alone indicating that at least a part of the health benefit of MeD depends on the gut microbial composition.

Dietary fiber is an important macronutrient in the context of gut microbiota and CVD. Large population-based observational studies (Bazzano et al., 2003) found that a diet containing high dietary fiber was associated with a reduced CVD risk, which was largely mediated via the reduction of LDL-cholesterol. Clinical trials found that ingestion of soluble fiber (2-10 g/day) was associated with a significant 7% LDL-cholesterol reduction (Brown et al., 1999) in a dose-

dependent manner (Anderson et al., 2000). Several gut bacteria utilize dietary fiber and produce SCFAs which prevent CVD as discussed earlier in this manuscript. In addition to SCFAs production, dietary soluble fibers bind to the bile acids in the gut and blocked reabsorption of the bile acids, which leads to increased bile acid synthesis in the liver, and ultimately increased LDL clearance from blood (Andersson et al., 2002; Kritchevsky and Story, 1974).

Growing evidence showing that prebiotics supplementation reduces CVD, specifically blood pressure and atherosclerosis, by manipulating the gut microbiota, which supports prebiotic interventions to prevent or treat CVD. Prebiotics are the non-digestible food ingredients that beneficially affect the host by selectively stimulating the growth and/or the activity of one or a limited number of bacterial species already resident in the colon (Gibson and Roberfroid, 1995). Inulin, a linear β-2,1 fructosyl-fructose polydisperse carbohydrate material, feeding decreased atherosclerosis in the aortic root of mice and normalized the altered microbial abundance of *Bifidobacterium, Lactobacillus, Akkermansia, Allobaculum,* and *Coprococcus* (Hoffman et al., 2020). These results are supported by another study where inulin supplementation increased *Akkermansia* and *Bifidobacterium* abundance, decreased bacterial taxa involved in secondary bile acid metabolism, and reversed endothelial dysfunction (Catry et al., 2018). β-glucan is a glucose polysaccharide that can sequester cholesterol, scavenges reactive oxygen species, and produces SCFAs when digested by gut microbiota (Nakashima et al., 2018). In addition to serving as an energy source for gut bacteria, β-glucans are immunostimulatory through activation of β-glucan receptors, such as dectin-1 or CR3 on the intestinal macrophages (Brown and Gordon, 2001; Chan et al., 2009). Oat β-glucan supplementation increased high-density lipoprotein (HDL)-cholesterol and decreased plasma triglyceride (TG) and atherosclerosis alone with enrichment of the genus *Akkermansia* in the gut (Ryan et al., 2017).

2.8.2. Targeting the Microbiota to Affect CVD using Small Molecules

The approach of manipulating intestinal microbial communities and their metabolic pathways has not yet reached clinical practice, but some studies show promising results. Targeting microbial enzymes (TMA lyases) that convert nutrients such as choline or carnitine into TMA regulates TMAO levels and experimental studies have found several chemical compounds that can trigger the modulatory effects on CVD. One example is 3,3-dimethyl-1-butanol (DMB), which is a TMA lyase inhibitor found in natural products such as olive oil, which can decrease plasma TMAO levels without perturbing microbial cell viability in *in-vivo* mouse models (Chen et al., 2016b; Wang et al., 2015). Mice receiving DMB have been shown to have reduced atherosclerotic lesion, attenuated foam cell formation, and alleviated progression of CVD (Wang et al., 2015). Chemically synthesized compounds have been used to inhibit TMA lyase in mice and these molecules also showed the selectively targeting and accumulating in the gut microbiota, allowing for sustained inhibition of TMA lyase in the host (Roberts et al., 2018).

In addition to TMAO there are a number of other microbial pathways whose modulation or inhibition may affect CVD risk. One possible pathway to target is the microbial cholesterol dehydrogenase enzyme (*ism*A gene), which converts cholesterol to the sterol coprostanol (Kriaa et al., 2019). Coprostanol is not absorbed as efficiently as cholesterol in the gastrointestinal tract and thus this pathway contributes to lowering blood cholesterol levels (Kenny et al., 2020). Individuals carrying coprostanol-forming microorganisms have significantly lower cholesterol levels in stools and lower plasma total cholesterol, with effects comparable to those attributed to variations in human genes involved in lipid homeostasis. Therefore, altering the abundance of the bacteria-containing *ism*A or increasing its expression could reduce CVD risk by lowering intestinal and serum cholesterol levels.

An alternative approach is to use small molecules, such as cyclic d,l-α-peptides, which modulate the growth of specific bacteria. Initial experiments have been promising as these compounds remodel the microbiota of high-fat diet-fed Ldlr-/- mice to resemble a low-fat diet gut microbiota and inhibited atherosclerosis development (Chen et al., 2020). The effects of this modulation are broad and include decreased plasma cholesterol, suppressed the pro-inflammatory cytokines such as IL-6, TNFα, and IL-1B, and altered levels of SCFAs and bile acids in the feces and plasma. Identification of similar compounds that affect host phenotype by targeting microbial processes is an exciting and active area of research.

2.8.3. Fecal Microbiota Transplantation

While most often used for mechanistic studies, fecal microbiota transplantation (FMT) has also been used as a therapy for patients with *Clostridium difficile* infection (Bakken et al., 2011) and ulcerative colitis (Moayyedi et al., 2015). This process includes the collection of stools collected from healthy donors or the recipients themselves (self FMT) prior to administration into the intestine of patients suffering from disease or related dysbiosis. These studies are complex and to date have not been extensively studied with regards to CVD endpoints. While the clinical utility of such approach is under debate for CVD (Woodworth et al., 2017), the therapeutic effect of FMT has been reported effective against insulin resistance and small intestinal permeability (Craven et al., 2020; de Groot et al., 2020).

2.9. Concluding Remarks and Future Perspectives

Many studies have confirmed the link between gut microbiota and CVD and we are beginning to understand the underlying mechanisms of these associations. One exciting aspect of this work is related to metabolomics. In some sense, the gut microbiota is an intermediate trait between diet (environmental factor) and CVD risk (clinical trait) that produces metabolites, some

of which play an important role in the pathogenesis of CVD. Recent studies have confirmed that specific microbial taxa are associated with CVD, and that various gut-derived metabolites, including TMAO or SCFAs, can promote or attenuate CVD. However, much still remains to be investigated. For example, only bacterial communities are being studied extensively and other members in gut microbiota such as virus, fungal, or archaea are not widely studied and thus their roles in human disease remain underappreciated. Therefore, combining these members in the analysis of the microbiota-CVD association might open us a new door for potential therapeutics.

Perhaps the most important questions remain -- how will these exciting results be applied clinically (if at all)? Studies utilizing the LifeLines-DEEP population cohort have identified that candidate SNPs explain 3-7% of the variation in HDL and triglycerides while adding 16S microbial diversity explains another 4-6% of the variation in HDL and triglycerides (Fu et al., 2015). Interestingly, when genetic and microbiome data are combined they explain significantly more variation in HDL and triglycerides. Thus, there is an indication that obtaining microbiota data may further characterize patients as we move towards a model of precision medicine. For example, the application of machine learning to datasets containing gut microbiota, genetics, and diet together (Berry et al., 2020) predicted better postprandial plasma TG, glucose, and insulin responses. In addition to direct causal pathways, it is likely that specific bacteria or microbiota components will be casually implicated or simply act as moderating variables for both genetic studies (Hughes et al., 2020b; Yang et al., 2018) and drug therapies (Tuteja and Ferguson, 2019; Vieira-Silva et al., 2020). Thus, the microbiota remains a critical part of the movement towards personalized medicine. Ultimately utilizing microbiota data that has been vetted to be clinically relevant may in fact refine our risk predictions and therapeutic interventions.

2.10. Acknowledgments

This research was supported in part by NIH grant 5R01HL128572 (BJB), USDA project 2032-51530-025-00D (BJB). The USDA is an equal opportunity employer. Figures were created using Biorender.com.

2.11. References

Adelibieke, Y., Yisireyili, M., Ng, H.-Y., Saito, S., Nishijima, F., and Niwa, T. (2014). Indoxyl sulfate induces IL-6 expression in vascular endothelial and smooth muscle cells through OAT3-mediated uptake and activation of AhR/NF-κB pathway. Nephron Experimental Nephrology *128*, 1-8.

Adnan, S., Nelson, J.W., Ajami, N.J., Venna, V.R., Petrosino, J.F., Bryan, R.M., Jr., and Durgan, D.J. (2017). Alterations in the gut microbiota can elicit hypertension in rats. Physiol Genomics *49*, 96-104.

Amos, G.C.A., Logan, A., Anwar, S., Fritzsche, M., Mate, R., Bleazard, T., and Rijpkema, S. (2020). Developing standards for the microbiome field. Microbiome *8*, 98.

Anderson, J.W., Allgood, L.D., Lawrence, A., Altringer, L.A., Jerdack, G.R., Hengehold, D.A., and Morel, J.G. (2000). Cholesterol-lowering effects of psyllium intake adjunctive to diet therapy in men and women with hypercholesterolemia: meta-analysis of 8 controlled trials. The American journal of clinical nutrition *71*, 472-479.

Andersson, M., Ellegard, L., and Andersson, H. (2002). Oat bran stimulates bile acid synthesis within 8 h as measured by 7alpha-hydroxy-4-cholesten-3-one. Am J Clin Nutr *76*, 1111-1116.

Atarashi, K., Tanoue, T., Shima, T., Imaoka, A., Kuwahara, T., Momose, Y., Cheng, G., Yamasaki, S., Saito, T., Ohba, Y., *et al.* (2011). Induction of colonic regulatory T cells by indigenous Clostridium species. Science *331*, 337-341.

Bakken, J.S., Borody, T., Brandt, L.J., Brill, J.V., Demarco, D.C., Franzos, M.A., Kelly, C., Khoruts, A., Louie, T., Martinelli, L.P., *et al.* (2011). Treating Clostridium difficile infection with fecal microbiota transplantation. Clin Gastroenterol Hepatol *9*, 1044-1049.

Barreto, F.C., Barreto, D.V., Liabeuf, S., Meert, N., Glorieux, G., Temmar, M., Choukroun, G., Vanholder, R., Massy, Z.A., and Group, E.U.T.W. (2009). Serum indoxyl sulfate is associated with vascular disease and mortality in chronic kidney disease patients. Clinical Journal of the American Society of Nephrology *4*, 1551-1558.

Bartolomaeus, H., Balogh, A., Yakoub, M., Homann, S., Marko, L., Hoges, S., Tsvetkov, D., Krannich, A., Wundersitz, S., Avery, E.G., *et al.* (2019). Short-Chain Fatty Acid Propionate Protects From Hypertensive Cardiovascular Damage. Circulation *139*, 1407-1421.

Bazzano, L.A., He, J., Ogden, L.G., Loria, C.M., and Whelton, P.K. (2003). Dietary fiber intake and reduced risk of coronary heart disease in US men and women: the National Health and Nutrition Examination Survey I Epidemiologic Follow-up Study. Archives of internal medicine *163*, 1897-1904.

Bennett, B.J., de Aguiar Vallim, T.Q., Wang, Z., Shih, D.M., Meng, Y., Gregory, J., Allayee, H., Lee, R., Graham, M., Crooke, R., *et al.* (2013). Trimethylamine-N-oxide, a metabolite associated with atherosclerosis, exhibits complex genetic and dietary regulation. Cell Metab *17*, 49-60.

Berry, S.E., Valdes, A.M., Drew, D.A., Asnicar, F., Mazidi, M., Wolf, J., Capdevila, J., Hadjigeorgiou, G., Davies, R., Al Khatib, H., *et al.* (2020). Human postprandial responses to food and potential for precision nutrition. Nat Med *26*, 964-973.

Bier, A., Braun, T., Khasbab, R., Di Segni, A., Grossman, E., Haberman, Y., and Leibowitz, A. (2018). A High Salt Diet Modulates the Gut Microbiota and Short Chain Fatty Acids Production in a Salt-Sensitive Hypertension Rat Model. Nutrients *10*.

Blessing, E., Campbell, L.A., Rosenfeld, M.E., Chesebro, B., and Kuo, C.C. (2005). A 6 week course of azithromycin treatment has no beneficial effect on atherosclerotic lesion development in apolipoprotein E-deficient mice chronically infected with Chlamydia pneumoniae. J Antimicrob Chemother *55*, 1037-1040.

Boers, S.A., Jansen, R., and Hays, J.P. (2019). Understanding and overcoming the pitfalls and biases of next-generation sequencing (NGS) methods for use in the routine clinical microbiological diagnostic laboratory. Eur J Clin Microbiol Infect Dis *38*, 1059-1070.

Brandsma, E., Kloosterhuis, N.J., Koster, M., Dekker, D.C., Gijbels, M.J.J., van der Velden, S., Rios-Morales, M., van Faassen, M.J.R., Loreti, M.G., de Bruin, A., *et al.* (2019). A Proinflammatory Gut Microbiota Increases Systemic Inflammation and Accelerates Atherosclerosis. Circ Res *124*, 94-100.

Brown, G.D., and Gordon, S. (2001). A new receptor for β-glucans. Nature *413*, 36-37.

Brown, L., Rosner, B., Willett, W.W., and Sacks, F.M. (1999). Cholesterol-lowering effects of dietary fiber: a meta-analysis. The American journal of clinical nutrition *69*, 30-42.

Brumfield, K.D., Huq, A., Colwell, R.R., Olds, J.L., and Leddy, M.B. (2020). Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. PLoS One *15*, e0228899.

Catry, E., Bindels, L.B., Tailleux, A., Lestavel, S., Neyrinck, A.M., Goossens, J.F., Lobysheva, I., Plovier, H., Essaghir, A., Demoulin, J.B., *et al.* (2018). Targeting the gut microbiota with inulin-type fructans: preclinical demonstration of a novel approach in the management of endothelial dysfunction. Gut *67*, 271-283.

Cervantes-Barragan, L., Chai, J.N., Tianero, M.D., Di Luccia, B., Ahern, P.P., Merriman, J., Cortez, V.S., Caparon, M.G., Donia, M.S., and Gilfillan, S. (2017). Lactobacillus reuteri induces gut intraepithelial CD4+ CD8αα+ T cells. Science *357*, 806-810.

Chan, G.C.-F., Chan, W.K., and Sze, D.M.-Y. (2009). The effects of β-glucan on human immune and cancer cells. Journal of hematology & oncology *2*, 1-11.

Chen, M.L., Yi, L., Zhang, Y., Zhou, X., Ran, L., Yang, J., Zhu, J.D., Zhang, Q.Y., and Mi, M.T. (2016). Resveratrol Attenuates Trimethylamine-N-Oxide (TMAO)-Induced Atherosclerosis by Regulating TMAO Synthesis and Bile Acid Metabolism via Remodeling of the Gut Microbiota. mBio *7*, e02210-02215.

Chen, P.B., Black, A.S., Sobel, A.L., Zhao, Y., Mukherjee, P., Molparia, B., Moore, N.E., Aleman Muench, G.R., Wu, J., Chen, W., *et al.* (2020). Directed remodeling of the mouse gut microbiome inhibits the development of atherosclerosis. Nat Biotechnol.

Chen, S., Henderson, A., Petriello, M.C., Romano, K.A., Gearing, M., Miao, J., Schell, M., Sandoval-Espinola, W.J., Tao, J., Sha, B., *et al.* (2019). Trimethylamine N-Oxide Binds and Activates PERK to Promote Metabolic Dysfunction. Cell Metab *30*, 1141-1151 e1145.

Cluff, L.E., Reynolds, R.C., Page, D.L., and Breckenridge, J.L. (1968). Staphylococcal Bacteremia and Altered Host Resistance. Ann Intern Med *69*, 859-873.

Craven, L., Rahman, A., Nair Parvathy, S., Beaton, M., Silverman, J., Qumosani, K., Hramiak, I., Hegele, R., Joy, T., Meddings, J., *et al.* (2020). Allogenic Fecal Microbiota Transplantation in Patients With Nonalcoholic Fatty Liver Disease Improves Abnormal Small Intestinal Permeability. American Journal of Gastroenterology *Publish Ahead of Print*.

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

De Filippis, F., Pellegrini, N., Vannini, L., Jeffery, I.B., La Storia, A., Laghi, L., Serrazanetti, D.I., Di Cagno, R., Ferrocino, I., Lazzi, C., *et al.* (2016). High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. Gut *65*, 1812-1821.

de Groot, P., Scheithauer, T., Bakker, G.J., Prodan, A., Levin, E., Khan, M.T., Herrema, H., Ackermans, M., Serlie, M.J.M., de Brauw, M., *et al.* (2020). Donor metabolic characteristics drive effects of faecal microbiota transplantation on recipient insulin sensitivity, energy expenditure and intestinal transit time. Gut *69*, 502-512.

De Lorgeril, M., Renaud, S., Salen, P., Monjaud, I., Mamelle, N., Martin, J., Guidollet, J., Touboul, P., and Delaye, J. (1994). Mediterranean alpha-linolenic acid-rich diet in secondary prevention of coronary heart disease. The Lancet *343*, 1454-1459.

de Oliveira, C., Watt, R., and Hamer, M. (2010). Toothbrushing, inflammation, and risk of cardiovascular disease: results from Scottish Health Survey. Bmj *340*.

Depommier, C., Everard, A., Druart, C., Plovier, H., Van Hul, M., Vieira-Silva, S., Falony, G., Raes, J., Maiter, D., Delzenne, N.M., *et al.* (2019). Supplementation with Akkermansia muciniphila in overweight and obese human volunteers: a proof-of-concept exploratory study. Nat Med *25*, 1096-1103.

Derrien, M., Vaughan, E.E., Plugge, C.M., and de Vos, W.M. (2004). Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. International Journal of Systematic and Evolutionary Microbiology *54*, 1469-1476.

Dinakaran, V., Rathinavel, A., Pushpanathan, M., Sivakumar, R., Gunasekaran, P., and Rajendhran, J. (2014). Elevated levels of circulating DNA in cardiovascular disease patients: metagenomic profiling of microbiome in the circulation. PLoS One *9*, e105221.

Dodd, D., Spitzer, M.H., Van Treuren, W., Merrill, B.D., Hryckowian, A.J., Higginbottom, S.K., Le, A., Cowan, T.M., Nolan, G.P., and Fischbach, M.A. (2017). A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. Nature *551*, 648-652.

Duncan, S.H., Belenguer, A., Holtrop, G., Johnstone, A.M., Flint, H.J., and Lobley, G.E. (2007). Reduced dietary intake of carbohydrates by obese subjects results in decreased concentrations of butyrate and butyrate-producing bacteria in feces. Appl Environ Microbiol *73*, 1073-1078.

Emoto, T., Yamashita, T., Sasaki, N., Hirota, Y., Hayashi, T., So, A., Kasahara, K., Yodoi, K., Matsumoto, T., Mizoguchi, T., *et al.* (2016). Analysis of Gut Microbiota in Coronary Artery Disease Patients: a Possible Link between Gut Microbiota and Coronary Artery Disease. J Atheroscler Thromb *23*, 908-921.

Estruch, R., Ros, E., Salas-Salvadó, J., Covas, M.-I., Corella, D., Arós, F., Gómez-Gracia, E., Ruiz-Gutiérrez, V., Fiol, M., and Lapetra, J. (2013). Primary prevention of cardiovascular disease with a Mediterranean diet. New England Journal of Medicine *368*, 1279-1290.

Feng, Q., Liu, Z., Zhong, S., Li, R., Xia, H., Jie, Z., Wen, B., Chen, X., Yan, W., Fan, Y., *et al.* (2016). Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease. Sci Rep *6*, 22525.

Fu, J., Bonder, M.J., Cenit, M.C., Tigchelaar, E.F., Maatman, A., Dekens, J.A., Brandsma, E., Marczynska, J., Imhann, F., Weersma, R.K., *et al.* (2015). The Gut Microbiome

Contributes to a Substantial Proportion of the Variation in Blood Lipids. Circ Res *117*, 817-824.
Furusawa, Y., Obata, Y., Fukuda, S., Endo, T.A., Nakato, G., Takahashi, D., Nakanishi, Y., Uetake, C., Kato, K., and Kato, T. (2013). Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. Nature *504*, 446-450.
Garcia-Mantrana, I., Selma-Royo, M., Alcantara, C., and Collado, M.C. (2018). Shifts on Gut Microbiota Associated to Mediterranean Diet Adherence and Specific Dietary Intakes on General Adult Population. Front Microbiol *9*, 890.
Ge, X., Zheng, L., Zhuang, R., Yu, P., Xu, Z., Liu, G., Xi, X., Zhou, X., and Fan, H. (2020). The Gut Microbial Metabolite Trimethylamine N-Oxide and Hypertension Risk: A Systematic Review and Dose-Response Meta-analysis. Adv Nutr *11*, 66-76.
Gibson, G.R., and Roberfroid, M.B. (1995). Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. The Journal of nutrition *125*, 1401-1412.
Gil-Cruz, C., Perez-Shibayama, C., De Martin, A., Ronchi, F., van der Borght, K., Niederer, R., Onder, L., Lutge, M., Novkovic, M., Nindl, V., *et al.* (2019). Microbiota-derived peptide mimics drive lethal inflammatory cardiomyopathy. Science *366*, 881-886.
Golebiewski, M., and Tretyn, A. (2020). Generating amplicon reads for microbial community assessment with next-generation sequencing. J Appl Microbiol *128*, 330-354.
Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J.T., *et al.* (2014). Human genetics shape the gut microbiome. Cell *159*, 789-799.
Gregory, J.C., Buffa, J.A., Org, E., Wang, Z., Levison, B.S., Zhu, W., Wagner, M.A., Bennett, B.J., Li, L., DiDonato, J.A., *et al.* (2015). Transmission of atherosclerosis susceptibility with gut microbial transplantation. J Biol Chem *290*, 5647-5660.
Ha, S.K. (2014). Dietary salt intake and hypertension. Electrolyte Blood Press *12*, 7-18.
Hassan, A., Din, A.U., Zhu, Y., Zhang, K., Li, T., Wang, Y., Xu, S., Lei, H., Yu, X., and Wang, G. (2020). Anti-atherosclerotic effects of Lactobacillus plantarum ATCC 14917 in ApoE(-/-) mice through modulation of proinflammatory cytokines and oxidative stress. Appl Microbiol Biotechnol *104*, 6337-6350.
Hayashi, C., Viereck, J., Hua, N., Phinikaridou, A., Madrigal, A.G., Gibson III, F.C., Hamilton, J.A., and Genco, C.A. (2011). Porphyromonas gingivalis accelerates inflammatory atherosclerosis in the innominate artery of ApoE deficient mice. Atherosclerosis *215*, 52-59.
Hayashi, T., Yamashita, T., Watanabe, H., Kami, K., Yoshida, N., Tabata, T., Emoto, T., Sasaki, N., Mizoguchi, T., Irino, Y., *et al.* (2018). Gut Microbiome and Plasma Microbiome-Related Metabolites in Patients With Decompensated and Compensated Heart Failure. Circ J *83*, 182-192.
Hoffman, J.B., Petriello, M.C., Morris, A.J., Mottaleb, M.A., Sui, Y., Zhou, C., Deng, P., Wang, C., and Hennig, B. (2020). Prebiotic inulin consumption reduces dioxin-like PCB 126-mediated hepatotoxicity and gut dysbiosis in hyperlipidemic Ldlr deficient mice. Environ Pollut *261*, 114183.
Hsu, C.C., Wei, T.S., Huang, C.C., and Chen, Y.M. (2017). Endotoxemia is associated with acute coronary syndrome in patients with end stage kidney disease. BMC Nephrol *18*, 235.

Hug, H., Mohajeri, M.H., and La Fata, G. (2018). Toll-Like Receptors - Regulators of the Immune Response in the Human Gut. Nutrients, 203.

Hughes, D.A., Bacigalupe, R., Wang, J., Ruhlemann, M.C., Tito, R.Y., Falony, G., Joossens, M., Vieira-Silva, S., Henckaerts, L., Rymenans, L., *et al.* (2020). Genome-wide associations of human gut microbiome variation and implications for causal inference analyses. Nat Microbiol.

Ibrahim, M., Behairy, M., El-Ashry, M., and Mostafa, A.E. (2018). Cardiovascular risk of circulating endotoxin level in prevalent hemodialysis patients. Egypt Heart J *70*, 27-33.

Jama, H.A., Beale, A., Shihata, W.A., and Marques, F.Z. (2019). The effect of diet on hypertensive pathology: is there a link via gut microbiota-driven immunometabolism? Cardiovasc Res *115*, 1435-1447.

Jaworska, K., Huc, T., Samborowska, E., Dobrowolski, L., Bielinska, K., Gawlak, M., and Ufnal, M. (2017). Hypertension in rats is associated with an increased permeability of the colon to TMA, a gut bacteria metabolite. PLoS One *12*, e0189310.

Jie, Z., Xia, H., Zhong, S.L., Feng, Q., Li, S., Liang, S., Zhong, H., Liu, Z., Gao, Y., Zhao, H., *et al.* (2017). The gut microbiome in atherosclerotic cardiovascular disease. Nat Commun *8*, 845.

Karbach, S.H., Schonfelder, T., Brandao, I., Wilms, E., Hormann, N., Jackel, S., Schuler, R., Finger, S., Knorr, M., Lagrange, J., *et al.* (2016). Gut Microbiota Promote Angiotensin II-Induced Arterial Hypertension and Vascular Dysfunction. J Am Heart Assoc *5*.

Karlsson, F.H., Fak, F., Nookaew, I., Tremaroli, V., Fagerberg, B., Petranovic, D., Backhed, F., and Nielsen, J. (2012). Symptomatic atherosclerosis is associated with an altered gut metagenome. Nat Commun *3*, 1245.

Kasahara, K., Krautkramer, K.A., Org, E., Romano, K.A., Kerby, R.L., Vivas, E.I., Mehrabian, M., Denu, J.M., Backhed, F., Lusis, A.J., *et al.* (2018). Interactions between Roseburia intestinalis and diet modulate atherogenesis in a murine model. Nat Microbiol *3*, 1461-1471.

Kaye, D.M., Shihata, W.A., Jama, H.A., Tsyganov, K., Ziemann, M., Kiriazis, H., Horlock, D., Vijay, A., Giam, B., Vinh, A., *et al.* (2020). Deficiency of Prebiotic Fiber and Insufficient Signaling Through Gut Metabolite-Sensing Receptors Leads to Cardiovascular Disease. Circulation *141*, 1393-1403.

Keijser, B., Zaura, E., Huse, S., Van der Vossen, J., Schuren, F., Montijn, R., Ten Cate, J., and Crielaard, W. (2008). Pyrosequencing analysis of the oral microflora of healthy adults. Journal of dental research *87*, 1016-1020.

Kelly, T.N., Bazzano, L.A., Ajami, N.J., He, H., Zhao, J., Petrosino, J.F., Correa, A., and He, J. (2016). Gut Microbiome Associates With Lifetime Cardiovascular Disease Risk Profile Among Bogalusa Heart Study Participants. Circ Res *119*, 956-964.

Kennedy, E.A., King, K.Y., and Baldridge, M.T. (2018). Mouse Microbiota Models: Comparing Germ-Free Mice and Antibiotics Treatment as Tools for Modifying Gut Bacteria. Front Physiol *9*, 1534.

Kenny, D.J., Plichta, D.R., Shungin, D., Koppel, N., Hall, A.B., Fu, B., Vasan, R.S., Shaw, S.Y., Vlamakis, H., Balskus, E.P., *et al.* (2020). Cholesterol Metabolism by Uncultured Human Gut Bacteria Influences Host Cholesterol Level. Cell Host Microbe *28*, 245-257 e246.

Kim, D., Hofstaedter, C.E., Zhao, C., Mattei, L., Tanes, C., Clarke, E., Lauder, A., Sherrill-Mix, S., Chehoud, C., Kelsen, J., *et al.* (2017). Optimizing methods and dodging pitfalls in microbiome research. Microbiome *5*, 52.

Kim, S., Goel, R., Kumar, A., Qi, Y., Lobaton, G., Hosaka, K., Mohammed, M., Handberg, E.M., Richards, E.M., Pepine, C.J., *et al.* (2018a). Imbalance of gut microbiome and intestinal epithelial barrier dysfunction in patients with high blood pressure. Clin Sci (Lond) *132*, 701-718.

Kim, T.T., Parajuli, N., Sung, M.M., Bairwa, S.C., Levasseur, J., Soltys, C.M., Wishart, D.S., Madsen, K., Schertzer, J.D., and Dyck, J.R.B. (2018b). Fecal transplant from resveratrol-fed donors improves glycaemia and cardiovascular features of the metabolic syndrome in mice. Am J Physiol Endocrinol Metab *315*, E511-E519.

Kleinewietfeld, M., Manzel, A., Titze, J., Kvakan, H., Yosef, N., Linker, R.A., Muller, D.N., and Hafler, D.A. (2013). Sodium chloride drives autoimmune disease by the induction of pathogenic TH17 cells. Nature *496*, 518-522.

Koay, Y.C., Chen, Y.C., Wali, J.A., Luk, A.W.S., Li, M., Doma, H., Reimark, R., Zaldivia, M.T.K., Habtom, H.T., Frank, A.E., *et al.* (2020). Plasma Levels of TMAO can be Increased with 'Healthy' and 'Unhealthy' Diets and Do Not Correlate with the Extent of Atherosclerosis but with Plaque Instability. Cardiovasc Res.

Koeth, R.A., Wang, Z., Levison, B.S., Buffa, J.A., Org, E., Sheehy, B.T., Britt, E.B., Fu, X., Wu, Y., Li, L., *et al.* (2013). Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis. Nature Medicine *19*, 576-585.

Koizumi, M., Tatebe, J., Watanabe, I., Yamazaki, J., Ikeda, T., and Morita, T. (2014). Aryl hydrocarbon receptor mediates indoxyl sulfate-induced cellular senescence in human umbilical vein endothelial cells. Journal of atherosclerosis and thrombosis, 23663.

Koren, O., Spor, A., Felin, J., Fak, F., Stombaugh, J., Tremaroli, V., Behre, C.J., Knight, R., Fagerberg, B., Ley, R.E., *et al.* (2011). Human oral, gut, and plaque microbiota in patients with atherosclerosis. Proc Natl Acad Sci U S A *108 Suppl 1*, 4592-4598.

Kriaa, A., Bourgin, M., Mkaouar, H., Jablaoui, A., Akermi, N., Soussou, S., Maguin, E., and Rhimi, M. (2019). Microbial reduction of cholesterol to coprostanol: An old concept and new insights. Catalysts *9*, 167.

Krishnan, S., Ding, Y., Saedi, N., Choi, M., Sridharan, G.V., Sherr, D.H., Yarmush, M.L., Alaniz, R.C., Jayaraman, A., and Lee, K. (2018). Gut microbiota-derived tryptophan metabolites modulate inflammatory response in hepatocytes and macrophages. Cell reports *23*, 1099-1111.

Kritchevsky, D., and Story, J.A. (1974). Binding of bile salts in vitro by nonnutritive fiber. J Nutr *104*, 458-462.

Kuo, C.C., Grayston, J.T., Campbell, L.A., Goo, Y.A., Wissler, R.W., and Benditt, E.P. (1995). Chlamydia pneumoniae (TWAR) in coronary arteries of young adults (15-34 years old). Proc Natl Acad Sci U S A *92*, 6911-6914.

Ley, R.E., Peterson, D.A., and Gordon, J.I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell *124*, 837-848.

Li, J., Lin, S., Vanhoutte, P.M., Woo, C.W., and Xu, A. (2016). Akkermansia Muciniphila Protects Against Atherosclerosis by Preventing Metabolic Endotoxemia-Induced Inflammation in Apoe-/- Mice. Circulation *133*, 2434-2446.

Li, J., Zhao, F., Wang, Y., Chen, J., Tao, J., Tian, G., Wu, S., Liu, W., Cui, Q., Geng, B., *et al.* (2017). Gut microbiota dysbiosis contributes to the development of hypertension. Microbiome *5*, 14.

Libby, P., Ridker, P.M., Hansson, G.K., and Leducq Transatlantic Network on, A. (2009). Inflammation in atherosclerosis: from pathophysiology to practice. J Am Coll Cardiol *54*, 2129-2138.

Liu, J., Li, T., Wu, H., Shi, H., Bai, J., Zhao, W., Jiang, D., and Jiang, X. (2019a). Lactobacillus rhamnosus GG strain mitigated the development of obstructive sleep apnea-induced hypertension in a high salt diet via regulating TMAO level and CD4(+) T cell induced-type I inflammation. Biomed Pharmacother *112*, 108580.

Liu, M., Han, Q., and Yang, J. (2019b). Trimethylamine-N-oxide (TMAO) increased aquaporin-2 expression in spontaneously hypertensive rats. Clin Exp Hypertens *41*, 312-322.

Louis, P., and Flint, H.J. (2009). Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS Microbiol Lett *294*, 1-8.

Louis, P., and Flint, H.J. (2017). Formation of propionate and butyrate by the human colonic microbiota. Environ Microbiol *19*, 29-41.

Louis, P., Scott, K.P., Duncan, S.H., and Flint, H.J. (2007). Understanding the effects of diet on bacterial metabolism in the large intestine. J Appl Microbiol *102*, 1197-1208.

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

Mandal, R.S., Saha, S., and Das, S. (2015). Metagenomic surveys of gut microbiota. Genomics Proteomics Bioinformatics *13*, 148-158.

Marques, F.Z., Jama, H.A., Tsyganov, K., Gill, P.A., Rhys-Jones, D., Muralitharan, R.R., Muir, J., Holmes, A., and Mackay, C.R. (2019). Guidelines for Transparency on Gut Microbiome Studies in Essential and Experimental Hypertension. Hypertension *74*, 1279-1293.

Marques, F.Z., Nelson, E., Chu, P.Y., Horlock, D., Fiedler, A., Ziemann, M., Tan, J.K., Kuruppu, S., Rajapakse, N.W., El-Osta, A., *et al.* (2017). High-Fiber Diet and Acetate Supplementation Change the Gut Microbiota and Prevent the Development of Hypertension and Heart Failure in Hypertensive Mice. Circulation *135*, 964-977.

Mell, B., Jala, V.R., Mathew, A.V., Byun, J., Waghulde, H., Zhang, Y., Haribabu, B., Vijay-Kumar, M., Pennathur, S., and Joe, B. (2015). Evidence for a link between gut microbiota and hypertension in the Dahl rat. Physiol Genomics *47*, 187-197.

Meyer, K.A., Benton, T.Z., Bennett, B.J., Jacobs, D.R., Jr., Lloyd-Jones, D.M., Gross, M.D., Carr, J.J., Gordon-Larsen, P., and Zeisel, S.H. (2016). Microbiota-Dependent Metabolite Trimethylamine N-Oxide and Coronary Artery Calcium in the Coronary Artery Risk Development in Young Adults Study (CARDIA). J Am Heart Assoc *5*.

Moayyedi, P., Surette, M.G., Kim, P.T., Libertucci, J., Wolfe, M., Onischi, C., Armstrong, D., Marshall, J.K., Kassam, Z., Reinisch, W., *et al.* (2015). Fecal Microbiota Transplantation Induces Remission in Patients With Active Ulcerative Colitis in a Randomized Controlled Trial. Gastroenterology *149*, 102-109 e106.

Morrison, D.J., and Preston, T. (2016). Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. Gut Microbes *7*, 189-200.

Muhlestein, J.B., Anderson, J.L., Hammond, E.H., Zhao, L., Trehan, S., Schwobe, E.P., and Carlquist, J.F. (1998). Infection with Chlamydia pneumoniae accelerates ... lerosis and treatment with azithromycin prevents it in a rabbit model. Circulation *97*, 633-636.

Nakashima, A., Yamada, K., Iwata, O., Sugimoto, R., Atsuji, K., Ogawa, T., Ishibashi-Ohgo, N., and Suzuki, K. (2018). beta-Glucan in Foods and Its Physiological Functions. J Nutr Sci Vitaminol (Tokyo) *64*, 8-17.

Natarajan, N., Hori, D., Flavahan, S., Steppan, J., Flavahan, N.A., Berkowitz, D.E., and Pluznick, J.L. (2016). Microbial short chain fatty acid metabolites lower blood pressure via endothelial G protein-coupled receptor 41. Physiol Genomics *48*, 826-834.

Norlander, A.E., Madhur, M.S., and Harrison, D.G. (2018). The immunology of hypertension. J Exp Med *215*, 21-33.

Palmu, J., Salosensaari, A., Havulinna, A.S., Cheng, S., Inouye, M., Jain, M., Salido, R.A., Sanders, K., Brennan, C., Humphrey, G.C., *et al.* (2020). Association Between the Gut Microbiota and Blood Pressure in a Population Cohort of 6953 Individuals. J Am Heart Assoc *9*, e016641.

Pankey, G.A. (1965). EFFECT OF VIRUSES ON THE CARDIOVASCULAR SYSTEM. Am J Med Sci.

Pant, S., Deshmukh, A., Gurumurthy, G.S., Pothineni, N.V., Watts, T.E., Romeo, F., and Mehta, J.L. (2014). Inflammation and atherosclerosis--revisited. J Cardiovasc Pharmacol Ther *19*, 170-178.

Pellegrini, M., Ippolito, M., Monge, T., Violi, R., Cappello, P., Ferrocino, I., Cocolin, L.S., De Francesco, A., Bo, S., and Finocchiaro, C. (2020). Gut microbiota composition after diet and probiotics in overweight breast cancer survivors: a randomized open-label pilot intervention trial. Nutrition *74*, 110749.

Peschel, T., Schönauer, M., Thiele, H., Anker, S.D., Schuler, G., and Niebauer, J. (2003). Invasive assessment of bacterial endotoxin and inflammatory cytokines in patients with acute heart failure. Eur J Heart Fail, 609-614.

Pevsner-Fischer, M., Blacher, E., Tatirovsky, E., Ben-Dov, I.Z., and Elinav, E. (2017). The gut microbiome and hypertension. Curr Opin Nephrol Hypertens *26*, 1-8.

Pluznick, J.L., Protzko, R.J., Gevorgyan, H., Peterlin, Z., Sipos, A., Han, J., Brunet, I., Wan, L.X., Rey, F., Wang, T., *et al.* (2013). Olfactory receptor responding to gut microbiota-derived signals plays a role in renin secretion and blood pressure regulation. Proc Natl Acad Sci U S A *110*, 4410-4415.

Pothineni, N.V.K., Subramany, S., Kuriakose, K., Shirazi, L.F., Romeo, F., Shah, P.K., and Mehta, J.L. (2017). Infections, atherosclerosis, and coronary heart disease. Eur Heart J *38*, 3195-3201.

Qi, J., You, T., Li, J., Pan, T., Xiang, L., Han, Y., and Zhu, L. (2018). Circulating trimethylamine N-oxide and the risk of cardiovascular diseases: a systematic review and meta-analysis of 11 prospective cohort studies. J Cell Mol Med *22*, 185-194.

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

Riviere, A., Selak, M., Lantin, D., Leroy, F., and De Vuyst, L. (2016). Bifidobacteria and Butyrate-Producing Colon Bacteria: Importance and Strategies for Their Stimulation in the Human Gut. Front Microbiol *7*, 979.

Roager, H.M., and Licht, T.R. (2018). Microbial tryptophan catabolites in health and disease. Nature communications *9*, 1-10.

Roberts, A.B., Gu, X., Buffa, J.A., Hurd, A.G., Wang, Z., Zhu, W., Gupta, N., Skye, S.M., Cody, D.B., Levison, B.S., *et al.* (2018). Development of a gut microbe-targeted nonlethal therapeutic to inhibit thrombosis potential. Nat Med *24*, 1407-1417.

Roshanravan, N., Mahdavi, R., Alizadeh, E., Jafarabadi, M.A., Hedayati, M., Ghavami, A., Alipour, S., Alamdari, N.M., Barati, M., and Ostadrahimi, A. (2017). Effect of Butyrate

and Inulin Supplementation on Glycemic Status, Lipid Profile and Glucagon-Like Peptide 1 Level in Patients with Type 2 Diabetes: A Randomized Double-Blind, Placebo-Controlled Trial. Horm Metab Res *49*, 886-891.

Ross, R. (1999). Atherosclerosis—an inflammatory disease. New England journal of medicine *340*, 115-126.

Round, J.L., and Mazmanian, S.K. (2010). Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. Proceedings of the National Academy of Sciences *107*, 12204-12209.

Russell, W.R., Duncan, S.H., Scobbie, L., Duncan, G., Cantlay, L., Calder, A.G., Anderson, S.E., and Flint, H.J. (2013). Major phenylpropanoid-derived metabolites in the human gut can arise from microbial fermentation of protein. Molecular nutrition & food research *57*, 523-535.

Ryan, P.M., London, L.E., Bjorndahl, T.C., Mandal, R., Murphy, K., Fitzgerald, G.F., Shanahan, F., Ross, R.P., Wishart, D.S., Caplice, N.M., *et al.* (2017). Microbiome and metabolome modifying effects of several cardiovascular disease interventions in apo-E(-/-) mice. Microbiome *5*, 30.

Saikku, P., Leinonen, M., Mattila, K., Ekman, M.R., Nieminen, M.S., Mäkelä, P.H., Huttunen, J.K., and Valtonen, V. (1988). Serological evidence of an association of a novel chlamydia, twar, with chronic coronary heart disease and acute myocardial infarction. Lancet *2*, 983-986.

Sanchez-Alcoholado, L., Castellano-Castillo, D., Jordan-Martinez, L., Moreno-Indias, I., Cardila-Cruz, P., Elena, D., Munoz-Garcia, A.J., Queipo-Ortuno, M.I., and Jimenez-Navarro, M. (2017). Role of Gut Microbiota on Cardio-Metabolic Parameters and Immunity in Coronary Artery Disease Patients with and without Type-2 Diabetes Mellitus. Front Microbiol *8*, 1936.

Sata, Y., Marques, F.Z., and Kaye, D.M. (2020). The Emerging Role of Gut Dysbiosis in Cardio-metabolic Risk Factors for Heart Failure. Curr Hypertens Rep *22*, 38.

Sato, B., Yoshikawa, D., Ishii, H., Kikuchi, R., Arima, T., Takeshita, K., Inoue, Y., Suzuki, S., Tanaka, M., and Kumagai, S. (2012). Indoxyl sulfate, a uremic toxin, and carotid intima-media thickness in patients with coronary artery disease. International journal of cardiology *163*, 214-216.

Sekirov, I., Russell, S.L., Antunes, L.C., and Finlay, B.B. (2010). Gut microbiota in health and disease. Physiol Rev *90*, 859-904.

Shimizu, H., Hirose, Y., Nishijima, F., Tsubakihara, Y., and Miyazaki, H. (2009). ROS and PDFG-β receptors are critically involved in indoxyl sulfate actions that promote vascular smooth muscle cell proliferation and migration. American Journal of Physiology-Cell Physiology *297*, C389-C396.

Shkoporov, A.N., Ryan, F.J., Draper, L.A., Forde, A., Stockdale, S.R., Daly, K.M., McDonnell, S.A., Nolan, J.A., Sutton, T.D.S., Dalmasso, M., *et al.* (2018). Reproducible protocols for metagenomic analysis of human faecal phageomes. Microbiome *6*, 68.

Sinha, T., Vich Vila, A., Garmaeva, S., Jankipersadsing, S.A., Imhann, F., Collij, V., Bonder, M.J., Jiang, X., Gurry, T., Alm, E.J., *et al.* (2019). Analysis of 1135 gut metagenomes identifies sex-specific resistome profiles. Gut Microbes *10*, 358-366.

Smits, L.P., Kootte, R.S., Levin, E., Prodan, A., Fuentes, S., Zoetendal, E.G., Wang, Z., Levison, B.S., Cleophas, M.C., and Kemper, E.M. (2018). Effect of vegan fecal microbiota transplantation on carnitine-and choline-derived trimethylamine-N-oxide production and

vascular inflammation in patients with metabolic syndrome. Journal of the American Heart Association *7*, e008342.
Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermúdez-Humarán, L.G., Gratadoux, J.J., Blugeon, S., Bridonneau, C., Furet, J.P., Corthier, G., *et al.* (2008). Faecalibacterium prausnitzii is an anti-inflammatory ... erium identified by gut microbiota analysis of Crohn disease patients. Proc Natl Acad Sci U S A *105*, 16731-16736.
Streppel, M.T., Arends, L.R., van 't Veer, P., Grobbee, D.E., and Geleijnse, J.M. (2005). Effect of dietary fiber intake on blood pressure - a meta-analysis of randomized, controlled clinical trials. Arch Intern Med *165*, 150-156.
Sun, S., Lulla, A., Sioda, M., Winglee, K., Wu, M.C., Jacobs, D.R., Jr., Shikany, J.M., Lloyd-Jones, D.M., Launer, L.J., Fodor, A.A., *et al.* (2019). Gut Microbiota Composition and Blood Pressure. Hypertension *73*, 998-1006.
Takagi, T., Naito, Y., Inoue, R., Kashiwagi, S., Uchiyama, K., Mizushima, K., Tsuchiya, S., Dohi, O., Yoshida, N., Kamada, K., *et al.* (2019). Differences in gut microbiota associated with age, sex, and stool consistency in healthy Japanese subjects. J Gastroenterol *54*, 53-63.
Tan, J., McKenzie, C., Potamitis, M., Thorburn, A.N., Mackay, C.R., and Macia, L. (2014). The role of short-chain fatty acids in health and disease. Adv Immunol *121*, 91-119.
Tang, W.H., Wang, Z., Levison, B.S., Koeth, R.A., Britt, E.B., Fu, X., Wu, Y., and Hazen, S.L. (2013). Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. N Engl J Med *368*, 1575-1584.
Tang, W.H., Wang, Z., Li, X.S., Fan, Y., Li, D.S., Wu, Y., and Hazen, S.L. (2017). Increased Trimethylamine N-Oxide Portends High Mortality Risk Independent of Glycemic Control in Patients with Type 2 Diabetes Mellitus. Clin Chem *63*, 297-306.
Taur, Y., and Pamer, E.G. (2013). The intestinal microbiota and susceptibility to infection in immunocompromised patients. Curr Opin Infect Dis *26*, 332-337.
Thomas, T., Gilbert, J., and Meyer, F. (2012). Metagenomics-a guide from sampling to data analysis. Microbial informatics and experimentation *2*, 3.
Tierney, B.T., Yang, Z., Luber, J.M., Beaudin, M., Wibowo, M.C., Baek, C., Mehlenbacher, E., Patel, C.J., and Kostic, A.D. (2019). The Landscape of Genetic Content in the Gut and Oral Human Microbiome. Cell Host Microbe *26*, 283-295 e288.
Toral, M., Robles-Vera, I., de la Visitacion, N., Romero, M., Yang, T., Sanchez, M., Gomez-Guzman, M., Jimenez, R., Raizada, M.K., and Duarte, J. (2019). Critical Role of the Interaction Gut Microbiota - Sympathetic Nervous System in the Regulation of Blood Pressure. Front Physiol *10*, 231.
Torres, N., Guevara-Cruz, M., Velazquez-Villegas, L.A., and Tovar, A.R. (2015). Nutrition and Atherosclerosis. Arch Med Res *46*, 408-426.
Tuteja, S., and Ferguson, J.F. (2019). Gut Microbiome and Response to Cardiovascular Drugs. Circ Genom Precis Med *12*, 421-429.
Van Daele, E., Knol, J., and Belzer, C. (2019). Microbial transmission from mother to child: improving infant intestinal microbiota development by identifying the obstacles. Crit Rev Microbiol *45*, 613-648.
Velmurugan, G., Ramprasath, T., Gilles, M., Swaminathan, K., and Ramasamy, S. (2017). Gut Microbiota, Endocrine-Disrupting Chemicals, and the Diabetes Epidemic. Trends Endocrinol Metab *28*, 612-625.

Verhaar, B.J.H., Collard, D., Prodan, A., Levels, J.H.M., Zwinderman, A.H., Backhed, F., Vogt, L., Peters, M.J.L., Muller, M., Nieuwdorp, M., *et al.* (2020). Associations between gut microbiota, faecal short-chain fatty acids, and blood pressure across ethnic groups: the HELIUS study. Eur Heart J.

Vieira-Silva, S., Falony, G., Belda, E., Nielsen, T., Aron-Wisnewsky, J., Chakaroun, R., Forslund, S.K., Assmann, K., Valles-Colomer, M., Nguyen, T.T.D., *et al.* (2020). Statin therapy is associated with lower prevalence of gut microbiota dysbiosis. Nature *581*, 310-315.

Vital, M., Howe, A.C., and Tiedje, J.M. (2014). Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. mBio *5*, e00889.

Wang, C., Huang, Z., Yu, K., Ding, R., Ye, K., Dai, C., Xu, X., Zhou, G., and Li, C. (2017). High-Salt Diet Has a Certain Impact on Protein Digestion and Gut Microbiota: A Sequencing and Proteome Combined Study. Front Microbiol *8*, 1838.

Wang, Z., Roberts, A.B., Buffa, J.A., Levison, B.S., Zhu, W., Org, E., Gu, X., Huang, Y., Zamanian-Daryoush, M., Culley, M.K., *et al.* (2015). Non-lethal Inhibition of Gut Microbial Trimethylamine Production for the Treatment of Atherosclerosis. Cell *163*, 1585-1595.

Wang, Z., Tang, W.H., Buffa, J.A., Fu, X., Britt, E.B., Koeth, R.A., Levison, B.S., Fan, Y., Wu, Y., and Hazen, S.L. (2014). Prognostic value of choline and betaine depends on intestinal microbiota-generated metabolite trimethylamine-N-oxide. Eur Heart J *35*, 904-910.

Watanabe, I., Tatebe, J., Namba, S., Koizumi, M., Yamazaki, J., and Morita, T. (2013). Activation of aryl hydrocarbon receptor mediates indoxyl sulfate-induced monocyte chemoattractant protein-1 expression in human umbilical vein endothelial cells. Circulation Journal *77*, 224-230.

Wiedermann, C.J., Kiechl, S., Dunzendorfer, S., Schratzberger, P., Egger, G., Oberhollenzer, F., and Willeit, J. (1999). Association of endotoxemia with carotid atherosclerosis and cardiovascular disease. Journal of the American College of Cardiology *34*, 1975-1981.

Wilck, N., Matus, M.G., Kearney, S.M., Olesen, S.W., Forslund, K., Bartolomaeus, H., Haase, S., Mahler, A., Balogh, A., Marko, L., *et al.* (2017). Salt-responsive gut commensal modulates TH17 axis and disease. Nature *551*, 585-589.

Witkowski, M., Weeks, T.L., and Hazen, S.L. (2020). Gut Microbiota and Cardiovascular Disease. Circ Res *127*, 553-570.

Woodworth, M.H., Carpentieri, C., Sitchenko, K.L., and Kraft, C.S. (2017). Challenges in fecal donor selection and screening for fecal microbiota transplantation: A review. Gut Microbes *8*, 225-237.

Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.Y., Keilbaugh, S.A., Bewtra, M., Knights, D., Walters, W.A., Knight, R., *et al.* (2011). Linking long-term dietary patterns with gut microbial enterotypes. Science *334*, 105-108.

Wu, M., Yang, S., Wang, S., Cao, Y., Zhao, R., Li, X., Xing, Y., and Liu, L. (2020). Effect of Berberine on Atherosclerosis and Gut Microbiota Modulation and Their Correlation in High-Fat Diet-Fed ApoE-/- Mice. Front Pharmacol *11*, 223.

Yan, Q., Gu, Y., Li, X., Yang, W., Jia, L., Chen, C., Han, X., Huang, Y., Zhao, L., Li, P., *et al.* (2017). Alterations of the Gut Microbiome in Hypertension. Front Cell Infect Microbiol *7*, 381.

Yan, X., Jin, J., Su, X., Yin, X., Gao, J., Wang, X., Zhang, S., Bu, P., Wang, M., Zhang, Y., *et al.* (2020). Intestinal Flora Modulates Blood Pressure by Regulating the Synthesis of

Intestinal-Derived Corticosterone in High Salt-Induced Hypertension. Circ Res *126*, 839-853.

Yang, Q., Lin, S.L., Kwok, M.K., Leung, G.M., and Schooling, C.M. (2018). The Roles of 27 Genera of Human Gut Microbiota in Ischemic Heart Disease, Type 2 Diabetes Mellitus, and Their Risk Factors: A Mendelian Randomization Study. Am J Epidemiol *187*, 1916-1922.

Yang, T., Santisteban, M.M., Rodriguez, V., Li, E., Ahmari, N., Carvajal, J.M., Zadeh, M., Gong, M., Qi, Y., Zubcevic, J., *et al.* (2015). Gut dysbiosis is linked to hypertension. Hypertension *65*, 1331-1340.

Yi, B., Titze, J., Rykova, M., Feuerecker, M., Vassilieva, G., Nichiporuk, I., Schelling, G., Morukov, B., and Chouker, A. (2015). Effects of dietary salt levels on monocytic cells and immune responses in healthy human subjects: a longitudinal study. Transl Res *166*, 103-110.

Yoshida, N., Emoto, T., Yamashita, T., Watanabe, H., Hayashi, T., Tabata, T., Hoshi, N., Hatano, N., Ozawa, G., Sasaki, N., *et al.* (2018). Bacteroides vulgatus and Bacteroides dorei Reduce Gut Microbial Lipopolysaccharide Production and Inhibit Atherosclerosis. Circulation *138*, 2486-2498.

Yu, E., Ruiz-Canela, M., Guasch-Ferré, M., Zheng, Y., Toledo, E., Clish, C.B., Salas-Salvadó, J., Liang, L., Wang, D.D., and Corella, D. (2017). Increases in plasma tryptophan are inversely associated with incident cardiovascular disease in the Prevencion con Dieta Mediterranea (PREDIMED) Study. The Journal of nutrition *147*, 314-322.

Zelante, T., Iannitti, R.G., Cunha, C., De Luca, A., Giovannini, G., Pieraccini, G., Zecchi, R., D'Angelo, C., Massi-Benedetti, C., and Fallarino, F. (2013). Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. Immunity *39*, 372-385.

Zhang, C., Zhang, M., Wang, S., Han, R., Cao, Y., Hua, W., Mao, Y., Zhang, X., Pang, X., Wei, C., *et al.* (2010a). Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice. ISME J *4*, 232-241.

Zhang, T., Kurita-Ochiai, T., Hashizume, T., Du, Y., Oguchi, S., and Yamamoto, M. (2010b). Aggregatibacter actinomycetemcomitans accelerates atherosclerosis with an increase in atherogenic factors in spontaneously hyperlipidemic mice. FEMS Immunology & Medical Microbiology *59*, 143-151.

Zhu, W., Gregory, J.C., Org, E., Buffa, J.A., Gupta, N., Wang, Z., Li, L., Fu, X., Wu, Y., Mehrabian, M., *et al.* (2016). Gut Microbial Metabolite TMAO Enhances Platelet Hyperreactivity and Thrombosis Risk. Cell *165*, 111-124.

CHAPTER 3.

Hepatic transcriptional profile reveals the role of diet and genetic backgrounds on metabolic traits in female progenitor strains of the Collaborative Cross

3.1. Author Contributions

This manuscript was published in the Physiological Genomics journal on May 13th, 2021 and selected as a cover page. B.J.B designed all the experiments. M.K., A.O.C., and J.A. performed the experiments. M.K., M.N.H., A.O.C., J.A., and B.D.J. analyzed raw data. M.K. and B.J.B. wrote the manuscript, which was critically reviewed by all authors.

3.2. Abstract

Mice have provided critical mechanistic understandings of clinical traits underlying Metabolic Syndrome (MetSyn) and susceptibility to MetSyn in mice is known to vary among inbred strains. We investigated the diet- and strain-dependent effects on metabolic traits in the eight Collaborative Cross (CC) founder strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HILtJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ). Liver transcriptomics analysis showed that both atherogenic diet and host genetics have profound effects on the liver transcriptome, which may be related to differences in metabolic traits observed between strains. We found strain differences in circulating trimethylamine N-Oxide (TMAO) concentration and liver triglyceride content, both of which are traits associated with metabolic diseases. Using a network approach, we identified a module of transcripts associated with TMAO and liver triglyceride content which was enriched in functional pathways. Interrogation of the module related to metabolic traits identified NADPH oxidase 4 (*Nox4),* a gene for a key enzyme in the production of reactive oxygen species, which showed a strong association with plasma TMAO and liver triglyceride. Interestingly, *Nox4* was identified as the highest expressed in the C57BL/6J and NZO/HILtJ strains and the lowest expressed in the CAST/EiJ strain. Based on these results, we suggest that there may be genetic variation in the contribution of *Nox4* to the regulation of plasma TMAO and liver triglyceride content. In summary, we show that liver transcriptomic analysis identified diet- or strain-specific pathways for metabolic traits in the Collaborative Cross (CC) founder strains.

3.3. Introduction

Metabolic Syndrome (MetSyn) is a cluster of clinical traits (including elevated blood lipids and glucose concentrations, increased blood pressure, and central obesity (Huang, 2009)) that is highly associated with risk of diabetes and cardiovascular disease. In particular, the liver plays a central role in regulating these clinical traits and thus metabolic imbalance in the liver can affect susceptibility to MetSyn. For example, alterations in hepatic metabolism can induce dyslipidemia (Hamaguchi et al., 2005; Marchesini et al., 2003) and production of novel metabolites associated with increased risk of MetSyn such as trimethylamine N-oxide (TMAO) (Barrea et al., 2018; Chen et al., 2016; Koeth et al., 2013; Lent-Schochet et al., 2018; Schugar et al., 2017; Stubbs et al., 2016). Specifically, TMAO is converted from trimethylamine (TMA) produced in the intestine by the activity of liver flavin monooxygenase 3 (FMO3) (Wang et al., 2011). Plasma levels of TMAO are primarily determined by genetic variation (Hartiala et al., 2014; Lambert et al., 2001) and diet (Cho et al., 2017), and FMO3 may promote dyslipidemia by regulating several genes involved in hepatic gluconeogenesis and lipogenesis (Shih et al., 2015; Warrier et al., 2015). There are a number of factors that affect susceptibility to MetSyn including genetics and environmental queues such as diet. In particular, genetics has been shown to be an important factor significantly affecting the susceptibility of MetSyn in humans (Altshuler et al., 2000; Grant et al., 2006; Hani et al., 1998). Identifying the genetic architecture and biological pathways that modulate the risk of MetSyn is essential to discovering more effective therapeutics approaches.

The mouse has been indispensable for the study of MetSyn (Attie et al., 2017; Getz and Reardon, 2006) as both genetic and environmental factors can be well controlled in mouse models. The phenotype spectrum present in various mouse strains provides an opportunity to

discover genetic functions related to metabolic traits. However, most studies have been conducted with a small number of mouse strains with limited genetic variation. Almost exclusively mouse gene knockout studies are conducted in C57BL/6J mice, with a smaller number performed in FVB and 129/Sv. Studies of mice with targeted overexpression or inactivation of a gene often report specific phenotypic changes, but these effects are highly influenced by the background strain of the mice harboring the genetic mutation(s) (Meng et al., 2007). Thus, understanding the underlying genetic architecture remains important to further our understanding of MetSyn.

An alternative approach is to use forward genetic studies utilizing a wide variety of mice to investigate how natural variants affect MetSyn. Classically these have been done in F2 crosses but more recently multiparent advanced generation intercross populations have been developed, such as the Collaborative Cross (CC). The CC mouse population is derived from five classic inbred mouse strains A/J, C57BL/6J (B6), 129S1/SvImJ (129), NOD/ShiLtJ (NOD), NZO/HILtJ (NZO), and three wild-derived strains CAST/EiJ (CAST), PWK/PhJ (PWK), and WSB/EiJ (WSB) (Churchill et al., 2012). These eight CC founder strains are highly diverse genetically as they contain ~40 million SNPs and numerous insertions and deletions. The genetic diversity across the CC founder strains provides an opportunity to assess the effect of host genetics on metabolic traits. The genetic and phenotypic diversity of the combined eight founder strains is similar to the inter-individual diversity of the human population. The utilization of the genomic sequence of the eight CC founder strains provides an unprecedented unique resource for genetic mapping and correlation studies (Chick et al., 2016; Coffey et al., 2019; Coffey et al., 2017; Huda et al., 2020; Keller et al., 2018; Keller et al., 2019; Kemis et al., 2019; O'Connor et al., 2014; Que et al., 2020; Smallwood et al., 2014; Tyler et al., 2017; Winter et al., 2017)

Here, we evaluate the variability of metabolic traits and perform global liver transcriptomics from the eight CC founder strains fed a standard purified diet or an atherogenic diet. Our studies focus on female mice as a number of studies have reported that female mice are generally more susceptible to atherosclerosis than male mice (AlSiraj et al., 2019; Bennett et al., 2015; Daugherty, 2002; Hsu and Smith, 2013; O'Connor et al., 2014; Smallwood et al., 2014). Our study shows that both diet and genetic background have a profound influence on metabolic traits and the associated transcriptional network. For example, analyses of differentially expressed genes (DEGs) and gene networks suggest that diet- or strain-specific DEGs and gene clusters are enriched for biological pathways known to affect metabolic traits. We identify a novel co-expression module associated with plasma TMAO and describe a candidate gene, NADPH oxidase 4 (*Nox4*), which is a hydrogen peroxide NADPH oxidase isoform. We found that the *Nox4* gene showed a strong association with plasma TMAO and liver triacylglycerol (TG) in the liver transcriptome. Our results demonstrate the utility of leveraging the CC to understand dietary influences on the liver transcriptome and disease-associated traits.

3.4. Methods

3.4.1. Ethics Statement

We followed all NIH animal welfare guidelines and animal care. The study protocols were approved by the North Carolina Research Campus (NCRC) Animal Care and Use Committee.

3.4.2. Study Design

Eight female mice from each of the eight different CC founder strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HILtJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ) were purchased from Jackson Laboratories (Bar Harbor, ME, USA) at 4 weeks of age. The study design was reported previously (O'Connor et al., 2014). Briefly, mice were housed under standard

conditions (12 h light: dark, temperature- and humidity-controlled conditions) with free access to water and a nutritionally purified AIN-93M diet (#D10012M; Research Diets Inc, New Brunswick, NJ, USA). Four weeks after AIN-93M administration, mice were assigned to either the AIN-93M diet or high-fat cholic acid (HFCA) diet (#D12109C; Research Diets Inc) for an additional 16 weeks (n = 4 per diet per strain) (**Table 3.1**). After feeding this diet for 16 weeks, mice were euthanized for tissue collection. Euthanasia of all mice was performed by cervical dislocation after anesthesia with isoflurane. For eight CC founder strains of mice, liver, gonadal fat, spleen, and heart were collected and weighed upon euthanasia.

3.4.3. Body Composition

Body composition (fat mass and lean mass) was assessed using EchoMRI™-100H (Echo MRI LLC, Houston, TX, USA) at 8 weeks and 24 weeks, respectively. Body fat and lean mass percentages were calculated by dividing fat mass by body weight and lean mass by body weight, respectively.

Table 3.1. Nutrient constituents in AIN-93M, high-fat cholic acid (HFCA), and high-fat and high-cholesterol (HFHC) diet.

Class description	Ingredient	AIN-93M (Grams)	HFCA (Grams)
Protein	Casein, Lactic, 30 Mesh	200.00 g	200.00 g
Protein	Cystine, L	3.00 g	3.00 g
Carbohydrate	Starch, Corn	150.00 g	212.00 g
Carbohydrate	Sucrose, Fine Granulated	500.00 g	113.00 g
Carbohydrate	Lodex 10 (Maltodextrin)	0.00 g	71.00 g
Fiber	Solka Floc (Cellulose)	50.00 g	50.00 g
Fat	Cocoa Butter, Deodorized	0.00 g	155.00 g
Fat	Soybean Oil	50.00 g	25.00 g
Mineral	Potassium Citrate, Monohydrate		16.50 g
Mineral	Calcium Phosphate, Dibasic		13.00 g
Mineral	Calcium Carbonate, Light	37.00 g	5.50 g
Mineral	Mineral mixture		10.00 g
Mineral	Sodium Chloride		
Vitamin	Choline Bitartrate	2.00 g	2.00 g
Vitamin	Vitamin mixture	10.00 g	10.00 g
Special	Cholesterol	0.0 g	11.25 g
Special	Sodium Cholate	0.0 g	4.5 g
	Total:	**1002.0 g**	**901.75 g**

The two diets used in this study were manufactured by Research Diets. AIN-93M was fed to the Diversity Outbred (DO) founder strains from 6 weeks of age in order to ensure that there were no spurious effects due to the potential variable composition of standard laboratory chow. CC founder strains were then fed either AIN-93M or high-fat cholic acid (HFCA) diet for 16 weeks from 8-24 weeks of age. HFCA diets are considered atherogenic and were intended to induce the formation of atherosclerosis in the CC founder strains mice.

3.4.4. Plasma Clinical Metabolic Markers

Mice at 8 weeks or 24 weeks of age fasted for 4 h before blood collection via retro-orbital bleeding. Blood was collected into EDTA-containing tubes and plasma was separated by centrifugation at 10,000×g for 10 min. Plasma alanine aminotransferase (ALT), aspartate aminotransferase (AST), TG, total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), glucose, and urea were measured by the Biolis 24i Analyzer (Carolina Liquid Chemistries, Winston-Salem, NC). Very low-density lipoprotein cholesterol/low-density lipoprotein cholesterol (VLDL-C/LDL-C) levels were determined by subtracting high-density lipoprotein cholesterol (HDL-C) from total cholesterol. Insulin was assessed using the Alpco Mouse Ultrasensitive Insulin ELISA assay (Alpco, Salem, NH) and measured at 450 nm using a microplate reader (Bio-Tek, Winooski, VT, USA).

3.4.5. Plasma Metabolite Analysis using LC/MS/MS

TMAO analytes were measured by the Metabolomics Core Facility in the NCRC (Coffey et al., 2019; O'Connor et al., 2014). Briefly, plasma was extracted with internal standards TMAO-d9 (Cambridge Isotope Laboratories, Tewksbury, MA), creatinine-d3 (CDN Isotopes Inc., Quebec, Canada), choline-d9 (Cambridge Isotope Laboratories) and betaine-d9 (Sigma-Aldrich, St. Louis, MO), incubated on ice for 10 min, and centrifuged at 15,000 g for 2 minutes. The concentrations of TMAO, creatinine, choline, and betaine were quantified by using liquid chromatography-stable isotope dilution-multiple reaction monitoring mass spectrometry (LC-SID-MRM/MS). Chromatographic separations were conducted on an Atlantis Silica HILIC 3µm 4.6×150mm column (Waters Corp, Milford, MA) using a Waters ACQUITY UPLC system. The metabolites and their corresponding isotopes were monitored on a Waters TQ detector using characteristic precursor-product ion transitions: 76→58 for TMAO, 85→66 for TMAO-d9,

114→86 for creatinine, 117→89 for creatinine-d9, 104→45 for choline, 113→45 for choline-d9, 118→59 for betaine, and 127→68 for betaine-d9. Concentrations of each metabolite in samples were determined by calculating the peak area ratio of the metabolite versus its isotope.

3.4.6. Hepatic Triacylglycerol

Hepatic TG levels were quantified via Folch extraction. Mouse liver was collected, frozen, and stored at −80°C prior to analysis. Frozen mouse liver tissue was thoroughly homogenized for 5 min in 500 μL of a 2:1 v/v chloroform/methanol mix and then equilibrated for 15 min at room temperature. After adding 100 μL of 0.9% w/v NaCl to each sample, the samples were vortexed for 1 min and centrifuged at 2000 × g for 15 min at 4°C. The lower organic phase was separated and evaporated in Eppendorf tubes under a stream of nitrogen for 1 h. After evaporation, each tube was resuspended with 500 μL of a 0.5% Triton X-100/PBS solution, sonicated for 5 minutes using Bioruptor, and placed in a drying bath at 55°C for 5 min. Hepatic TG was measured using a colorimetric assay, (Infinity™, Thermo Scientific, Waltham, MA, USA) according to the manufacturer's instructions as follows: 2 μL of the standards, samples, and blanks were pipetted into a 96-well plate in duplicate and 200 μL of the Infinity reagent was added to the 96-well plate. Absorbance (500/660 nm) was measured on a 96-well plate reader.

3.4.7. Metabolic Rate and Activity

Mice were placed into individual indirect calorimetry cages (Phenomaster, TSE SYSTEMS, Chesterfield, MO) at the week after 16 weeks of the experimental diet challenge to obtain O_2 consumption (VO_2), respiratory exchange ratio (RER), and feed consumption measurements. After a 24-hr acclimation period, data were collected during the following 24-hour period. Basal activity was measured in three dimensions (x, y, and z) as breaks in the two infrared light beam frames that surrounded each cage. Rearing was detected by beam breaks in the z-axis and total

physical activity was defined as the sum of beam breaks in all three axes in counts. Feed was available *ad libitum* and consumption was measured by weighing sensors that held containers for feed and water respectively and recorded the amount of feed or water consumed.

3.4.8. RNA-Seq Library Preparation and Sequencing

Total RNA was extracted from frozen liver samples with the Maxwell 16 LEV simplyRNA Tissue Kit (Promega, Madison, WI, USA) according to the manufacturer's protocol. The quality and amount of liver RNA were evaluated using the Qubit RNA HS assay kit (Thermo Fisher Scientific, Waltham, MA, USA) and the Biorad Bioanalyzer Chip (Hercules, CA, USA). RNA samples from 24 mice fed the AIN-93M diet and 24 mice fed the HFCA diet were submitted to the David H. Murdock Research Institute (Kannapolis, NC, USA). The RNA-seq libraries were constructed from total RNA following the Illumina TruSeq RNA library construction protocol. The size of the adapted fragments in the libraries was determined by running an Agilent DNA 1000 Chip. In parallel, the DNA concentration in the libraries was quantified using Real-Time PCR (Kapa Biosystems, Wilmington, MA, USA). The pooled libraries were sequenced on the Illumina HiSeq 2500 sequencing to achieve 100 bp paired-end reads in a total of 11 lanes (Illumina Inc., San Diego, CA, USA). Seven pools were created from 44 samples, with 6-7 samples per pool in equimolar concentrations. Two B6 and two CAST samples from the control diet were deep sequenced with one sample per lane because B6 strain is the most commonly used laboratory mice and CAST strain is a wild-derived mouse strain that is genetically distinct from the B6 strain. Raw data were deposited at National Center for Biotechnology Information's Gene Expression Omnibus (GEO accession GSE159992).

3.4.9. RNA-Seq Mapping and Quantification

Raw read data was filtered using HTStream (version 1.1.0, https://github.com/ibest/HTStream), which included screening for contaminants (such as PhiX and rRNA), PCR deduplication readout, quality-based trimming, adapter trimming, and overlapping paired-end reads. We randomly took 50% of the reads from the two B6 and two CAST deep sequenced samples. STAR (version 2.7.0f) (Dobin et al., 2013) was used to align the processed data to custom reference mouse genomes constructed by incorporating genetic variants of eight founder strains into reference mouse genome GRCm38 using g2gtools (https://github.com/churchill-lab/g2gtools). Custom R code was then used for sequence read and alignment quality assessment as well as collating counts into a single table for downstream analysis. We obtained median 32 million pass-filter reads, 20 million uniquely mapped reads, and 17 million reads mapped to genes per library (**Table 3.2**). We filtered in 12,502 transcripts with median counts per million (CPM) greater than 1 in 48 liver samples.

Table 3.2. RNA-seq Alignment Statistics.

Sample	Strain	Diet	Input Reads	Uniquely Mapped Reads	Uniquely Mapped Reads (%)	Reads Mapped to Genes	Reads Mapped to Genes (%)
1	A/J	HFCA	32,471,605	22,292,704	94.11%	18,729,116	57.68%
2	A/J	HFCA	34,804,995	24,291,853	94.21%	20,691,464	59.45%
3	A/J	HFCA	30,038,112	20,498,332	94.41%	16,936,644	56.38%
4	B6	HFCA	27,871,941	18,980,300	94.67%	16,098,026	57.76%
5	B6	HFCA	27,070,672	19,058,483	94.91%	15,950,314	58.92%
6	B6	HFCA	28,447,003	19,903,514	94.81%	16,896,889	59.40%
7	129	HFCA	30,804,898	19,668,288	93.81%	16,657,068	54.07%
8	129	HFCA	30,556,344	19,608,251	93.86%	16,343,494	53.49%
9	129	HFCA	33,207,972	21,769,521	93.86%	18,215,319	54.85%
10	NOD	HFCA	34,001,471	22,451,767	94.03%	18,917,822	55.64%
11	NOD	HFCA	32,488,275	22,166,097	94.29%	18,753,671	57.72%
12	NOD	HFCA	33,237,490	21,784,033	94.06%	18,335,202	55.16%
13	NZO	HFCA	44,287,463	28,332,120	93.84%	23,932,477	54.04%
14	NZO	HFCA	33,097,584	21,907,527	93.71%	18,676,327	56.43%
15	NZO	HFCA	34,765,193	22,263,323	93.75%	18,947,908	54.50%
16	CAST	HFCA	28,786,382	19,377,149	92.37%	16,409,642	57.00%
17	CAST	HFCA	25,631,990	17,024,883	92.16%	14,389,298	56.14%
18	CAST	HFCA	25,222,984	16,032,164	91.41%	13,480,814	53.45%
19	PWK	HFCA	37,541,176	25,145,942	93.23%	21,291,964	56.72%
20	PWK	HFCA	39,559,915	25,289,336	93.24%	21,432,494	54.18%
21	PWK	HFCA	29,808,966	20,661,198	92.86%	17,379,898	58.30%
22	WSB	HFCA	33,261,931	20,600,820	93.31%	17,343,786	52.14%
23	WSB	HFCA	32,515,770	21,424,801	93.39%	17,972,497	55.27%
24	WSB	HFCA	33,107,521	21,923,117	93.70%	18,672,909	56.40%
25	A/J	AIN93	30,227,183	18,666,507	93.64%	14,853,750	49.14%
26	A/J	AIN93	25,308,842	16,036,190	93.90%	13,040,075	51.52%
27	A/J	AIN93	32,469,136	20,185,467	93.65%	16,678,866	51.37%
28	B6	AIN93	128,126,695	63,742,372	95.12%	52,360,328	40.87%
29	B6	AIN93	138,117,965	66,367,245	94.99%	56,283,808	40.75%
30	B6	AIN93	28,799,118	18,315,994	94.18%	15,589,871	54.13%
31	129	AIN93	57,832,217	32,468,045	93.85%	27,276,730	47.17%
32	129	AIN93	30,050,195	18,326,391	93.42%	15,358,000	51.11%
33	129	AIN93	24,992,183	15,852,934	93.04%	12,819,341	51.29%
34	NOD	AIN93	30,360,789	19,281,162	93.87%	16,244,911	53.51%
35	NOD	AIN93	33,175,443	20,082,373	93.78%	16,916,497	50.99%
36	NOD	AIN93	31,313,942	19,677,173	93.60%	16,504,550	52.71%
37	NZO	AIN93	30,850,017	18,954,308	93.22%	16,143,752	52.33%

38	NZO	AIN93	31,391,882	19,436,609	92.57%	16,311,377	51.96%
39	NZO	AIN93	37,786,619	22,457,729	93.05%	18,929,694	50.10%
40	CAST	AIN93	130,363,721	60,730,820	92.35%	49,065,718	37.64%
41	CAST	AIN93	114,004,639	54,944,651	91.42%	39,856,181	34.96%
42	CAST	AIN93	26,720,731	16,276,939	91.04%	13,553,433	50.72%
43	PWK	AIN93	31,163,059	18,763,199	92.42%	15,454,495	49.59%
44	PWK	AIN93	35,128,092	21,216,708	92.13%	17,885,477	50.92%
45	PWK	AIN93	41,215,170	23,183,967	92.27%	19,252,736	46.71%
46	WSB	AIN93	31,788,919	18,844,353	92.77%	16,030,981	50.43%
47	WSB	AIN93	31,136,857	18,175,130	93.17%	15,470,437	49.69%
48	WSB	AIN93	35,471,949	20,252,433	92.51%	16,642,834	46.92%
	Median		32,470,371	20,375,383	93.65%	16,926,571	53.47%

3.4.10. Differential Gene Expression and Enrichment Analysis

Diet or Strain-specific differential expression genes (DEGs) analysis was performed using the R package ‘limma’ version 3.11 (Ritchie et al., 2015) from TMM (trimmed mean of M values) normalized log2 transformed count per million (CPM) values.

Enrichment analyses for DEGs or modules were performed using enrichR (Chen et al., 2013) to generate enrichment terms and pathways from the Gene Ontology (GO) Biological Process 2018, Kyoto Encyclopedia of Genes and Genomes (KEGG) 2019 Mouse, and Jenson Diseases (Pletscher-Frankild et al., 2015). This analysis identifies differential enrichment terms and pathways for the functional categories of DEGs or transcripts in the module. The GO Biological process 2018 contains 5,103 terms and 14,433 genes. While it is clear that individual GO terms can be found in related classes of ontology, GO terms do not occupy strictly fixed levels in a hierarchy. Each of the GO terms identified is associated with a unique GO annotation number that relates to a specific function. Both the Gene Ontology website (http://geneontology.org/docs/faq/) and specific tool (enrichR) do not utilize a specific hierarchy thus all available terms are used in the analysis.

3.4.11. Weighted Gene Co-expression Network Analyses (WGCNA)

Co-expression gene modules were calculated using Weighted Gene Co-expression Network Analysis (WGCNA) version 1.13 (Zhang and Horvath, 2005), which performs network construction by module detection. For the WGCNA analysis, log2 transformed 12,502 CPM measured in 48 liver samples from CC founder strains were included. We used a soft thresholding power of 9 by the scale-free topology criterion in the WGCNA package. We chose the "unsigned" network type to maintain the relationship of the negatively correlated gene and the "signed" topological overlap matrix (TOM) to exclude the connections influenced by noise

(Langfelder and Horvath, 2008; Zhang and Horvath, 2005). We set 20 as the minimum number of genes to form a module. The network connectivity of each gene was calculated as the sum of the intensity of connectivity with all genes in the other network.

For each module, the first principal component (PC1) and the module eigengene (ME) were calculated and used for the correlation with metabolic traits. The average number of transcripts per module was 595, ranging from 44 (royalblue module) to 4,020 (turquoise module).

3.4.12. Assessing Genetic Variation at FMO3 and NOX4 Loci

To better understand the SNPs of the *Fmo3* and *Nox4* genes discovered as candidate genes, we compared 658 SNPs (Chromosome 1: 162,954,207-162,984,416) in *Fmo3* and 2,641 SNPs (Chromosome 7: 87,246,136-87,398,699) in *Nox4* in eight CC founder mouse genomes available from the Sanger Institute's mouse database (www.sanger.ac.uk) and calculated SNP similarity between the reference genome, B6 strain and the other 7 strains. For example, in 2,641 *Nox4* SNPs, the number of SNPs difference between A/J and B6 strain is 663. Therefore, dissimilarity between A/J and B6 for *Nox4* SNPs is 663 divided by 2,641, which is 0.251 and the similarity (%) is 74.9% (100%-25.1%). The effects of SNP mutations on the protein function for the discovered candidate genes *Fmo3* and *Nox4* were determined by a web-based tool, Protein Variation Effect Analyzer-PROVEAN (Choi and Chan, 2015) and Sorting Intolerant From Tolerant-SIFT (Vaser et al., 2016).

3.4.13. Other statistical analysis

All statistical analyses were performed in R (v.3.5.3) (R Core Team) (R Core Team, 2016). Diet or strain effects were assessed using two-group Mann-Whitney U (Wilcoxon rank) or Kruskal-Wallis statistical test, respectively. Diet by strain interaction effect was assessed using a two-way

ANOVA test. Tukey's multiple comparison test was performed to compare groups with different diets and strains for plasma TMAO and liver TG traits. Spearman's correlation was used to correlate the clinical traits and liver transcripts. The p-values were adjusted using the Benjamini-Hochberg (BH) false discovery rate (FDR) procedure (Benjamini and Hochberg, 1995), and correlation coefficients and adjusted p-value were visualized using the 'pheatmap' package (Kolde et al., 2018). Significance was determined with a p-value < 0.05.

3.5. Results

3.5.1. Hepatic Transcriptomics Reveals Diet-Specific Differences in the CC Progenitors

A number of studies have reported that thousands of genes are differentially expressed by a high-fat diet as compared to a control diet (Cheng et al., 2018; Lan et al., 2015; Yoon et al., 2019; Zhou et al., 2020). We profiled global gene expression by RNA-Seq in three female mice fed the AIN-93M diet and three female mice fed a HFCA diet from each CC founder strains (n=48 in total). We found that 12,502 transcripts were expressed in the liver and performed in differential gene expression analysis to identify transcriptional responses to diet perturbation regardless of strain by combining all eight founder strains. Principal component analysis revealed distinct differences in global gene expression between diets (**Figure 3.1A**). A total of 6,411 genes showed significant differential expression (FDR adjusted P <0.05; **Figure 3.1B**) between mice fed the HFCA diet (3,157 genes were upregulated) compared to those fed the AIN-93M diet (3,254 genes were upregulated). These 6,411 genes we define as our "Core Diet DEGs". A volcano plot showed that genes such as *Gpnmb*, *Mmp12*, *Lpl*, and *Col1a1* previously reported as fatty liver-related genes (Cazanave et al., 2017; McGettigan et al., 2019; Remmerie et al., 2020) were also identified as HFCA diet-specific differentially expressed genes (DEGs) in this study (**Figure 3.1C**).

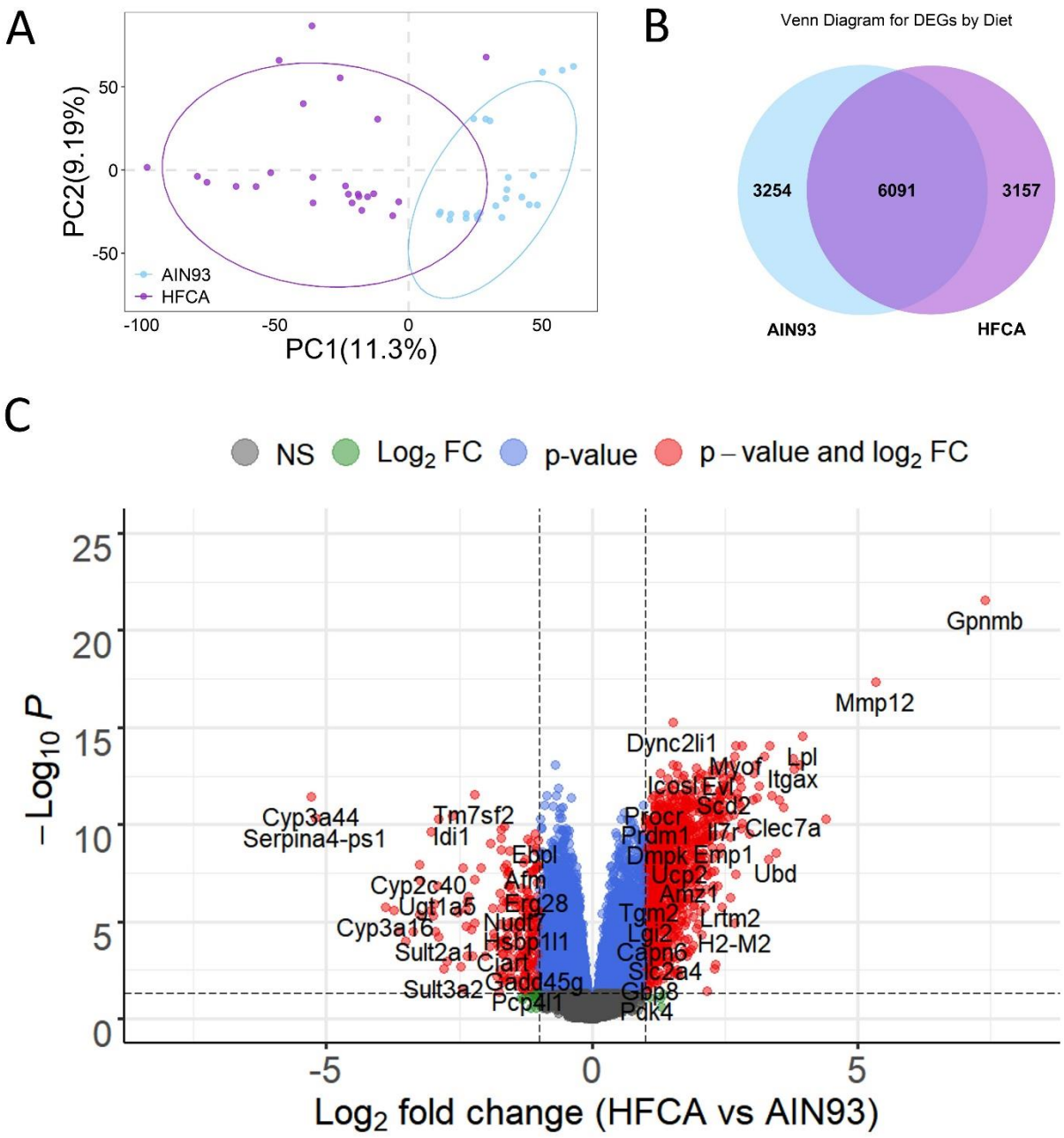


Figure 3.1. Effect of diet on liver gene expression in female eight Collaborative Cross (CC) founder strains mice.

We identified the effect of an atherogenic diet on global liver gene expression in eight CC founder strains (n = 48). PCA analysis (A), the number of differential expression genes (DEGs, adj.P < 0.05) (B), and Volcano plot (C) between high-fat and cholic acid (HFCA) diet and AIN-93M diet in liver gene expression in eight CC founder strains. (B) 3254 transcripts in lightblue circle were upregulated in AIN-93M diet fed mice and 3157 transcripts in purple circle were upregulated in HFCA diet fed mice. (C) Horizontal dotted lines indicate adj.P < 0.05, vertical dotted gray lines indicate a 2-fold difference.

We also determined which potential biological aspects of liver metabolism were reflected by the DEGs between the diets. Enrichment analysis revealed upregulated genes in a specific diet enriched in a number of GO Biological Processes and KEGG pathways (**Figure 3.2A**). In genes that were not identified as DEGs between the two diets (non-diet responsive genes: gray color denoted), a relatively small number of GO terms and KEGG pathways were identified compared to DEGs upregulated in a specific diet (**Figure 3.2A**). In particular, there was an upregulation of genes involved in immune response in the HFCA diet-fed mice while genes involved in mitochondrial function were upregulated in the AIN-93M diet-fed mice. GO Biological Processes and KEGG pathways that were highly enriched in HFCA diet-fed mice included ‘‘Extracellular matrix organization (GO: 00030196)” and “Neutrophil mediated immunity (GO: 0002446)’’ (-logP > 21) and “Osteoclast differentiation” (KEGG pathway) (-logP > 18). GO Biological Processes and KEGG pathways that were enriched in AIN-93M diet-fed mice included ‘‘Mitochondrial translation (GO: 0032543)” and “Respiratory electron transport chain (GO: 0022904)’’ (-logP > 34) and ‘‘Thermogenesis, Oxidative phosphorylation, and Non-alcoholic fatty liver disease’’ (KEGG pathway) (-logP > 19) (**Figures 3.2B and 3.2C, Table 3.3 and 3.4**).

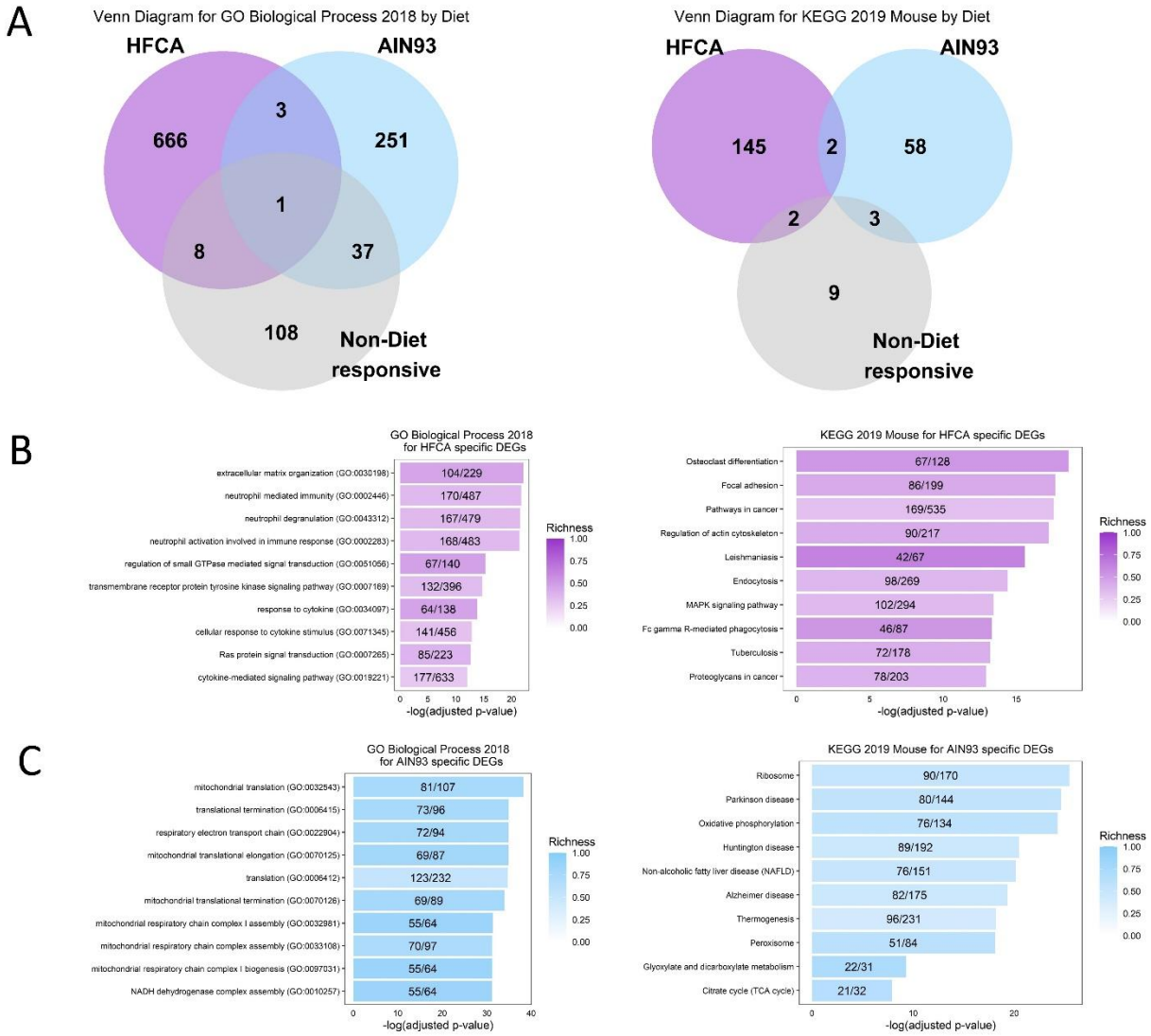


Figure 3.2. Effect of diet on functional enrichment for liver transcriptome in female eight Collaborative Cross (CC) founder strains mice.

(A) Venn diagram to identify overlapping gene ontology (GO) Biological Process 2018 terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) 2019 Mouse pathways between upregulated genes in HFCA diet, upregulated genes in AIN-93M diet, or non-diet responsive genes in enrichment analysis. Top 10 GO terms and KEGG pathways of upregulated genes in HFCA diet (B) and AIN-93M diet (C) identified in enrichment analysis. Pathways were ordered from top to bottom by significance (highest to lowest) and colored by gene richness.

Table 3.3. Top 30 Gene Ontology results for diet-specific DEGs in liver in eight CC founder strains. n = 48 (24 mice for AIN-93M diet and 24 mice for HFCA diet)

Diet	Term	Overlap (gene count)	Adjusted P-value	Odds Ratio	Combined Score[a]	Rank
HFCA	extracellular matrix organization (GO:0030198)	104/229	8.37E-23	2.88	171.04	1
HFCA	neutrophil mediated immunity (GO:0002446)	170/487	1.74E-22	2.21	128.31	2
HFCA	neutrophil degranulation (GO:0043312)	167/479	2.8E-22	2.21	125.57	3
HFCA	neutrophil activation involved in immune response (GO:0002283)	168/483	3.59E-22	2.21	125.36	4
HFCA	regulation of small GTPase mediated signal transduction (GO:0051056)	67/140	4.61E-16	3.04	128.23	5
HFCA	transmembrane receptor protein tyrosine kinase signaling pathway (GO:0007169)	132/396	1.77E-15	2.11	86.09	6
HFCA	response to cytokine (GO:0034097)	64/138	1.5E-14	2.94	113.03	7
HFCA	cellular response to cytokine stimulus (GO:0071345)	141/456	1.34E-13	1.96	70.81	8
HFCA	Ras protein signal transduction (GO:0007265)	85/223	2.15E-13	2.42	85.85	9
HFCA	cytokine-mediated signaling pathway (GO:0019221)	177/633	9.0E-13	1.77	60.25	10
HFCA	regulation of intracellular signal transduction (GO:1902531)	129/422	5.0E-12	1.94	62.38	11
HFCA	regulation of cell migration (GO:0030334)	104/316	9.51E-12	2.09	65.62	12
HFCA	protein phosphorylation (GO:0006468)	137/470	3.53E-11	1.85	55.40	13
HFCA	positive regulation of intracellular signal transduction (GO:1902533)	139/479	3.67E-11	1.84	55.23	14
HFCA	vascular endothelial growth factor receptor signaling pathway (GO:0048010)	38/70	4.1E-11	3.44	102.43	15
HFCA	Rho protein signal transduction (GO:0007266)	37/72	7.46E-10	3.26	87.30	16
HFCA	platelet degranulation (GO:0002576)	52/124	8.27E-10	2.66	70.80	17
HFCA	regulated exocytosis (GO:0045055)	58/148	1.33E-09	2.49	64.84	18
HFCA	cell-matrix adhesion (GO:0007160)	42/90	1.33E-09	2.96	77.05	19
HFCA	endocytosis (GO:0006897)	86/263	1.48E-09	2.07	53.66	20
HFCA	phosphorylation (GO:0016310)	113/386	2.59E-09	1.86	46.91	21
HFCA	plasma membrane bounded cell projection assembly (GO:0120031)	80/241	3.13E-09	2.11	52.70	22
HFCA	actin filament organization (GO:0007015)	49/120	8.19E-09	2.59	62.22	23
HFCA	regulation of apoptotic process (GO:0042981)	198/815	1.64E-08	1.54	35.88	24
HFCA	regulation of GTPase activity (GO:0043087)	65/188	3.21E-08	2.19	49.51	25
HFCA	regulation of cytoskeleton organization (GO:0051493)	41/95	3.67E-08	2.74	61.32	26
HFCA	ephrin receptor signaling pathway (GO:0048013)	37/82	5.75E-08	2.86	62.72	27
HFCA	positive regulation of cell migration (GO:0030335)	72/221	6.61E-08	2.07	44.92	28
HFCA	positive regulation of I-kappaB kinase/NF-kappaB signaling (GO:0043123)	58/163	7.47E-08	2.26	48.71	29
HFCA	pattern recognition receptor signaling pathway (GO:0002221)	26/48	1.63E-07	3.44	71.34	30
AIN-93M	mitochondrial translation (GO:0032543)	81/107	6.67E-39	4.71	454.44	1
AIN-93M	translational termination (GO:0006415)	73/96	1.2E-35	4.73	414.46	2

AIN-93M	respiratory electron transport chain (GO:0022904)	72/94	1.32E-35	4.77	415.94	3
AIN-93M	mitochondrial translational elongation (GO:0070125)	69/87	1.35E-35	4.94	433.12	4
AIN-93M	translation (GO:0006412)	123/232	2.0E-35	3.30	289.61	5
AIN-93M	mitochondrial translational termination (GO:0070126)	69/89	9.9E-35	4.83	410.41	6
AIN-93M	mitochondrial respiratory chain complex I assembly (GO:0032981)	55/64	4.85E-32	5.35	419.63	7
AIN-93M	mitochondrial respiratory chain complex assembly (GO:0033108)	70/97	5.34E-32	4.49	351.47	8
AIN-93M	mitochondrial respiratory chain complex I biogenesis (GO:0097031)	55/64	5.45E-32	5.35	419.63	9
AIN-93M	NADH dehydrogenase complex assembly (GO:0010257)	55/64	6.23E-32	5.35	419.63	10
AIN-93M	mitochondrial ATP synthesis coupled electron transport (GO:0042775)	64/85	5.1E-31	4.69	355.69	11
AIN-93M	translational elongation (GO:0006414)	71/105	1.26E-29	4.21	305.58	12
AIN-93M	peptide biosynthetic process (GO:0043043)	93/174	2.0E-27	3.33	224.44	13
AIN-93M	cotranslational protein targeting to membrane (GO:0006613)	60/93	3.11E-23	4.02	231.81	14
AIN-93M	protein targeting to ER (GO:0045047)	61/97	8.26E-23	3.91	221.61	15
AIN-93M	SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	58/89	8.69E-23	4.06	229.71	16
AIN-93M	cellular protein metabolic process (GO:0044267)	166/484	5.68E-21	2.13	111.70	17
AIN-93M	mitochondrial electron transport, NADH to ubiquinone (GO:0006120)	38/46	1.06E-20	5.14	265.52	18
AIN-93M	gene expression (GO:0010467)	143/411	1.38E-18	2.17	101.18	19
AIN-93M	cellular macromolecule biosynthetic process (GO:0034645)	129/367	4.23E-17	2.19	94.61	20
AIN-93M	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (GO:0000184)	57/112	3.06E-15	3.17	123.27	21
AIN-93M	viral transcription (GO:0019083)	57/113	5.0E-15	3.14	120.51	22
AIN-93M	viral gene expression (GO:0019080)	55/110	2.87E-14	3.11	113.86	23
AIN-93M	fatty acid catabolic process (GO:0009062)	40/65	2.88E-14	3.83	139.96	24
AIN-93M	nuclear-transcribed mRNA catabolic process (GO:0000956)	73/174	6.41E-14	2.61	93.22	25
AIN-93M	mitochondrion organization (GO:0007005)	71/167	6.75E-14	2.65	94.23	26
AIN-93M	cellular respiration (GO:0045333)	36/57	2.58E-13	3.93	134.56	27
AIN-93M	mitochondrial transport (GO:0006839)	59/135	4.44E-12	2.72	85.27	28
AIN-93M	fatty acid oxidation (GO:0019395)	32/50	5.29E-12	3.98	124.04	29
AIN-93M	mitochondrial gene expression (GO:0140053)	30/45	6.38E-12	4.15	128.29	30

[a]Combined score is described as c = log(p) * z , where c = the combined score, p = Fisher exact test p-value, and z = z-score for deviation from expected rank.

Table 3.4. Top 30 KEGG pathway results for diet-specific DEGs in liver in eight CC founder strains. n = 48 (24 mice for AIN-93M diet and 24 mice for HFCA diet

Diet	Term	Overlap (gene count)	Adjusted P-value	Odds Ratio	Combined Score[a]	Rank
HFCA	Osteoclast differentiation	67/128	2.5E-19	3.32	161.16	1
HFCA	Focal adhesion	86/199	2.1E-18	2.74	124.22	2
HFCA	Pathways in cancer	169/535	2.7E-18	2.00	91.13	3
HFCA	Regulation of actin cytoskeleton	90/217	5.9E-18	2.63	115.76	4
HFCA	Leishmaniasis	42/67	2.6E-16	3.98	159.00	5
HFCA	Endocytosis	98/269	3.7E-15	2.31	85.83	6
HFCA	MAPK signaling pathway	102/294	3.4E-14	2.20	76.52	7
HFCA	Fc gamma R-mediated phagocytosis	46/87	4.7E-14	3.35	115.14	8
HFCA	Tuberculosis	72/178	6.2E-14	2.57	87.07	9
HFCA	Proteoglycans in cancer	78/203	1.1E-13	2.44	80.99	10
HFCA	AGE-RAGE signaling pathway in diabetic complications	49/101	3.5E-13	3.08	98.49	11
HFCA	Bacterial invasion of epithelial cells	40/74	8.4E-13	3.43	106.42	12
HFCA	PI3K-Akt signaling pathway	112/357	2.1E-12	1.99	59.78	13
HFCA	Phagosome	69/180	3.5E-12	2.43	71.61	14
HFCA	Salmonella infection	40/78	6.7E-12	3.25	93.47	15
HFCA	Lysosome	53/124	1.3E-11	2.71	75.94	16
HFCA	Toxoplasmosis	48/108	2.6E-11	2.82	76.85	17
HFCA	C-type lectin receptor signaling pathway	49/112	2.9E-11	2.78	75.13	18
HFCA	Epstein-Barr virus infection	79/229	3.0E-11	2.19	58.98	19
HFCA	Ras signaling pathway	80/233	3.0E-11	2.18	58.79	20
HFCA	Chronic myeloid leukemia	38/76	5.1E-11	3.17	83.61	21
HFCA	Kaposi sarcoma-associated herpesvirus infection	75/216	6.8E-11	2.20	57.34	22
HFCA	Apoptosis	56/141	7.6E-11	2.52	65.19	23
HFCA	Small cell lung cancer	42/92	1.5E-10	2.90	72.77	24
HFCA	Hepatitis B	61/163	1.6E-10	2.37	59.42	25
HFCA	B cell receptor signaling pathway	36/72	1.6E-10	3.17	79.47	26
HFCA	ECM-receptor interaction	39/83	2.4E-10	2.98	73.17	27
HFCA	Leukocyte transendothelial migration	48/115	2.5E-10	2.65	64.85	28
HFCA	Fluid shear stress and atherosclerosis	55/143	4.1E-10	2.44	58.46	29
HFCA	Human cytomegalovirus infection	82/255	4.7E-10	2.04	48.52	30
AIN-93M	Ribosome	90/170	3.3E-26	3.30	212.16	1
AIN-93M	Parkinson disease	80/144	2.4E-25	3.46	213.41	2
AIN-93M	Oxidative phosphorylation	76/134	5.3E-25	3.53	213.66	3
AIN-93M	Huntington disease	89/192	3.5E-21	2.89	148.43	4
AIN-93M	Non-alcoholic fatty liver disease (NAFLD)	76/151	7.3E-21	3.13	158.12	5
AIN-93M	Alzheimer disease	82/175	4.9E-20	2.92	141.11	6

AIN-93M	Thermogenesis	96/231	6.5E-19	2.59	118.08	7
AIN-93M	Peroxisome	51/84	8.0E-19	3.78	171.22	8
AIN-93M	Glyoxylate and dicarboxylate metabolism	22/31	4.9E-10	4.42	110.26	9
AIN-93M	Citrate cycle (TCA cycle)	21/32	1.3E-08	4.08	88.22	10
AIN-93M	Valine, leucine and isoleucine degradation	29/56	1.9E-08	3.22	68.03	11
AIN-93M	Propanoate metabolism	20/31	4.4E-08	4.02	81.04	12
AIN-93M	Protein processing in endoplasmic reticulum	57/163	6.1E-08	2.18	43.03	13
AIN-93M	Pyruvate metabolism	22/38	1.1E-07	3.60	68.86	14
AIN-93M	Steroid biosynthesis	14/19	7.8E-07	4.59	78.29	15
AIN-93M	Proteasome	23/46	1.8E-06	3.11	50.39	16
AIN-93M	Butanoate metabolism	16/27	7.4E-06	3.69	54.20	17
AIN-93M	Glycine, serine and threonine metabolism	20/40	1.1E-05	3.11	44.38	18
AIN-93M	Complement and coagulation cascades	33/88	1.5E-05	2.33	32.43	19
AIN-93M	Porphyrin and chlorophyll metabolism	20/41	1.6E-05	3.04	41.76	20
AIN-93M	Drug metabolism	39/114	2.2E-05	2.13	28.54	21
AIN-93M	Protein export	15/28	7.2E-05	3.33	40.42	22
AIN-93M	Tryptophan metabolism	21/48	7.2E-05	2.72	33.12	23
AIN-93M	Retrograde endocannabinoid signaling	46/150	7.2E-05	1.91	23.04	24
AIN-93M	Fatty acid elongation	15/29	1.2E-04	3.22	37.20	25
AIN-93M	Fatty acid degradation	21/50	1.3E-04	2.61	29.76	26
AIN-93M	Ascorbate and aldarate metabolism	14/27	2.1E-04	3.23	35.19	27
AIN-93M	Folate biosynthesis	13/26	6.5E-04	3.11	30.23	28
AIN-93M	Selenocompound metabolism	10/17	7.7E-04	3.66	34.85	29
AIN-93M	Cholesterol metabolism	19/49	1.1E-03	2.41	22.02	30

[a]Combined score is described as c = log(p) * z , where c = the combined score, p = Fisher exact test p-value, and z = z-score for deviation from expected rank.

3.5.2. DEGs between Two Diets by Eight Found Strains Vary in Biological Pathways.

To investigate the effect of genetic background on the global hepatic gene expression, we next performed transcriptomic analysis of strain differences among progenitor strains of the CC. Hierarchical clustering and principal component analysis of the liver transcriptome demonstrate that transcript abundance was highly variable across the eight founder strains. This analysis reveals several interesting findings. First, the wild-derived founder strains, CAST and PWK, are distinct in their response to the other CC progenitor strains. Moreover, this difference is apparent for both diets (**Figures 3.3A and 3.3B**). Second, the B6, 129, NOD, and WSB strains showed distinct liver transcriptome patterns by diet and the NZO, CAST, and PWK strains were less affected by diet than other strains based on the hierarchical clustering (**Figure 3.3B)**. Finally, the AJ strain showed the most extreme diet response (**Figure 3.3B**). Based on these results, we performed a DEG analysis between the HFCA diet-fed mice and the AIN-93M diet-fed mice for each strain to identify which strain was most or less responsive to the diet. As shown in **Table 3.5**, among the eight strains, the A/J strain showed the most significant difference in hepatic gene expression between the two diets (2,862 genes in the HFCA diet and 3,186 genes in the AIN-93M diet), and CAST was found to have a subdued diet response (64 genes in HFCA diet and 49 genes in AIN-93M diet), although we acknowledge a variable response to the HFCA in the CAST replicates. Furthermore, among diet-specific upregulated genes for each strain, the ratio of uniquely upregulated genes only in that strain was the highest in A/J strain (18.6% in HFCA diet and 27.7% in AIN-93M diet).

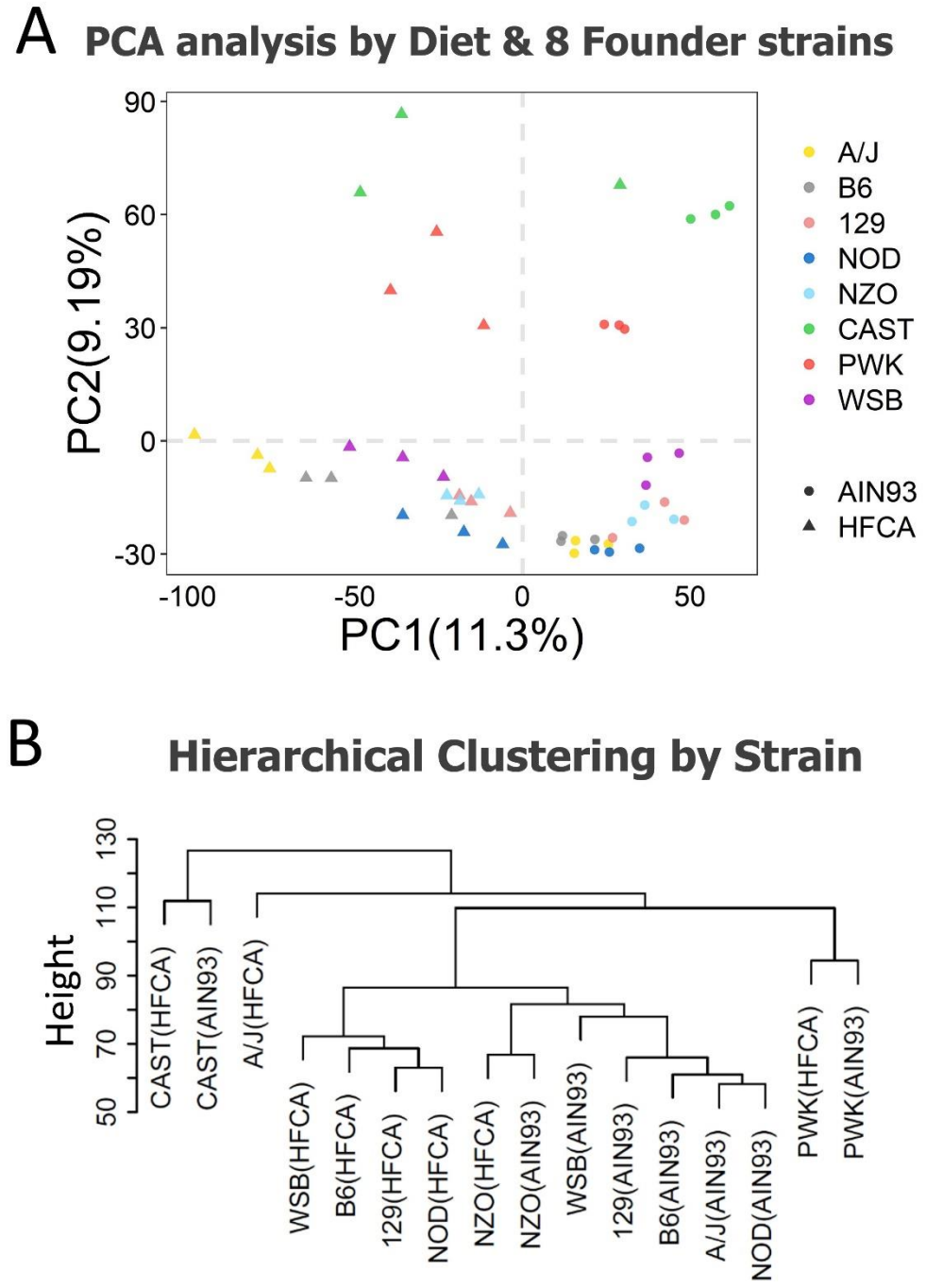


Figure 3.3. Effect of genetic background on hepatic gene expression in the eight CC founder strains in female mice.
(A) Principal component analysis from AIN-93M diet or HFCA diet and (B) hierarchical clustering determined that the major source of variation in gene expression was due to genetic variation among the eight strains.

Based on the result of strain-specific DEGs between HFCA diet and AIN-93M diet, we also compared the differences in GO Biological Process 2018 terms and KEGG 2019 Mouse pathways between the two diets for each strain. We detected the highest number of 616 GO terms and 154 KEGG pathways in A/J while the number of GO (9 terms) and KEGG (9 pathways) in CAST was the least (**Table 3.6**). Therefore, we targeted the A/J and CAST strains, which showed the most extreme differences in DEG analysis. The top upregulated genes in each diet differed significantly between the two strains. For example, only 4 genes were overlapped in the top 100 upregulated genes between the two strains in both diets (**Figures 3.4A and 3.4B**). In the comparison of the A/J strain (most diet responsive) and the B6 strain (most utilized resource), we also found that 66.3% of the upregulated genes (1,899 out of 2,862 genes) in the HFCA diet-fed A/J strain and 76.5% of the upregulated genes (2,437 out of 3,186 genes) in the AIN-93M diet-fed A/J strain were uniquely identified only in the A/J strain (**Figure 3.4C**). Likewise, 60.9% of upregulated genes (39 out of 64 genes) in HFCA diet-fed CAST strain and 61.2% of upregulated genes (30 out of 49 genes) in AIN-93M diet-fed CAST strain were uniquely identified only in the least diet responsive CAST strain (**Figure 3.4D**).

Table 3.5. The number of DEGs between high-fat cholic acid (HFCA) diet and AIN-93M diet for each strain in liver gene expression in eight CC founder strains. n = 48 (6 mice per eight strains and 24 mice per two diets)

Strain	HFCA diet-upregulated genes	HFCA diet-upregulated genes that are unique to the strain	AIN-93M diet-upregulated genes	AIN-93M diet-upregulated genes that are unique to the strain
A/J	2,862	532 (18.6%)	3,186	881 (27.7%)
B6	1,209	86 (7.1%)	962	72 (7.5%)
129	1,352	77 (5.7%)	772	85 (11.0%)
NOD	1,141	103 (9.0%)	750	103 (13.7%)
NZO	2,438	452 (18.5%)	2,225	362 (16.3%)
CAST	64	6 (9.4%)	49	9 (18.4%)
PWK	414	63 (15.2%)	490	88 (18.0%)
WSB	2,482	322 (13.0%)	2,006	380 (18.9%)
At least one strain	4,503		4,725	
Overlap between all-strain and individual strain	2,958		2,885	

Table 3.6. The number of GO Biological Process 2018 terms and KEGG 2019 Mouse pathways between upregulated genes in HFCA diet, upregulated genes in AIN-93M diet, or non-diet responsive genes in enrichment analysis for each strain. n = 48 (6 mice per eight strains and 24 mice per two diets)

Strain	HFCA diet-upregulated genes		AIN-93M diet-upregulated genes	
	#GO	#KEGG	#GO	#KEGG
A/J	616	154	230	64
B6	155	86	69	28
129	309	108	48	28
NOD	216	89	21	13
NZO	479	119	191	53
CAST	9	9	1	4
PWK	82	35	59	20
WSB	477	124	163	55

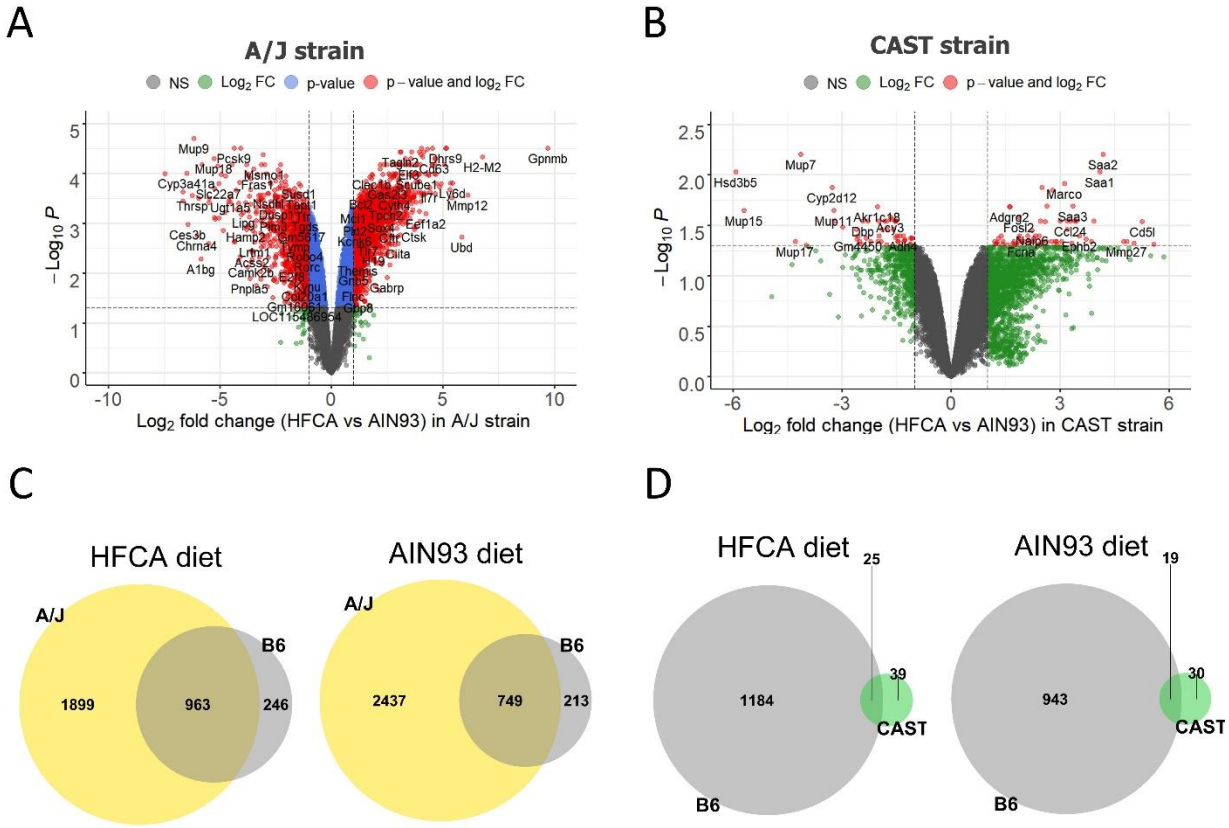


Figure 3.4. Comparison of hepatic gene expression between A/J and CAST strains in female mice.

(A,B) Venn diagrams to identify overlapping upregulated genes between A/J strain and B6 strain (A) and between CAST strain and B6 strain (B) in each diet. (C,D) Volcano plots between high-fat and cholic acid (HFCA) diet and AIN-93M diet in liver gene expression in A/J strain (C) and CAST strain (D). Horizontal dotted lines indicate adj.P < 0.05, vertical dotted gray lines indicate a 2-fold difference.

In terms of the GO Biological Process between the two strains, we found the immune response-related GO terms in both strains (**Table 3.7**) and the A/J strain had a high richness in each GO term and the neutrophil related immune response GO terms were mainly enriched (**Figure 3.5A**) as shown in **Figure 3.2B**. On the other hand, the CAST strain had a low richness in GO term and mainly enriched in GO terms such as calcium signaling, hydrogen peroxide, and acute inflammatory response (**Figure 3.5B**). These results show that the hepatic gene expression pattern responding to the diet is contains both Core Diet DEGs and DEGs unique to each strain.

Table 3.7. Top 10 Gene Ontology results for diet-specific DEGs in each strain in eight CC founder strains. n = 48 (3 mice in HFCA diet and 3 mice in AIN-93M diet per strain)

Strain	Diet	Term	Overlap (gene count)	Adjusted P-value	Odds Ratio	Combined Score[a]	Rank
AJ	HFCA	neutrophil mediated immunity (GO:0002446)	185/487	9.31E-36	3.86	344.20	1
AJ	HFCA	neutrophil activation involved in immune response (GO:0002283)	182/483	4.94E-35	3.81	328.97	2
AJ	HFCA	neutrophil degranulation (GO:0043312)	181/479	4.94E-35	3.83	330.26	3
AJ	HFCA	extracellular matrix organization (GO:0030198)	97/229	2.82E-22	4.53	256.77	4
AJ	HFCA	transmembrane receptor protein tyrosine kinase signaling pathway (GO:0007169)	119/396	2.78E-13	2.65	94.59	5
AJ	HFCA	regulation of cell migration (GO:0030334)	98/316	1.11E-11	2.76	87.87	6
AJ	HFCA	positive regulation of intracellular signal transduction (GO:1902533)	128/479	2.62E-10	2.25	64.07	7
AJ	HFCA	protein phosphorylation (GO:0006468)	126/470	2.73E-10	2.25	63.93	8
AJ	HFCA	Ras protein signal transduction (GO:0007265)	74/223	2.79E-10	3.04	85.39	9
AJ	HFCA	Rho protein signal transduction (GO:0007266)	36/72	2.79E-10	6.07	170.20	10
AJ	AIN-93M	respiratory electron transport chain (GO:0022904)	69/94	4.3E-32	15.17	1221.84	1
AJ	AIN-93M	mitochondrial translation (GO:0032543)	71/107	7.89E-29	10.84	784.15	2
AJ	AIN-93M	mitochondrial ATP synthesis coupled electron transport (GO:0042775)	61/85	1.17E-27	13.93	964.80	3
AJ	AIN-93M	mitochondrial respiratory chain complex assembly (GO:0033108)	65/97	5.85E-27	11.14	749.25	4
AJ	AIN-93M	NADH dehydrogenase complex assembly (GO:0010257)	51/64	5.85E-27	21.45	1432.58	5
AJ	AIN-93M	mitochondrial respiratory chain complex I biogenesis (GO:0097031)	51/64	5.85E-27	21.45	1432.58	6
AJ	AIN-93M	mitochondrial respiratory chain complex I assembly (GO:0032981)	51/64	5.85E-27	21.45	1432.58	7
AJ	AIN-93M	mitochondrial translational termination (GO:0070126)	58/89	3.61E-23	10.24	593.31	8
AJ	AIN-93M	mitochondrial translational elongation (GO:0070125)	56/87	5.69E-22	9.88	544.06	9
AJ	AIN-93M	translational termination (GO:0006415)	59/96	1.01E-21	8.73	474.63	10
B6	HFCA	neutrophil activation involved in immune response (GO:0002283)	84/483	2.88E-15	3.45	143.94	1
B6	HFCA	neutrophil degranulation (GO:0043312)	83/479	2.96E-15	3.44	140.73	2
B6	HFCA	neutrophil mediated immunity (GO:0002446)	83/487	5.59E-15	3.37	134.38	3
B6	HFCA	extracellular matrix organization (GO:0030198)	53/229	8.82E-15	4.87	190.59	4
B6	HFCA	cytokine-mediated signaling pathway (GO:0019221)	77/633	2.09E-06	2.24	44.01	5
B6	HFCA	cellular response to cytokine stimulus (GO:0071345)	60/456	5.78E-06	2.43	44.93	6
B6	HFCA	collagen fibril organization (GO:0030199)	12/29	2.21E-05	11.11	188.43	7
B6	HFCA	extracellular matrix disassembly (GO:0022617)	19/78	5.72E-05	5.09	80.76	8
B6	HFCA	protein complex subunit organization (GO:0071822)	14/45	8.31E-05	7.12	109.06	9

B6	HFCA	platelet aggregation (GO:0070527)	12/33	8.31E-05	8.99	137.24	10
B6	AIN-93M	regulation of steroid biosynthetic process (GO:0050810)	22/44	1.23E-14	20.57	822.15	1
B6	AIN-93M	regulation of alcohol biosynthetic process (GO:1902930)	19/34	7.49E-14	25.99	973.49	2
B6	AIN-93M	cholesterol biosynthetic process (GO:0006695)	19/35	8.24E-14	24.36	894.63	3
B6	AIN-93M	sterol biosynthetic process (GO:0016126)	20/40	8.24E-14	20.53	748.40	4
B6	AIN-93M	regulation of cholesterol biosynthetic process (GO:0045540)	20/40	8.24E-14	20.53	748.40	5
B6	AIN-93M	secondary alcohol biosynthetic process (GO:1902653)	19/36	1.06E-13	22.93	825.79	6
B6	AIN-93M	regulation of cholesterol metabolic process (GO:0090181)	20/41	1.1E-13	19.55	700.53	7
B6	AIN-93M	respiratory electron transport chain (GO:0022904)	27/94	5.46E-12	8.32	264.36	8
B6	AIN-93M	cholesterol metabolic process (GO:0008203)	23/68	8.53E-12	10.51	328.30	9
B6	AIN-93M	mitochondrial ATP synthesis coupled electron transport (GO:0042775)	24/85	1.91E-10	8.10	226.73	10
129	HFCA	neutrophil mediated immunity (GO:0002446)	113/487	1.11E-28	4.47	324.58	1
129	HFCA	neutrophil activation involved in immune response (GO:0002283)	112/483	1.12E-28	4.47	320.95	2
129	HFCA	neutrophil degranulation (GO:0043312)	111/479	1.52E-28	4.46	317.31	3
129	HFCA	cytokine-mediated signaling pathway (GO:0019221)	107/633	4.65E-16	2.97	125.20	4
129	HFCA	cellular response to cytokine stimulus (GO:0071345)	83/456	4.66E-14	3.22	119.95	5
129	HFCA	extracellular matrix organization (GO:0030198)	46/229	1.12E-08	3.57	88.20	6
129	HFCA	inflammatory response (GO:0006954)	48/252	2.5E-08	3.34	79.39	7
129	HFCA	toll-like receptor signaling pathway (GO:0002224)	25/86	1.1E-07	5.76	127.65	8
129	HFCA	response to cytokine (GO:0034097)	32/138	2.02E-07	4.25	91.19	9
129	HFCA	regulation of small GTPase mediated signal transduction (GO:0051056)	32/140	2.6E-07	4.18	87.87	10
129	AIN-93M	cholesterol biosynthetic process (GO:0006695)	19/35	4.72E-15	30.89	1262.05	1
129	AIN-93M	regulation of alcohol biosynthetic process (GO:1902930)	18/34	3.46E-14	29.23	1115.50	2
129	AIN-93M	sterol biosynthetic process (GO:0016126)	19/40	4.25E-14	23.53	883.71	3
129	AIN-93M	secondary alcohol biosynthetic process (GO:1902653)	18/36	6.64E-14	25.98	956.54	4
129	AIN-93M	regulation of cholesterol biosynthetic process (GO:0045540)	18/40	5.75E-13	21.25	731.83	5
129	AIN-93M	regulation of cholesterol metabolic process (GO:0090181)	18/41	8.24E-13	20.33	688.95	6
129	AIN-93M	regulation of steroid biosynthetic process (GO:0050810)	18/44	3.23E-12	17.98	582.01	7
129	AIN-93M	cholesterol metabolic process (GO:0008203)	19/68	1.5E-09	10.07	262.87	8
129	AIN-93M	dicarboxylic acid metabolic process (GO:0043648)	14/59	8.83E-06	8.03	138.91	9
129	AIN-93M	monocarboxylic acid metabolic process (GO:0032787)	18/108	3.14E-05	5.18	82.45	10
NOD	HFCA	neutrophil activation involved in immune response (GO:0002283)	92/483	1.32E-21	4.15	233.14	1
NOD	HFCA	neutrophil degranulation (GO:0043312)	91/479	1.48E-21	4.13	228.88	2
NOD	HFCA	neutrophil mediated immunity (GO:0002446)	91/487	3.39E-21	4.05	219.17	3

NOD	HFCA	cytokine-mediated signaling pathway (GO:0019221)	93/633	2E-14	3.02	115.38	4
NOD	HFCA	cellular response to cytokine stimulus (GO:0071345)	66/456	1.62E-09	2.91	77.86	5
NOD	HFCA	pattern recognition receptor signaling pathway (GO:0002221)	19/48	2.53E-09	11.02	287.48	6
NOD	HFCA	regulation of small GTPase mediated signal transduction (GO:0051056)	31/140	2.31E-08	4.81	114.24	7
NOD	HFCA	inflammatory response (GO:0006954)	42/252	1.35E-07	3.40	74.25	8
NOD	HFCA	toll-like receptor signaling pathway (GO:0002224)	22/86	6.37E-07	5.78	116.64	9
NOD	HFCA	positive regulation of cytokine production (GO:0001819)	37/220	9.38E-07	3.43	67.41	10
NOD	AIN-93M	sterol biosynthetic process (GO:0016126)	13/40	2.55E-06	12.77	264.54	1
NOD	AIN-93M	secondary alcohol biosynthetic process (GO:1902653)	12/36	3.2E-06	13.25	259.03	2
NOD	AIN-93M	regulation of steroid biosynthetic process (GO:0050810)	13/44	3.2E-06	11.12	215.62	3
NOD	AIN-93M	regulation of cholesterol metabolic process (GO:0090181)	12/41	1.06E-05	10.96	195.98	4
NOD	AIN-93M	regulation of alcohol biosynthetic process (GO:1902930)	11/34	1.06E-05	12.66	223.68	5
NOD	AIN-93M	cholesterol biosynthetic process (GO:0006695)	11/35	1.25E-05	12.13	210.19	6
NOD	AIN-93M	regulation of cholesterol biosynthetic process (GO:0045540)	11/40	5.01E-05	10.03	158.41	7
NOD	AIN-93M	steroid biosynthetic process (GO:0006694)	13/79	0.001857	5.22	62.79	8
NOD	AIN-93M	monocarboxylic acid metabolic process (GO:0032787)	15/108	0.002804	4.28	49.22	9
NOD	AIN-93M	purine nucleobase metabolic process (GO:0006144)	5/11	0.006518	21.89	231.20	10
NZO	HFCA	neutrophil mediated immunity (GO:0002446)	130/487	2.28E-15	2.72	113.06	1
NZO	HFCA	neutrophil activation involved in immune response (GO:0002283)	129/483	2.28E-15	2.72	112.39	2
NZO	HFCA	neutrophil degranulation (GO:0043312)	128/479	2.28E-15	2.72	111.72	3
NZO	HFCA	regulation of cell migration (GO:0030334)	95/316	1.3E-14	3.19	124.35	4
NZO	HFCA	regulation of small GTPase mediated signal transduction (GO:0051056)	56/140	4.85E-14	4.90	183.63	5
NZO	HFCA	extracellular matrix organization (GO:0030198)	75/229	1.37E-13	3.60	130.29	6
NZO	HFCA	regulation of intracellular signal transduction (GO:1902531)	110/422	1.75E-12	2.62	87.81	7
NZO	HFCA	response to cytokine (GO:0034097)	50/138	1.38E-10	4.17	120.98	8
NZO	HFCA	positive regulation of intracellular signal transduction (GO:1902533)	115/479	1.5E-10	2.34	67.59	9
NZO	HFCA	cytokine-mediated signaling pathway (GO:0019221)	139/633	6.67E-10	2.09	57.02	10
NZO	AIN-93M	respiratory electron transport chain (GO:0022904)	64/94	6.14E-36	17.85	1594.28	1
NZO	AIN-93M	mitochondrial ATP synthesis coupled electron transport (GO:0042775)	54/85	6.5E-28	14.51	1017.52	2
NZO	AIN-93M	mitochondrial respiratory chain complex assembly (GO:0033108)	56/97	4.46E-26	11.38	745.41	3
NZO	AIN-93M	NADH dehydrogenase complex assembly (GO:0010257)	44/64	4.81E-25	18.25	1139.27	4
NZO	AIN-93M	mitochondrial respiratory chain complex I biogenesis (GO:0097031)	44/64	4.81E-25	18.25	1139.27	5

NZO	AIN-93M	mitochondrial respiratory chain complex I assembly (GO:0032981)	44/64	4.81E-25	18.25	1139.27	6
NZO	AIN-93M	mitochondrial electron transport, NADH to ubiquinone (GO:0006120)	33/46	2.06E-19	20.96	1033.42	7
NZO	AIN-93M	mitochondrial transport (GO:0006839)	51/135	1.3E-13	5.03	180.36	8
NZO	AIN-93M	cholesterol biosynthetic process (GO:0006695)	23/35	6.24E-12	15.75	500.21	9
NZO	AIN-93M	mitochondrial translational termination (GO:0070126)	38/89	6.24E-12	6.15	195.21	10
CAST	HFCA	calcium-mediated signaling using intracellular calcium source (GO:0035584)	3/18	0.013	65.31	693.06	1
CAST	HFCA	calcium-mediated signaling (GO:0019722)	4/70	0.020	20.07	190.74	2
CAST	HFCA	response to hydrogen peroxide (GO:0042542)	3/43	0.038	24.46	194.52	3
CAST	HFCA	regulation of protein activation cascade (GO:2000257)	4/108	0.038	12.71	99.52	4
CAST	HFCA	regulation of complement activation (GO:0030449)	4/109	0.038	12.59	98.13	5
CAST	HFCA	regulation of humoral immune response (GO:0002920)	4/113	0.038	12.13	92.85	6
CAST	HFCA	regulation of immune effector process (GO:0002697)	4/114	0.038	12.02	91.60	7
CAST	HFCA	regulation of acute inflammatory response (GO:0002673)	4/121	0.042	11.29	83.55	8
CAST	HFCA	regulation of protein processing (GO:0070613)	4/128	0.046	10.65	76.56	9
CAST	AIN-93M	sterol biosynthetic process (GO:0016126)	3/40	0.028	35.10	314.60	1
PWK	HFCA	actin cytoskeleton reorganization (GO:0031532)	10/61	0.001	9.48	138.44	1
PWK	HFCA	cortical actin cytoskeleton organization (GO:0030866)	6/21	3.22.E-03	19.19	242.99	2
PWK	HFCA	cellular response to organic cyclic compound (GO:0071407)	13/135	3.22.E-03	5.17	63.42	3
PWK	HFCA	leukocyte aggregation (GO:0070486)	4/7	3.22.E-03	63.68	765.43	4
PWK	HFCA	negative regulation of intracellular signal transduction (GO:1902532)	14/161	3.22.E-03	4.63	55.08	5
PWK	HFCA	negative regulation of small GTPase mediated signal transduction (GO:0051058)	6/27	6.14.E-03	13.70	151.71	6
PWK	HFCA	myelin assembly (GO:0032288)	4/9	6.44.E-03	38.21	411.53	7
PWK	HFCA	cortical cytoskeleton organization (GO:0030865)	6/29	6.44.E-03	12.51	133.02	8
PWK	HFCA	neutrophil degranulation (GO:0043312)	25/479	6.44.E-03	2.71	28.76	9
PWK	HFCA	neutrophil activation involved in immune response (GO:0002283)	25/483	6.65.E-03	2.68	28.13	10
PWK	AIN-93M	regulation of alcohol biosynthetic process (GO:1902930)	19/34	2.14E-19	53.47	2682.75	1
PWK	AIN-93M	sterol biosynthetic process (GO:0016126)	20/40	2.14E-19	42.30	2103.84	2
PWK	AIN-93M	cholesterol biosynthetic process (GO:0006695)	19/35	2.14E-19	50.13	2476.82	3
PWK	AIN-93M	secondary alcohol biosynthetic process (GO:1902653)	19/36	3.32E-19	47.18	2296.65	4
PWK	AIN-93M	regulation of steroid biosynthetic process (GO:0050810)	20/44	1.08E-18	35.24	1666.19	5
PWK	AIN-93M	regulation of cholesterol biosynthetic process (GO:0045540)	19/40	3.09E-18	38.18	1758.12	6
PWK	AIN-93M	regulation of cholesterol metabolic process (GO:0090181)	19/41	4.83E-18	36.45	1656.24	7

PWK	AIN-93M	cholesterol metabolic process (GO:0008203)	18/68	5.11E-12	15.14	475.91	8
PWK	AIN-93M	IRE1-mediated unfolded protein response (GO:0036498)	13/57	1.46E-07	12.29	258.91	9
PWK	AIN-93M	acetyl-CoA metabolic process (GO:0006084)	7/14	2.43E-06	41.16	746.81	10
WSB	HFCA	neutrophil mediated immunity (GO:0002446)	160/487	2.06E-29	3.63	270.38	1
WSB	HFCA	neutrophil activation involved in immune response (GO:0002283)	158/483	3.78E-29	3.61	263.04	2
WSB	HFCA	neutrophil degranulation (GO:0043312)	157/479	3.78E-29	3.62	263.00	3
WSB	HFCA	cytokine-mediated signaling pathway (GO:0019221)	167/633	2.23E-19	2.65	132.19	4
WSB	HFCA	regulation of small GTPase mediated signal transduction (GO:0051056)	54/140	2.3E-12	4.52	151.71	5
WSB	HFCA	regulation of cell migration (GO:0030334)	89/316	2E-11	2.84	88.68	6
WSB	HFCA	response to cytokine (GO:0034097)	51/138	7.11E-11	4.22	125.57	7
WSB	HFCA	cellular response to cytokine stimulus (GO:0071345)	113/456	8.94E-11	2.40	70.49	8
WSB	HFCA	extracellular matrix organization (GO:0030198)	68/229	1.03E-09	3.05	81.83	9
WSB	HFCA	platelet degranulation (GO:0002576)	45/124	2.84E-09	4.09	105.24	10
WSB	AIN-93M	respiratory electron transport chain (GO:0022904)	68/94	4.17E-44	24.70	2668.91	1
WSB	AIN-93M	mitochondrial ATP synthesis coupled electron transport (GO:0042775)	59/85	1.41E-36	21.33	1920.24	2
WSB	AIN-93M	mitochondrial respiratory chain complex assembly (GO:0033108)	52/97	3.69E-24	10.81	659.75	3
WSB	AIN-93M	mitochondrial electron transport, NADH to ubiquinone (GO:0006120)	34/46	4.83E-22	26.31	1467.97	4
WSB	AIN-93M	NADH dehydrogenase complex assembly (GO:0010257)	40/64	4.83E-22	15.51	858.01	5
WSB	AIN-93M	mitochondrial respiratory chain complex I biogenesis (GO:0097031)	40/64	4.83E-22	15.51	858.01	6
WSB	AIN-93M	mitochondrial respiratory chain complex I assembly (GO:0032981)	40/64	4.83E-22	15.51	858.01	7
WSB	AIN-93M	fatty acid beta-oxidation (GO:0006635)	33/50	4.11E-19	18.01	872.20	8
WSB	AIN-93M	fatty acid catabolic process (GO:0009062)	36/65	2.59E-17	11.53	509.13	9
WSB	AIN-93M	fatty acid oxidation (GO:0019395)	31/50	8.83E-17	15.12	647.70	10

[a]Combined score is described as c = log(p) * z , where c = the combined score, p = Fisher exact test p-value, and z = z-score for deviation from expected rank.

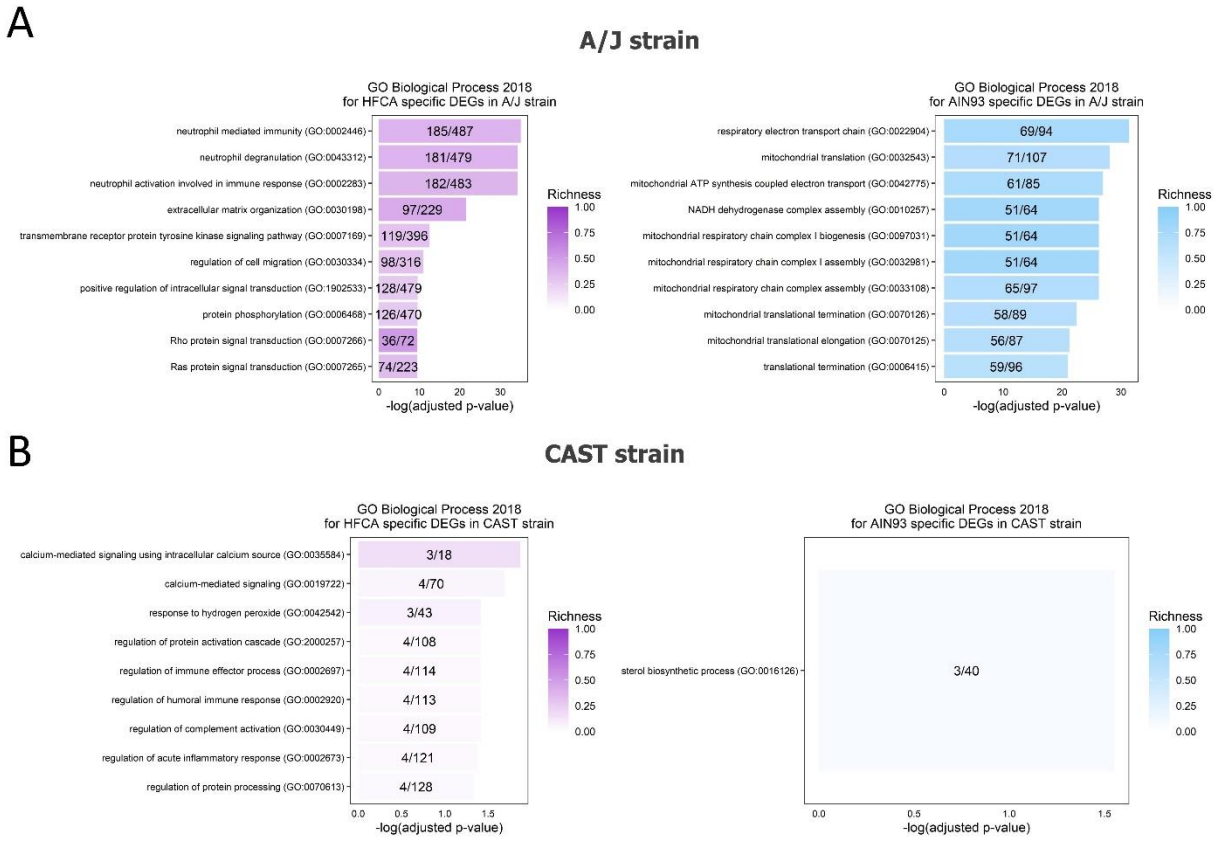


Figure 3.5. Comparison of functional enrichment in liver transcriptome between A/J and CAST strains in female mice.

(A,B) Top 10 GO terms of upregulated genes in HFCA diet and AIN-93M diet identified in enrichment analysis in A/J strain (A) and CAST strain (B). Pathways were ordered from top to bottom by significance (highest to lowest) and colored by gene richness. N = 6 mice for each founder, 3 for AIN-93M diet and 3 for HFCA diet-fed mice. A/J (yellow), B6 (gray), C57BL/6J; 129 (pink), 129S1/SvlmJ; NOD (blue), NOD/ShiLtJ; NZO (lightblue), NZO/HILtJ; CAST (green), CAST/EiJ; PWJ (red), PWK/PhJ; WSB (purple), WSB/EiJ.

3.5.3. Hepatic Transcriptional Network of CC Progenitors Enriched in Specific Functional Pathways

In addition to identifying transcripts whose abundance is affected by diet or genetic background, we were also interested in understanding the hepatic transcriptional network in the CC progenitors and its relationship to disease-related traits. For this reason, we performed weighted gene co-expression network analysis (WGCNA) to identify modules of highly co-expressed genes. Using WGCNA we identified 20 co-expression gene modules in the liver transcriptome data as indicated by color names excluding a grey module containing unclustered genes in the network (**Figures 3.6A and 3.6B**). The modules contain varying numbers of transcripts ranging from 44 to 4,020. A cluster dendrogram marks the modules as downward branches (**Figure 3.6A**). We calculated module eigengenes (ME) of each module to assess the transcript abundance pattern among the eight CC founder strains and two diets. Using GO and KEGG enrichment analysis we identify 10 modules enriched for specific biological pathways (**Table 3.8, 3.9, and 3.10)**. The functional annotations of these ten modules include: “Fatty acid catabolic process (GO: 0009062)” for the brown module, “Steroid biosynthesis degradation” (KEGG pathway) for the green module, "Type 1 interferon signaling pathway (GO: 0060337)” for the magenta module, “mRNA processing (GO: 0006397)” for the red module, and “Neutrophil activation involved in immune response (GO:0002283)” for the turquoise module.

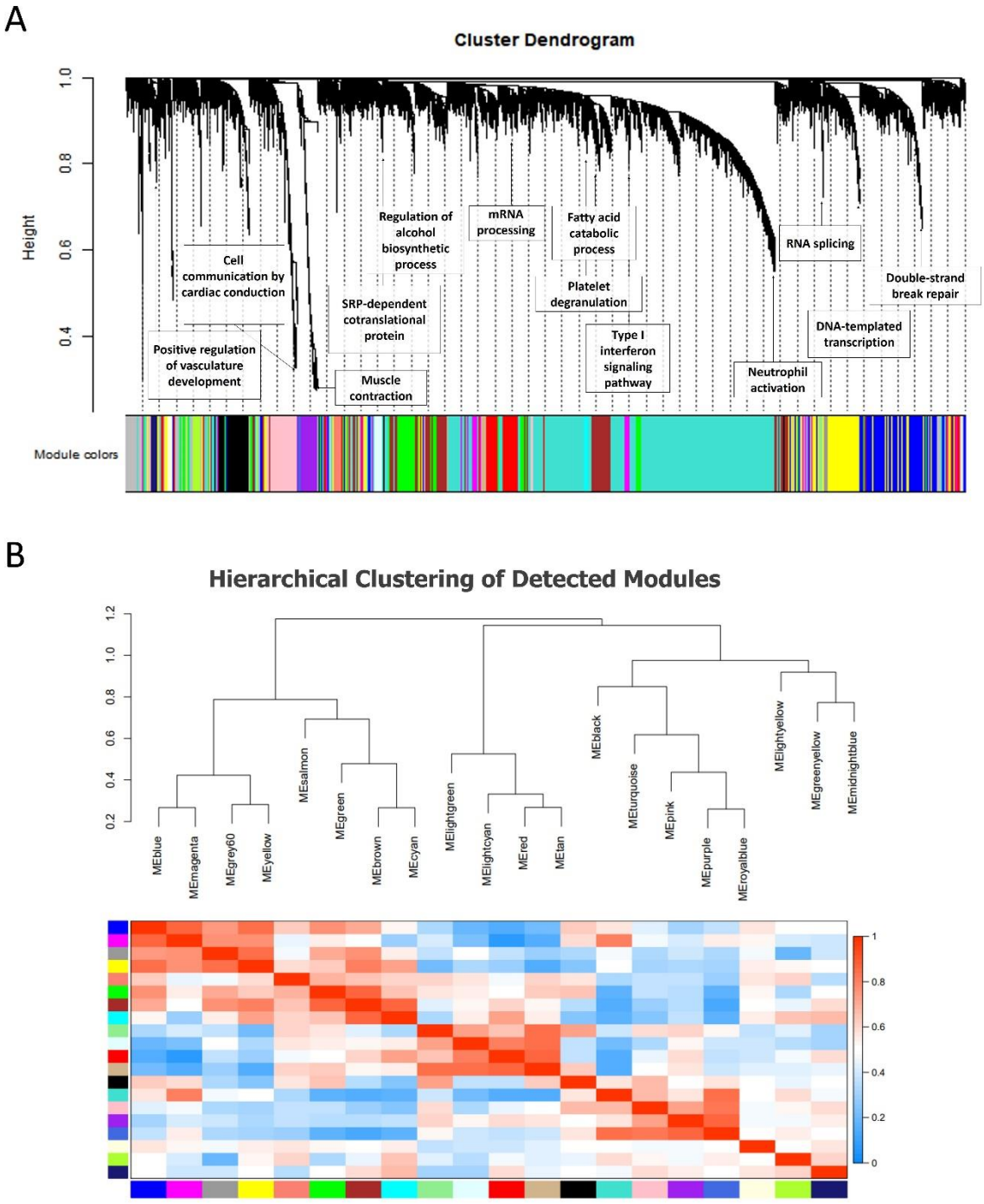


Figure 3.6. Effect of genetic background on hepatic co-expression gene modules in the eight CC founder strains in female mice.
(A) Cluster dendrogram illustrates modules (denoted by color and shown as downward branches) of highly correlated transcripts in livers. (B) Unsupervised hierarchical clustering dendrogram of 19 module eigengenes.

Table 3.8. Gene set enrichment determined that each of the clusters was specifically enriched for GO biological process, KEGG pathway, and Jensen Disease.

Module (#Genes)	GO Biological Process KEGG pathway Jensen Disease	Counts	Adjusted P-value
turquoise (4020)	Neutrophil activation involved in immune response (GO:0002283) Non-alcoholic fatty liver disease (NAFLD) Arthritis	187 81 63	3.1×10^{-18} 1.8×10^{-17} 1.3×10^{-2}
brown (1151)	Fatty acid catabolic process (GO:0009062) Valine, leucine and isoleucine degradation 3-Methylcrotonyl-CoA carboxylase deficiency	21 21 6	2.1×10^{-7} 4.5×10^{-10} 1.1×10^{-2}
green (843)	Regulation of alcohol biosynthetic process (GO:1902930) Steroid biosynthesis	20 19	1.0×10^{-15} 1.2×10^{-8}
red (700)	mRNA processing (GO:0006397) FoxO signaling pathway	29 14	1.3×10^{-3} 2.2×10^{-2}
pink (476)	Cell communication by electrical coupling involved in cardiac conduction (GO:0086064) Cortisol synthesis and secretion Congenital adrenal hyperplasia	6 20 9	7.6×10^{-4} 2.7×10^{-14} 1.9×10^{-5}
magenta (441)	Type I interferon signaling pathway (GO:0060337) DNA replication Aicardi-Goutieres syndrome	18 10 4	5.3×10^{-12} 8.3×10^{-7} 1.4×10^{-2}

purple (259)	Muscle contraction (GO:0006936) Dilated cardiomyopathy (DCM) Cardiomyopathy	50 14 9	4.4×10^{-56} 3.0×10^{-9} 9.2×10^{-12}
cyan (136)	Platelet degranulation (GO:0002576) Complement and coagulation cascades Congenital afibrinogenemia	13 13 8	1.4×10^{-8} 8.7×10^{-12} 1.4×10^{-6}
midnightblue (127)	Positive regulation of vasculature development (GO:1904018) Cell adhesion molecules (CAMs) Ovarian hyperstimulation syndrome	10 7 4	3.0×10^{-6} 1.6×10^{-2} 3.6×10^{-3}
lightcyan (125)	SSRP-dependent cotranslational protein targeting to membrane (GO:0006614) Ribosome Diamond-Blackfan anemia	38 40 4	2.8×10^{-58} 1.5×10^{-50} 3.4×10^{-3}

Table 3.9. Top 10 Gene Ontology results for 10 hepatic coexpression gene modules in liver in eight CC founder strains. n = 48 (6 mice per eight strains and 24 mice per two diets).

Module	Term	Overlap (gene count)	Adjusted P-value	Odds Ratio	Combined Score[a]	Rank
brown	fatty acid catabolic process (GO:0009062)	21/65	2.08E-07	5.61	134.30	1
brown	fatty acid oxidation (GO:0019395)	18/50	2.23E-07	6.26	142.36	2
brown	branched-chain amino acid catabolic process (GO:0009083)	12/21	2.88E-07	9.93	220.53	3
brown	fatty acid beta-oxidation (GO:0006635)	18/50	3.34E-07	6.26	142.36	4
brown	branched-chain amino acid metabolic process (GO:0009081)	12/22	4.81E-07	9.48	203.55	5
brown	cellular amino acid catabolic process (GO:0009063)	12/23	7.94E-07	9.07	188.50	6
brown	alpha-amino acid catabolic process (GO:1901606)	15/39	1.14E-06	6.68	135.51	7
brown	fatty acid beta-oxidation using acyl-CoA dehydrogenase (GO:0033539)	8/15	3.35E-04	9.27	133.99	8
brown	lysine catabolic process (GO:0006554)	6/12	0.013	8.69	92.23	9
brown	lysine metabolic process (GO:0006553)	6/12	0.014	8.69	92.23	10
cyan	platelet degranulation (GO:0002576)	13/124	1.37E-08	15.42	410.81	1
cyan	regulated exocytosis (GO:0045055)	13/148	6.62E-08	12.92	314.85	2
cyan	negative regulation of blood coagulation (GO:0030195)	8/36	1.63E-07	32.68	753.91	3
cyan	fibrinolysis (GO:0042730)	6/16	8.55E-07	55.15	1164.92	4
cyan	negative regulation of peptidase activity (GO:0010466)	8/62	9.41E-06	18.98	351.08	5
cyan	post-translational protein modification (GO:0043687)	15/357	1.58E-05	6.18	109.02	6
cyan	regulation of endopeptidase activity (GO:0052548)	8/68	1.66E-05	17.30	307.16	7
cyan	cellular protein metabolic process (GO:0044267)	16/484	1.25E-04	4.86	75.07	8
cyan	negative regulation of endopeptidase activity (GO:0010951)	7/83	7.93E-04	12.40	165.88	9
cyan	regulation of vasoconstriction (GO:0019229)	5/29	7.97E-04	25.35	341.65	10
green	regulation of alcohol biosynthetic process (GO:1902930)	20/34	1.02E-15	13.96	600.87	1
green	regulation of cholesterol biosynthetic process (GO:0045540)	20/40	3.98E-14	11.86	459.07	2
green	regulation of cholesterol metabolic process (GO:0090181)	20/41	4.98E-14	11.57	440.59	3
green	regulation of steroid biosynthetic process (GO:0050810)	20/44	2.17E-13	10.78	391.59	4
green	sterol biosynthetic process (GO:0016126)	19/40	2.95E-13	11.27	401.17	5
green	cholesterol biosynthetic process (GO:0006695)	18/35	3.48E-13	12.20	434.55	6
green	secondary alcohol biosynthetic process (GO:1902653)	18/36	4.77E-13	11.86	414.73	7
green	cholesterol metabolic process (GO:0008203)	18/68	1.64E-07	6.28	138.67	8
green	regulation of primary metabolic process (GO:0080090)	20/139	8.81E-04	3.41	45.66	9
green	acetyl-CoA metabolic process (GO:0006084)	6/14	0.006	10.17	114.90	10
magenta	type I interferon signaling pathway (GO:0060337)	18/65	5.28E-12	12.56	424.64	1

magenta	cellular response to type I interferon (GO:0071357)	18/65	1.06E-11	12.56	424.64	2
magenta	regulation of viral genome replication (GO:0045069)	13/63	1.60E-06	9.36	194.52	3
magenta	negative regulation of viral genome replication (GO:0045071)	11/15	1.16E-05	9.98	184.72	4
magenta	positive regulation of type I interferon production (GO:0032481)	11/62	8.31E-05	8.05	129.87	5
magenta	negative regulation of viral life cycle (GO:1903901)	11/61	8.37E-05	8.18	133.44	6
magenta	regulation of type I interferon production (GO:0032479)	12/85	2.55E-04	6.40	95.19	7
magenta	regulation of interferon-beta production (GO:0032648)	8/34	3.67E-04	10.67	153.32	8
magenta	positive regulation of interferon-beta production (GO:0032728)	7/26	6.27E-04	12.21	167.46	9
magenta	negative regulation of type I interferon production (GO:0032480)	8/44	0.002	8.25	101.28	10
red	mRNA processing (GO:0006397)	29/283	0.001	2.93	44.51	1
red	mRNA splicing, via spliceosome (GO:0000398)	27/261	0.001	2.96	42.68	2
red	RNA splicing (GO:0008380)	16/106	0.001	4.31	60.33	3
red	RNA processing (GO:0006396)	22/193	0.002	3.26	44.43	4
red	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile (GO:0000377)	24/236	0.003	2.91	36.88	5
red	transcription from RNA polymerase II promoter (GO:0006366)	35/485	0.030	2.06	20.56	6
red	regulation of lipid metabolic process (GO:0019216)	13/100	0.033	3.71	37.12	7
red	cellular response to DNA damage stimulus (GO:0006974)	27/329	0.034	2.34	23.75	8
red	spliceosomal complex assembly (GO:0000245)	8/41	0.042	5.57	53.01	9
red	positive regulation of gene expression (GO:0010628)	48/771	0.044	1.78	16.64	10
lightcyan	SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	38/89	2.79E-58	68.31	9636.66	1
lightcyan	cotranslational protein targeting to membrane (GO:0006613)	38/93	1.17E-57	65.38	9082.90	2
lightcyan	protein targeting to ER (GO:0045047)	38/97	5.85E-57	62.68	8582.17	3
lightcyan	viral gene expression (GO:0019080)	39/110	1.26E-56	56.73	7707.30	4
lightcyan	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (GO:0000184)	39/112	2.36E-56	55.71	7522.17	5
lightcyan	viral transcription (GO:0019083)	39/113	2.99E-56	55.22	7432.45	6
lightcyan	rRNA metabolic process (GO:0016072)	44/200	2.98E-54	35.20	4570.32	7
lightcyan	rRNA processing (GO:0006364)	44/202	4.24E-54	34.85	4508.17	8
lightcyan	ribosome biogenesis (GO:0042254)	44/226	8.52E-52	31.15	3860.55	9
lightcyan	ncRNA processing (GO:0034470)	44/227	9.47E-52	31.01	3836.98	10
midnightblue	positive regulation of vasculature development (GO:1904018)	10/104	3.00E-06	15.14	311.37	1
midnightblue	vasculogenesis (GO:0001570)	7/32	5.30E-06	34.45	712.57	2
midnightblue	positive regulation of angiogenesis (GO:0045766)	9/103	2.44E-05	13.76	244.54	3
midnightblue	regulation of angiogenesis (GO:0045765)	11/177	2.97E-05	9.79	174.82	4
midnightblue	sprouting angiogenesis (GO:0002040)	6/44	3.44E-04	21.47	320.03	5
midnightblue	cellular response to vascular endothelial growth factor stimulus (GO:0035924)	5/25	3.89E-04	31.50	459.75	6

midnightblue	cellular response to growth factor stimulus (GO:0071363)	8/139	2.17E-03	9.06	115.34	7
midnightblue	heart trabecula morphogenesis (GO:0061384)	4/18	2.51E-03	35.00	431.41	8
midnightblue	positive regulation of epithelial cell proliferation (GO:0050679)	7/107	2.52E-03	10.30	124.90	9
midnightblue	regulation of endothelial cell migration (GO:0010594)	6/69	2.57E-03	13.69	167.07	10
pink	cell communication by electrical coupling involved in cardiac conduction (GO:0086064)	6/14	7.65E-04	18.01	263.17	1
pink	cardiac muscle cell action potential involved in contraction (GO:0086002)	8/28	1.01E-03	12.00	185.26	2
pink	cellular response to hormone stimulus (GO:0032870)	12/80	1.02E-03	6.30	92.85	3
pink	membrane depolarization during cardiac muscle cell action potential (GO:0086012)	6/17	1.27E-03	14.83	196.63	4
pink	potassium ion homeostasis (GO:0055075)	7/25	1.40E-03	11.76	158.81	5
pink	glucocorticoid biosynthetic process (GO:0006704)	5/10	1.45E-03	21.01	279.00	6
pink	regulation of heart rate by cardiac conduction (GO:0086091)	8/35	1.66E-03	9.60	130.17	7
pink	mineralocorticoid biosynthetic process (GO:0006705)	4/6	2.59E-03	28.01	344.39	8
pink	cardiac muscle cell action potential (GO:0086001)	7/29	2.62E-03	10.14	125.80	9
pink	regulation of secretion by cell (GO:1903530)	9/56	3.25E-03	6.75	80.78	10
purple	muscle contraction (GO:0006936)	50/137	4.44E-56	28.18	3832.59	1
purple	muscle filament sliding (GO:0030049)	25/38	1.56E-35	50.80	4449.64	2
purple	actin-myosin filament sliding (GO:0033275)	25/38	2.34E-35	50.80	4449.64	3
purple	striated muscle contraction (GO:0006941)	25/61	1.46E-28	31.65	2254.66	4
purple	myofibril assembly (GO:0030239)	21/47	9.51E-25	34.50	2147.47	5
purple	actomyosin structure organization (GO:0031032)	21/71	2.63E-20	22.84	1183.83	6
purple	heart contraction (GO:0060047)	17/44	1.75E-18	29.84	1416.57	7
purple	sarcomere organization (GO:0045214)	14/31	3.64E-16	34.87	1464.93	8
purple	cardiac muscle contraction (GO:0060048)	14/36	4.38E-15	30.03	1183.27	9
purple	muscle fiber development (GO:0048747)	12/27	1.29E-13	34.32	1232.67	10
turquoise	neutrophil activation involved in immune response (GO:0002283)	187/483	3.12E-18	1.93	92.75	1
turquoise	neutrophil degranulation (GO:0043312)	186/479	5.45E-18	1.93	93.29	2
turquoise	neutrophil mediated immunity (GO:0002446)	187/487	6.12E-18	1.91	89.93	3
turquoise	respiratory electron transport chain (GO:0022904)	59/94	2.65E-16	3.12	134.33	4
turquoise	mitochondrial ATP synthesis coupled electron transport (GO:0042775)	53/85	2.19E-14	3.10	119.05	5
turquoise	mitochondrial electron transport, NADH to ubiquinone (GO:0006120)	34/46	4.54E-12	3.68	120.85	6
turquoise	mitochondrial translational elongation (GO:0070125)	49/87	7.28E-11	2.80	83.88	7
turquoise	mitochondrial translation (GO:0032543)	56/107	7.75E-11	2.60	77.43	8
turquoise	mitochondrial respiratory chain complex assembly (GO:0033108)	51/97	6.71E-10	2.62	71.83	9
turquoise	mitochondrial translational termination (GO:0070126)	48/89	8.05E-10	2.68	72.92	10

[a]Combined score is described as c = log(p) * z , where c = the combined score, p = Fisher exact test p-value, and z = z-score for deviation from expected rank.

Table 3.10. Top 10 KEGG pathway results for 6 hepatic coexpression gene modules in liver in eight CC founder strains. n = 48 (6 mice per eight strains and 24 mice per two diets)

Module[a]	Term	Overlap (gene count)	Adjusted P-value	Odds Ratio	Combined Score[b]	Rank
brown	Valine, leucine and isoleucine degradation	21/56	4.51E-10	6.52	177.46	1
brown	Peroxisome	21/84	1.17E-06	4.34	81.16	2
brown	Fatty acid degradation	15/50	7.76E-06	5.21	85.39	3
brown	Drug metabolism	22/114	3.68E-05	3.35	48.75	4
brown	Lysine degradation	15/59	5.06E-05	4.42	61.82	5
brown	Tryptophan metabolism	13/48	1.05E-04	4.71	61.59	6
brown	Glyoxylate and dicarboxylate metabolism	10/31	2.42E-04	5.61	67.79	7
brown	Glycine, serine and threonine metabolism	11/40	4.08E-04	4.78	54.66	8
brown	Butanoate metabolism	9/27	4.12E-04	5.79	65.51	9
brown	Propanoate metabolism	8/31	0.009	4.48	36.65	10
cyan	Complement and coagulation cascades	13/88	8.73E-12	21.72	677.33	1
green	Steroid biosynthesis	11/19	1.18E-08	13.74	329.28	1
green	Terpenoid backbone biosynthesis	9/23	2.92E-05	9.28	143.53	2
green	Peroxisome	13/84	0.005	3.67	36.56	3
green	Glyoxylate and dicarboxylate metabolism	7/31	0.019	5.36	44.44	4
magenta	DNA replication	10/35	8.34E-07	12.96	255.41	1
magenta	Herpes simplex virus 1 infection	27/433	2.11E-04	2.83	38.13	2
magenta	Epstein-Barr virus infection	17/229	0.001	3.37	37.61	3
magenta	Hepatitis C	14/160	0.001	3.97	44.68	4
magenta	Cell cycle	11/123	0.005	4.06	37.83	5
magenta	Cellular senescence	13/185	0.012	3.19	26.49	6
magenta	Cytosolic DNA-sensing pathway	7/61	0.014	5.20	41.02	7
magenta	Influenza A	12/168	0.016	3.24	25.66	8
magenta	RIG-I-like receptor signaling pathway	7/68	0.025	4.67	33.69	9
magenta	Antigen processing and presentation	8/90	0.025	4.03	28.56	10
red	Spliceosome	14/132	0.022	3.03	25.51	1
red	FoxO signaling pathway	14/132	0.033	3.03	25.51	2
red	Non-small cell lung cancer	9/66	0.035	3.90	29.89	3
red	Choline metabolism in cancer	11/99	0.042	3.17	23.12	4
red	Glycerophospholipid metabolism	12/97	0.044	3.53	31.26	5
lightcyan	Ribosome	40/170	1.52E-50	37.65	4533.62	1
midnightblue	Cell adhesion molecules (CAMs)	7/170	0.016	6.48	59.25	1
midnightblue	Leukocyte transendothelial migration	6/115	0.028	8.22	76.26	2
midnightblue	Rap1 signaling pathway	7/209	0.039	5.27	41.52	3
pink	Cortisol synthesis and secretion	20/69	2.73E-14	12.18	449.94	1
pink	Aldosterone synthesis and secretion	21/102	4.33E-12	8.65	269.77	2
pink	Cushing syndrome	21/159	2.12E-08	5.55	123.66	3

pink	Insulin secretion	13/86	9.05E-06	6.35	101.24	4
pink	cAMP signaling pathway	19/211	4.65E-05	3.78	53.27	5
pink	Adrenergic signaling in cardiomyocytes	15/148	1.31E-04	4.26	54.77	6
pink	Circadian entrainment	12/99	1.77E-04	5.09	63.20	7
pink	Renin secretion	10/76	4.70E-04	5.53	62.45	8
pink	Dopaminergic synapse	13/135	6.33E-04	4.05	43.60	9
pink	Oxytocin signaling pathway	14/154	6.57E-04	3.82	41.42	10
purple	Dilated cardiomyopathy (DCM)	14/90	2.99E-09	12.01	304.39	1
purple	Hypertrophic cardiomyopathy (HCM)	13/86	1.21E-08	11.67	271.37	2
purple	Adrenergic signaling in cardiomyocytes	13/148	6.96E-06	6.78	111.85	3
purple	Cardiac muscle contraction	9/78	5.65E-05	8.91	125.71	4
purple	Calcium signaling pathway	13/189	7.08E-05	5.31	72.56	5
purple	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	8/72	2.10E-04	8.58	106.32	6
purple	Oxytocin signaling pathway	11/154	2.36E-04	5.52	66.84	7
purple	Apelin signaling pathway	10/138	4.89E-04	5.60	63.00	8
purple	Vascular smooth muscle contraction	9/140	2.71E-03	4.96	46.28	9
purple	cGMP-PKG signaling pathway	10/172	2.86E-03	4.49	42.08	10
turquoise	Non-alcoholic fatty liver disease (NAFLD)	81/151	1.75E-17	2.67	118.22	1
turquoise	Huntington disease	90/192	7.69E-15	2.33	87.50	2
turquoise	Oxidative phosphorylation	70/134	1.24E-14	2.60	95.21	3
turquoise	Alzheimer disease	82/175	9.60E-14	2.33	79.96	4
turquoise	Fc gamma R-mediated phagocytosis	48/87	3.08E-11	2.74	77.70	5
turquoise	Parkinson disease	67/144	3.80E-11	2.31	64.62	6
turquoise	Thermogenesis	93/231	6.09E-11	2.00	54.66	7
turquoise	Regulation of actin cytoskeleton	87/217	3.64E-10	1.99	50.60	8
turquoise	Osteoclast differentiation	59/128	9.42E-10	2.29	55.73	9
turquoise	Focal adhesion	80/199	1.67E-09	2.00	47.24	10

[a]There are fewer than 10 KEGG pathways identified in cyan, green and red modules.
[b]Combined score is described as c = log(p) * z , where c = the combined score, p = Fisher exact test p-value, and z = z-score for deviation from expected rank.

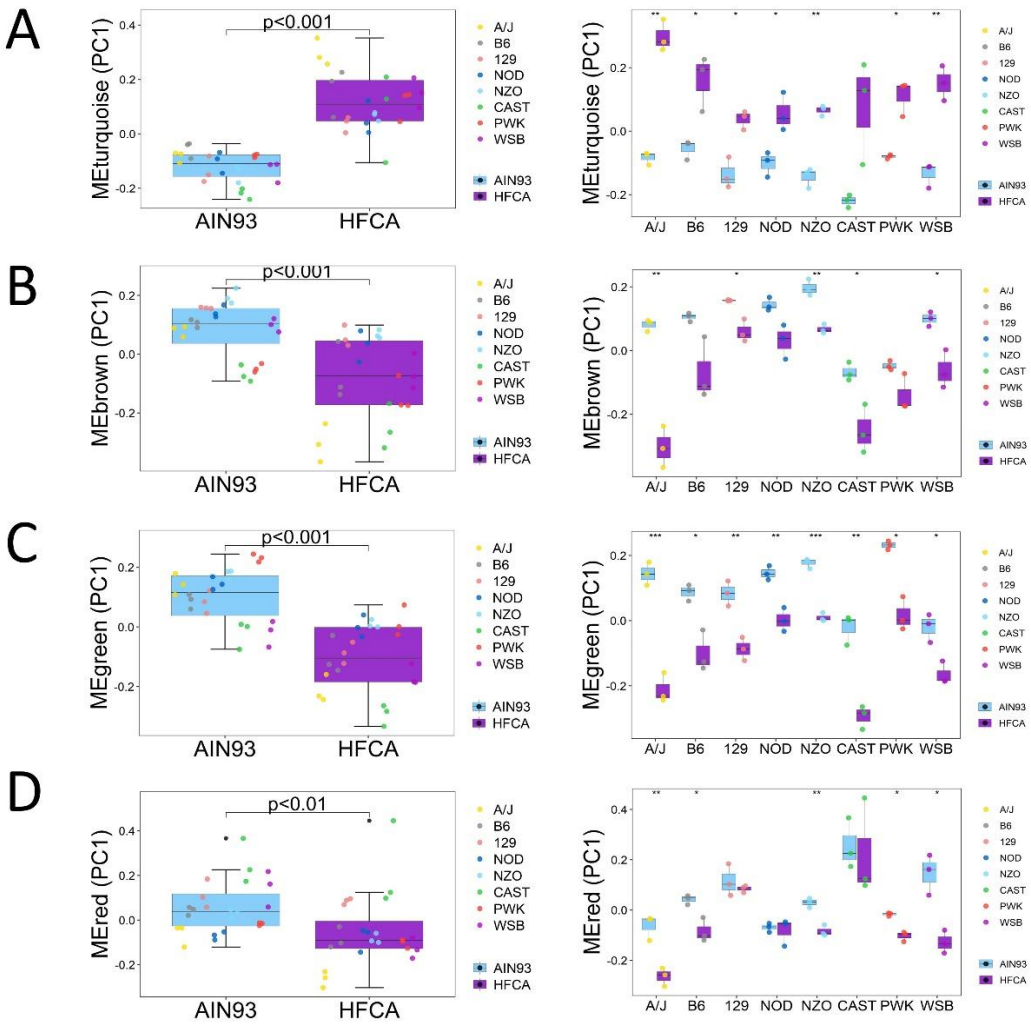
3.5.4. Effect of Diet and Strain on Hepatic Transcriptional Network of CC Progenitors

We next sought to delineate the effects of diet on these modules of highly connected genes. The modules can be characterized by their module eigengenes (MEs) which reflect the first principal component of the module. Non-parametric statistical tests of the effect of diet on ME levels were performed to identify diet-influenced modules**.** Among 20 co-expressed gene modules detected in WGCNA, 11 modules showed significant differences in MEs between diets (**Figures 3.7A-F and Table 3.11)**. Next, in order to investigate the connection between diet-specific DEGs and gene networks, we explored the proportion of "Core Diet DEGs" in each module (**Figure 3.8**). Of these 6,411 genes, 5,473 (85.3%) were contained in one of 5 modules within the network. Only 32 out of the 6,411 Core Diet DEGs were not associated with any module. The top 4 modules with the highest number of genes upregulated by the HFCA diet were turquoise (52.2%; 2,097 out of 4,020 genes), brown (20%; 230 out of 1151 genes), green (23.0%; 194 out of 843 genes), and magenta (29.7%; 131 out of 441 genes). The top 4 modules with the highest number of genes upregulated by the AIN-93M diet were turquoise (40.1%; 1,612 out of 4,020 genes), green (56.5%; 476 out of 843 genes), brown (41%; 475 out of 1151 genes), and red (22%; 154 out of 700 genes) (**Figure 3.7G and Tables 3.12 and 3.13**).

We next sought to identify strain-specific responses within the modules by examining both strain and strain by diet effects (**Table 3.11)**. Among the 20 modules, 18 were moduled by strain. One of the 2 modules without significant effects of strain was the turquoise module which contains the most diet-specific DEGs (P=0.48; **Figure 3.7A and Table 3.11**). As shown in **Table 3.5**, the number of diet-specific DEGs identified in the A/J, NZO, and WSB strains was the highest, and the number of diet-specific DEGs identified in the CAST strain was the lowest across all modules (**Tables 3.12 and 3.13**). Specifically, we highlighted 6 modules in which significant diet effects

were identified in at least 5 strains which may indicate a critical strain by diet interaction (**Figure. 3.7A-F**). We next described the brown and green modules with both strain and strain by diet effects as examples.

The brown module, which contains 1,151 transcripts, is highly enriched for genes involved in fatty acid catabolism (**Table 3.8-3.10**). A previous report identified that hepatic fatty acid oxidation-related genes are down-regulated by an atherogenic diet challenge in mice (Matsuzawa et al., 2007) and reflective of this, the brown module ME was higher in AIN-93M diet-fed mice than in HFCA diet-fed mice independent of strain ($P < 0.001$; **Figure 3.7B and Table 3.9**), and this trend was also observed in five individual strain analyses ($P < 0.01$ in A/J strain, $P < 0.05$ in 129 strain, $P < 0.01$ in NZO strain, $P < 0.05$ in CAST strain, and $P < 0.05$ in WSB strain). Furthermore, the number of genes upregulated by the AIN-93M diet (475 genes) in the brown module was 2.1 times higher than the number of genes upregulated by the HFCA diet (230 genes) (**Figure 3.7G and Tables 3.12 and 3.13**). In terms of strain-dependent effects, transcripts in this module were up-regulated in NZO mice and down-regulated in CAST mice, which may reflect differences in fatty acid catabolic process in the liver from these strains (**Figures 3.7B and Table 3.11)**. Genes involved in fatty acid oxidation have increased expression in patients with fatty liver, which is considered to reduce reactive oxygen species produced by fatty acid oxidation (Kohjima et al., 2007). Thus, the NZO strain has the highest expression of fatty acid oxidation-related genes and has the highest adiposity and liver TG content, while the CAST strain has the lowest expression of the genes and is generally resistant to MetSyn.


A
p<0.001
MEturquoise (PC1)
AIN93
HFCA
A/J
B6
129
NOD
NZO
CAST
PWK
WSB
B
p<0.001
MEbrown (PC1)
C
p<0.001
MEgreen (PC1)
D
p<0.01
MEred (PC1)

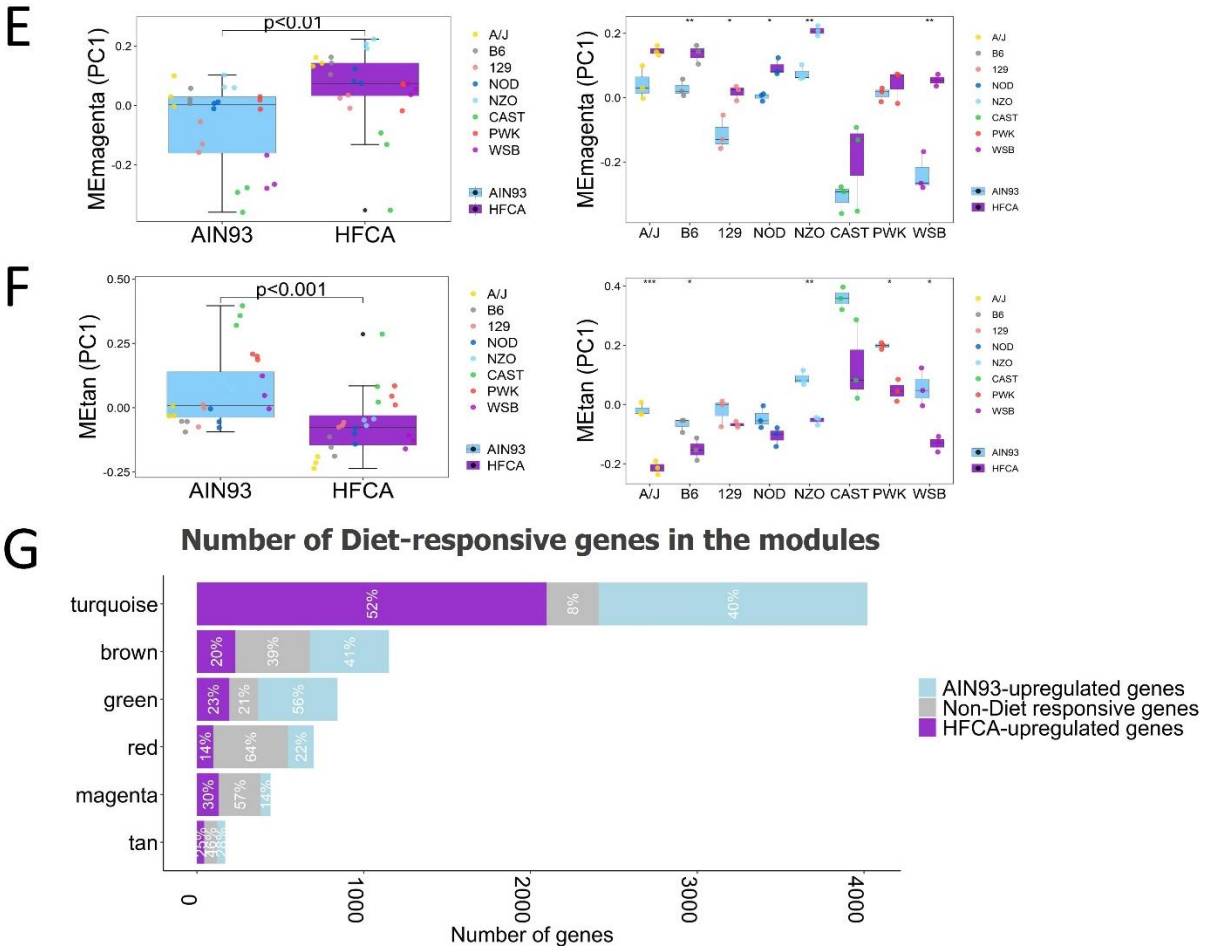


Figure 3.7. Effect of diet and genetic background on hepatic co-expression gene modules in the eight CC founder strains in female mice.
Differences of PC1 of module eigengenes (ME) including turquoise (A), brown (B), green (C), red (D), magenta (E), and tan (F) module ME by diet and genetic background were plotted. (G) Number and proportion of diet or non-diet responsive genes in six modules. The x-axis is the number of genes and the y-axis is the color of each module. Length of the bar corresponding to each module is the number of module genes, and portions corresponding to purple, grey, and lightblue colors in each bar are the number of genes upregulated by the HFCA diet (purple), non-diet responsive genes (grey), and genes upregulated by the AIN-93M diet (lightblue). Proportion of each color is written in white letters in the bar. "***"$P < 0.001$, "**"$P < 0.01$, "*"$P < 0.05$. N = 6 mice for each founder, 3 for AIN-93M diet (lightblue color) and 3 for HFCA diet (purple color)-fed mice. A/J (yellow), B6 (gray), C57BL/6J; 129 (pink), 129S1/SvlmJ; NOD (blue), NOD/ShiLtJ; NZO (lightblue), NZO/HILtJ; CAST (green), CAST/EiJ; PWJ (red), PWK/PhJ; WSB (purple), WSB/EiJ.

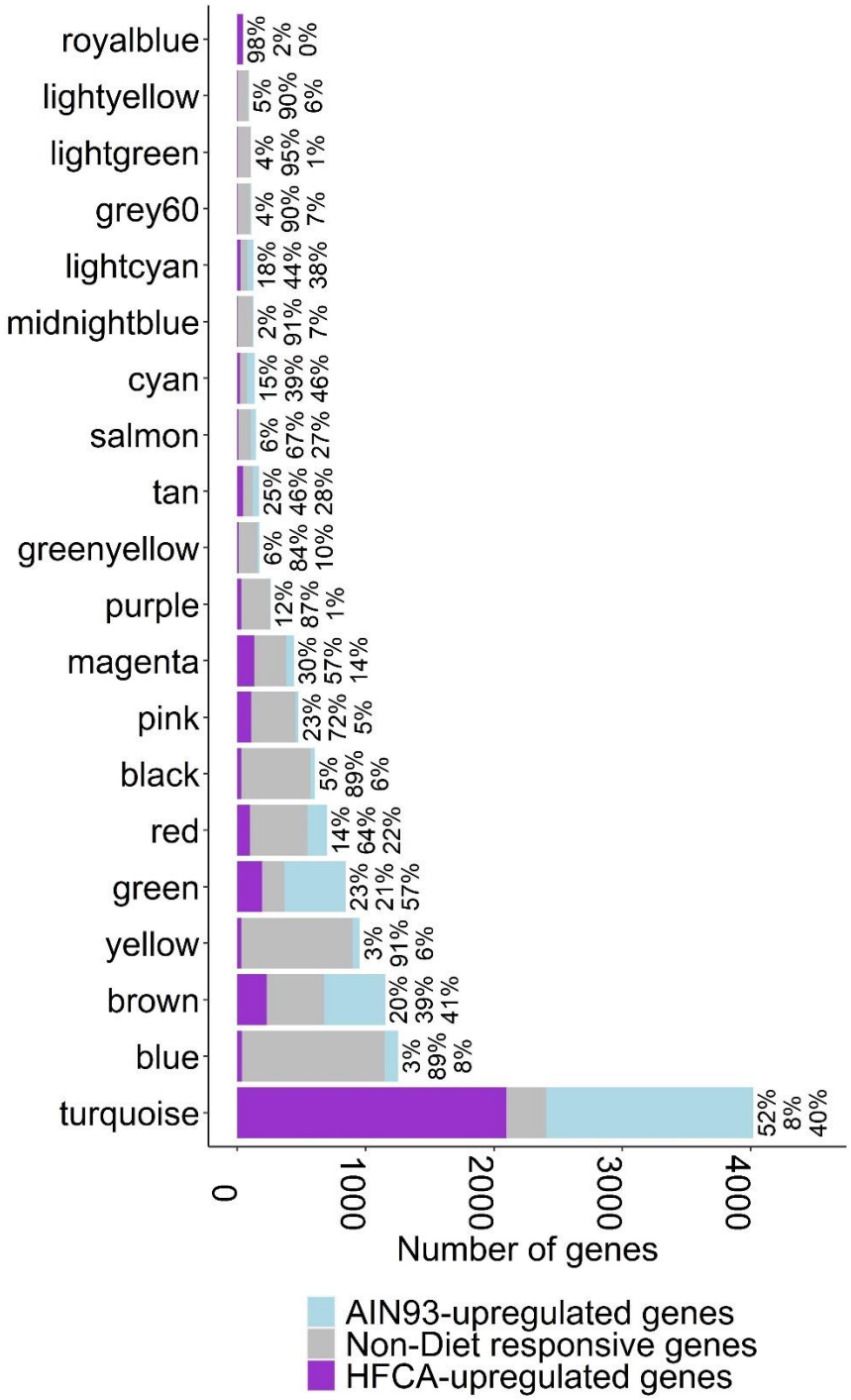


Figure 3.8. Number and proportion of diet or non-diet responsive genes in 20 modules. The x-axis is the number of genes and the y-axis is the color of each module. Length of the bar corresponding to each module is the number of module genes, and portions corresponding to purple, grey, and lightblue colors in each bar are the number of genes upregulated by the HFCA diet (purple), non-diet responsive genes (grey), and genes upregulated by the AIN-93M diet (lightblue). Proportion of each color is written in black on the right side of the bar.

Table 3.11. Diet-, strain-, and diet-by strain-dependent differences in module eigengenes for hepatic gene networks in eight CC founder strains. n = 48 (6 mice per eight strains and 24 mice per two diets)

Module	Diet effect	Strain effect			Diet by Strain effect		
	p-value for comparison in diets (Wilcoxon test)	p-value for comparison in strains (Kruskal-Wallis test)	Strain that has the highest value	Strain that has the lowest value	p-value for two-way ANOVA with interaction effect (Diet)	p-value for two-way ANOVA with interaction effect (Strain)	p-value for two-way ANOVA with interaction effect (Diet:Strain)
blue	NS	2.20E-05	NOD	CAST	NS	2.00E-16	NS
magenta	P < 0.01	2.30E-04	NZO	CAST	3.77E-10	4.57E-14	7.02E-03
grey60	NS	7.40E-07	NZO	WSB	9.82E-07	2.00E-16	5.02E-05
yellow	NS	1.30E-05	NZO	CAST	6.64E-04	2.00E-16	0.01
salmon	P < 0.05	1.90E-04	A/J	B6	1.38E-09	2.00E-16	NS
green	P < 0.001	2.50E-07	PWK	CAST	2.00E-16	5.46E-14	2.27E-04
brown	P < 0.001	1.40E-03	NZO	CAST	7.12E-14	5.06E-13	1.18E-04
cyan	P < 0.001	5.40E-03	WSB	PWK	4.86E-09	8.10E-08	NS
lightgreen	NS	2.20E-07	NZO	B6	5.00E-03	2.00E-16	NS
lightcyan	P < 0.001	1.50E-03	CAST	WSB	5.00E-03	2.00E-16	NS
red	P < 0.01	2.70E-04	CAST	A/J	2.69E-06	4.54E-10	0.02
tan	P < 0.001	2.10E-04	CAST	A/J	1.91E-11	2.10E-14	0.01
black	NS	1.70E-04	PWK	129	1.87E-02	2.00E-16	8.55E-06
turquoise	P < 0.001	NS	A/J	CAST	3.45E-15	1.88E-04	0.04
pink	P < 0.05	8.60E-03	PWK	NZO	0.03	NS	NS
purple	NS	0.033	CAST	NOD	NS	NS	NS
royalblue	P < 0.001	NS	A/J	NZO	7.28E-06	NS	NS
lightyellow	NS	5.90E-07	NOD	B6	5.73E-09	2.00E-16	1.66E-03
greenyellow	NS	5.00E-03	WSB	NZO	1.41E-08	2.00E-16	3.47E-03
midnightblue	NS	1.60E-03	B6	NZO	NS	5.32E-08	1.16E-05

Table 3.12. Number and proportion of HFCA diet-upregulated genes in modules.

Modulea	#Genes	#HFCA-upregulated genes									
		All mice (%)[b]	At least one strain	AJ	B6	129	NOD	NZO	CAST	PWK	WSB
turquoise	4020	2097 (52.2)	2122	1796	817	974	764	1418	45	267	1512
blue	1252	35 (2.8)	274	63	31	31	38	109	2	18	79
brown	1151	230 (20)	319	236	57	59	42	136	3	15	122
yellow	955	30 (3.1)	246	74	36	31	38	97	2	20	94
green	843	194 (23)	212	151	42	53	50	99	1	24	67
red	700	97 (13.9)	220	94	29	28	31	76	1	13	136
black	606	30 (5)	159	63	19	18	24	62	1	13	55
pink	476	108 (22.7)	158	96	38	14	6	59	1	2	41
magenta	441	131 (29.7)	222	77	47	67	68	115	3	9	173
purple	259	30 (11.6)	65	33	12	5	5	26	0	3	21
greenyellow	173	10 (5.8)	53	19	9	11	7	17	0	6	14
tan	169	43 (25.4)	53	36	8	7	11	32	0	6	26
salmon	144	8 (5.6)	27	4	5	4	6	10	0	2	11
cyan	136	21 (15.4)	26	18	9	6	3	13	0	2	17
midnightblue	127	3 (2.4)	53	1	0	0	1	51	0	0	3
lightcyan	125	23 (18.4)	29	12	9	4	3	21	1	1	8
grey60	108	4 (3.7)	28	14	4	5	5	10	0	1	8
lightgreen	107	4 (3.7)	26	9	6	5	5	17	0	0	12
lightyellow	89	4 (4.5)	36	3	6	6	10	16	1	1	20
royalblue	44	43 (97.7)	43	40	14	9	1	12	0	2	22

[a]Each module was listed in the order of the number of module genes.
[b]Percentage is the number of genes upregulated by the HFCA diet in the module divided by the number of module genes.

Table 3.13. Number and proportion of AIN-93M diet-upregulated genes in modules.

Modulea	#Genes	#HFCA-upregulated genes									
		All mice (%)[b]	At least one strain	AJ	B6	129	NOD	NZO	CAST	PWK	WSB
turquoise	4020	1612 (40.1)	1559	1211	390	262	248	905	8	156	868
blue	1252	103 (8.2)	378	170	57	52	65	120	9	34	93
brown	1151	475 (41.3)	638	539	129	136	128	300	11	37	240
yellow	955	56 (5.9)	302	164	33	38	47	96	3	24	70
green	843	476 (56.5)	477	382	148	154	105	246	9	130	166
red	700	154 (22)	332	183	54	17	16	154	0	18	213
black	606	36 (5.9)	188	92	18	24	23	76	0	33	54
pink	476	24 (5)	85	35	9	8	18	27	0	3	32
magenta	441	60 (13.6)	124	45	18	7	7	49	2	7	77
purple	259	3 (1.2)	42	15	7	10	5	15	0	1	8
greenyellow	173	17 (9.8)	52	33	9	5	12	14	0	3	12
tan	169	48 (28.4)	65	40	7	2	7	39	0	10	31
salmon	144	39 (27.1)	60	33	12	8	7	44	0	7	18
cyan	136	62 (45.6)	78	53	12	17	21	44	0	10	40
midnightblue	127	9 (7.1)	85	78	20	0	2	1	2	2	16
lightcyan	125	47 (37.6)	37	17	6	3	2	23	0	0	8
grey60	108	7 (6.5)	40	28	11	6	5	19	0	1	7
lightgreen	107	1 (0.9)	35	17	4	4	1	19	3	1	10
lightyellow	89	5 (5.6)	39	23	4	7	9	13	1	4	9
royalblue	44	0 (0)	0	0	0	0	0	0	0	0	0

[a]Each module was listed in the order of the number of module genes.
[b]Percentage is the number of genes upregulated by the HFCA diet in the module divided by the number of module genes.

As an additional example of strain and diet effects, we highlight the green module which contains 843 transcripts. A number of these are important for lipid processing such as low-density lipoprotein receptor (*Ldlr*), proprotein convertase subtilisin/kexin type 9 (*Pcsk9*), and many other steroid synthesis enzymes (**Tables 3.8-3.10**). The ME in the green module was higher in AIN-93M diet-fed mice than in HFCA diet-fed mice independent of strains ($P < 0.001$) and in all individual strains (**Figure 3.7C and Table 3.11)**. In addition, the number of genes upregulated by the AIN-93M diet (476 genes) in the green module was 2.5 times higher than the number of genes upregulated by the HFCA diet (194 genes) (**Figure 3.7G and Table 3.12 and 3.13**). A high-fat or an atherogenic diet is known to reduce the rate of synthesis and conversion of primary bile salts in humans and B6 mice (Bisschop et al., 2004; Renaud et al., 2014). In our study, HFCA-fed mice exhibited down-regulation of the cholesterol/steroid synthesis pathway potentially due to bile acid accumulation in the liver. In terms of strain-dependent effects, transcripts in the green module were down-regulated in the CAST strain and up-regulated in the PWK strain (**Figure 3.7C and Table 3.11)**. The CAST mice showed the highest plasma cholesterol concentration among all strains which may reflect a decreased expression of genes related to the cholesterol metabolism, including *Ldlr* which removes cholesterol from plasma LDL into the liver.

In addition to strain and diet effects, we also sought to identify a number of strain by diet interactions of the co-expression modules. Significant diet-by-strain interactions, assessed by a two-way ANOVA test, were observed in 12 modules. These include the above 6 modules, and the grey60, yellow, black, lightyellow, greenyellow, and midnightblue modules (**Table 3.11**).

For example, the black and midnightblue modules showed opposite diet effects by genetic background. The ME of the midnightblue module, which showed KEGG pathway for cell adhesion molecules, was significantly higher in the AIN-93M diet-fed mice than in the HFCA diet-fed mice

in the AJ and B6 strains, yet in the NZO and CAST strains the ME of the midnightblue module was higher in HFCA fed mice than in AIN-93 M fed mice (**Table 3.11**). These results show that genetic background has a significant influence on hepatic gene networks in response to a control diet or an atherogenic diet.

3.5.5. Liver Transcriptome Co-expression Modules Correlate with Metabolic Traits

Before assessing the relationship between gene modules and clinical traits, we determined the effect of diet on the 29 metabolic traits measured. Liver weight ($P < 0.05$), liver TG ($P < 0.05$), plasma ALT ($P < 0.01$), AST ($P < 0.01$), total cholesterol ($P < 0.001$), betaine ($P < 0.05$), and oxygen consumption per weight ($P < 0.05$) showed significant diet-dependent differences, assessed by the Wilcoxon test, regardless of genetic backgrounds during the 16-week HFCA diet challenge period (**Figure 3.9 and Table 3.14**). Plasma TMAO concentration also increased 1.35-fold HFCA diet-fed mice. This difference was not significant, but we noticed broad strain differences and speculated that the relationship between plasma TMAO and diet depends on the genetic background. We thus tested the clinical traits for strain-by-diet interactions and identified significant diet-by-strain interaction effects for food intake, plasma TG, glucose, and TMAO concentrations, as well as glucose/insulin ratio and oxygen consumption (**Table 3.14**). Furthermore, food intake ($P = 0.026$), plasma TMAO level ($P = 0.016$), and oxygen consumption ($P = 0.019$) were significantly different between the two diets only when considering the diet by strain interaction effect by two-way ANOVA (**Table 3.14**).

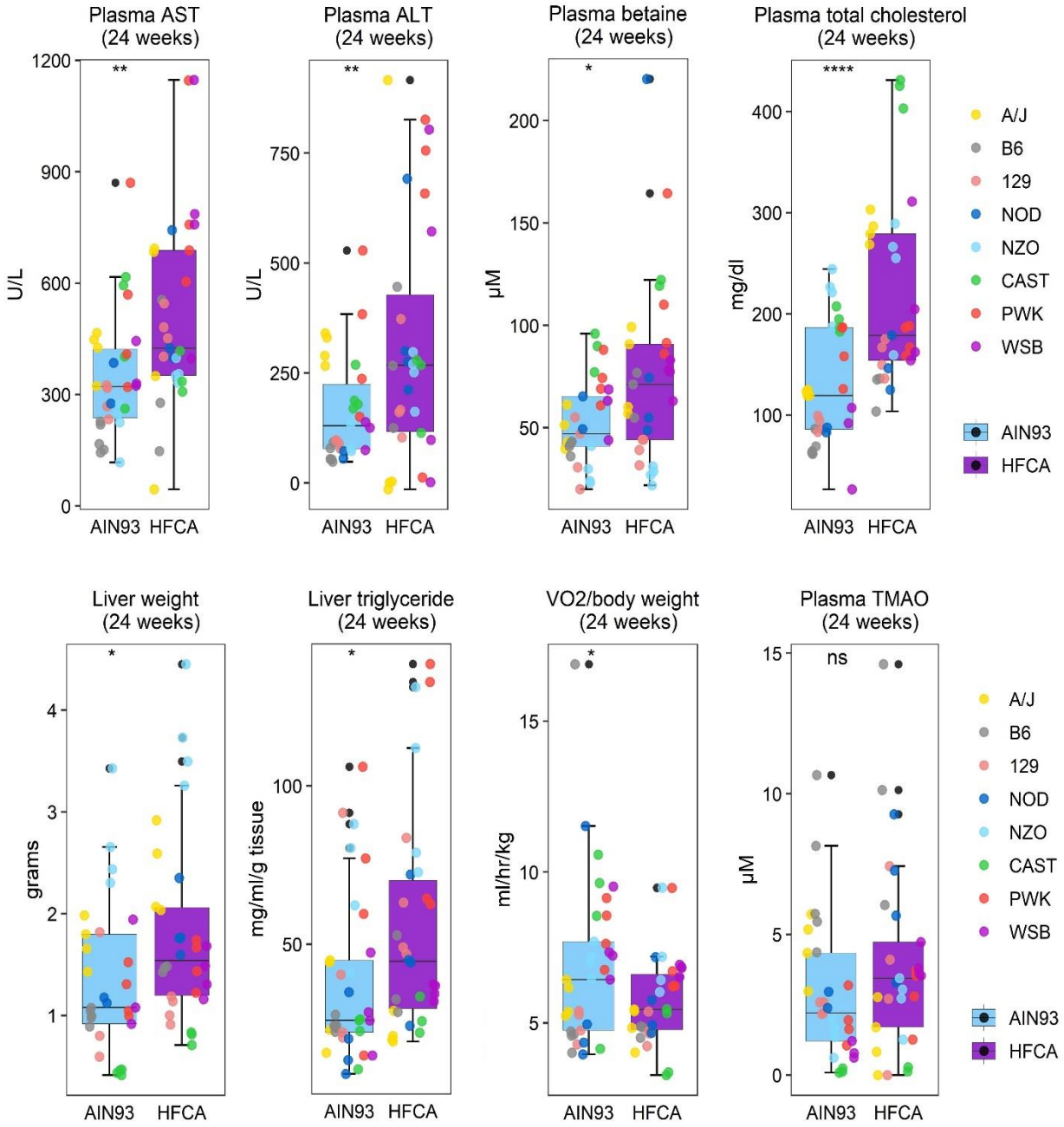


Figure 3.9. Diet-dependent differences in key metabolic traits in eight CC founder strains. Liver weight and triglyceride (TG) and plasma AST, ALT, total cholesterol, and betaine were higher in HFCA diet-fed mice (purple color), while oxygen consumption per body weight was higher in AIN-93M diet-fed mice (lightblue color). Plasma TMAO was higher in HFCA diet-fed mice, but not significant. A/J (yellow), B6 (gray), C57BL/6J; 129 (pink), 129S1/SvlmJ; NOD (blue), NOD/ShiLtJ; NZO (lightblue), NZO/HILtJ; CAST (green), CAST/EiJ; PWJ (red), PWK/PhJ; WSB (purple), WSB/EiJ. "****"$P < 0.001$, "**"$P < 0.01$, "*"$P < 0.05$. Data were mean ± S.E., $n \geq 4$ mice/diet/strain

Table 3.14. Diet-, strain-, and diet-by strain-dependent differences in cardio-metabolic traits in eight CC founder strains. n = 64 (33 mice for AIN-93M diet and 31 mice for HFCA diet).

Trait	Unit	Diet effect	Strain effect			Diet by Strain effect		
		p-value for comparison in diets (Wilcoxon test)	p-value for comparison in strains (Kruskal-Wallis test)	Strain that has the highest value	Strain that has the lowest value	p-value for two-way ANOVA with interaction effect (Diet)	p-value for two-way ANOVA with interaction effect (Strain)	p-value for two-way ANOVA with interaction effect (Diet:Strain)
Body weight	grams	NS	9.80E-08	NZO	CAST	NS	2.00E-16	NS
Fat mass	grams	NS	3.40E-08	NZO	CAST	NS	2.00E-16	NS
% Fat mass	%	NS	7.90E-09	NZO	CAST	NS	2.00E-16	NS
Lean mass	grams	NS	5.40E-08	NZO	CAST	NS	2.00E-16	NS
% Lean mass	%	NS	7.50E-09	CAST	NZO	NS	2.00E-16	NS
Liver weight	grams	P < 0.05	2.50E-07	NZO	CAST	3.90E-06	2.00E-16	NS
Spleen weight	grams	NS	1.40E-07	NZO	CAST	0.03	7.33E-12	NS
Gonadal fat weight	grams	NS	6.90E-08	NZO	CAST	NS	2.00E-16	NS
Heart weight	grams	NS	1.50E-04	NZO	CAST	4.10E-03	2.00E-16	NS
Food intake (g)	grams	NS	7.00E-08	NZO	A/J	0.03	2.00E-16	9.97E-03
Food intake (kcal)	kcal	NS	7.70E-08	NZO	A/J	0.02	2.00E-16	7.71E-04
Plasma ALT	U/L	P < 0.01	NS	PWK	129	1.07E-02	NS	NS
Plasma AST	U/L	P < 0.01	1.70E-03	PWK	B6	1.28E-03	1.04E-04	NS
Plasma total cholesterol	mg/dL	P < 0.001	5.80E-05	CAST	B6	9.52E-15	1.12E-15	5.39E-08
Plasma triglyceride	mg/dL	NS	NS	CAST	B6	NS	7.74E-03	7.26E-06
Liver triglyceride	mg/ml/g tissue	P < 0.05	1.90E-04	NZO	CAST	4.50E-03	5.55E-08	NS
Plasma glucose	mg/dL	NS	3.50E-06	CAST	129	NS	1.84E-09	2.73E-05
Plasma insulin	µlU/mL	NS	1.90E-04	NZO	NOD	NS	1.05E-05	NS
HOMA-IR	NA	NS	1.10E-04	NZO	A/J	NS	1.23E-05	NS
Glucose/Ins ratio	NA	NS	5.90E-05	CAST	NZO	NS	5.60E-10	8.84E-08
Plasma urea/BUN	mg/dL	NS	9.90E-05	B6	A/J	NS	3.04E-04	NS
Plasma creatinine	µM	NS	5.00E-06	B6	PWK	NS	1.53E-09	NS
Plasma TMAO	µM	NS	1.00E-04	B6	CAST	0.02	9.58E-09	0.01

Plasma choline	µM	NS	NS	**NOD**	**NZO**	NS	NS	NS
Plasma betaine	µM	P < 0.05	1.60E-06	**CAST**	**NZO**	2.72E-03	1.36E-05	NS
VO2	ml/hr	NS	1.00E-05	**CAST**	**129**	0.02	5.25E-07	0.02
VO2/body weight	ml/hr/kg	P < 0.05	5.60E-04	**PWK**	**129**	0.04	NS	NS
VO2/lean mass	ml/hr/kg	P < 0.05	1.10E-03	**CAST**	**129**	1.44E-02	0.04	NS
RER	NA	NS	0.046	**NZO**	**WSB**	NS	0.01	NS

These data reveal that genetic background has a profound effect on metabolic traits in response to a control diet or an atherogenic diet. To assess the physiological significance of the gene networks in liver tissue, we investigated whether the modules were correlated with metabolic traits (**Figure 3.10**). For example, transcripts in the red module were enriched for the GO terms "mRNA processing (GO: 0006397) (GO term: -logP > 3) and “FoxO signaling pathway” (KEGG pathway: -logP > 2) and included many NADH:ubiquinone oxidoreductase and forkhead box O-related proteins (**Tables 3.8-3.10**). The ME in the red module was higher in AIN-93M diet-fed mice than in HFCA diet-fed mice independent of strain ($P < 0.01$; **Figures 3.7D and Table 3.11)**, and this trend was also observed in five individual strain analyses ($P < 0.01$ in A/J strain, $P < 0.05$ in B6 strain, $P < 0.01$ in NZO strain, $P < 0.05$ in PWK strain, and $P < 0.05$ in WSB strain). The number of genes upregulated by the AIN-93M diet (154 genes) in the red module was 1.6 times higher than the number of genes upregulated by the HFCA diet (97 genes) (**Figure 3.7G and Tables 3.12 and 3.13**). In terms of strain-dependent effect, the ME in the red module was up-regulated in the CAST strain and down-regulated in the A/J strain (**Figure. 3.7D and Table 3.11)**. With regard to hepatic lipid metabolism, the FoxO signaling pathway has been shown to inhibit lipogenesis (Cook et al., 2015) and promote lipolysis (Zhang et al., 2016). Therefore, our results suggest that HFCA diet-fed mice with increased hepatic lipogenesis have relatively down-regulated FoxO signaling pathway, and CAST mice with the lowest liver TG have activated FoxO signaling pathway, resulting in decreased hepatic lipogenesis and increased lipolysis.

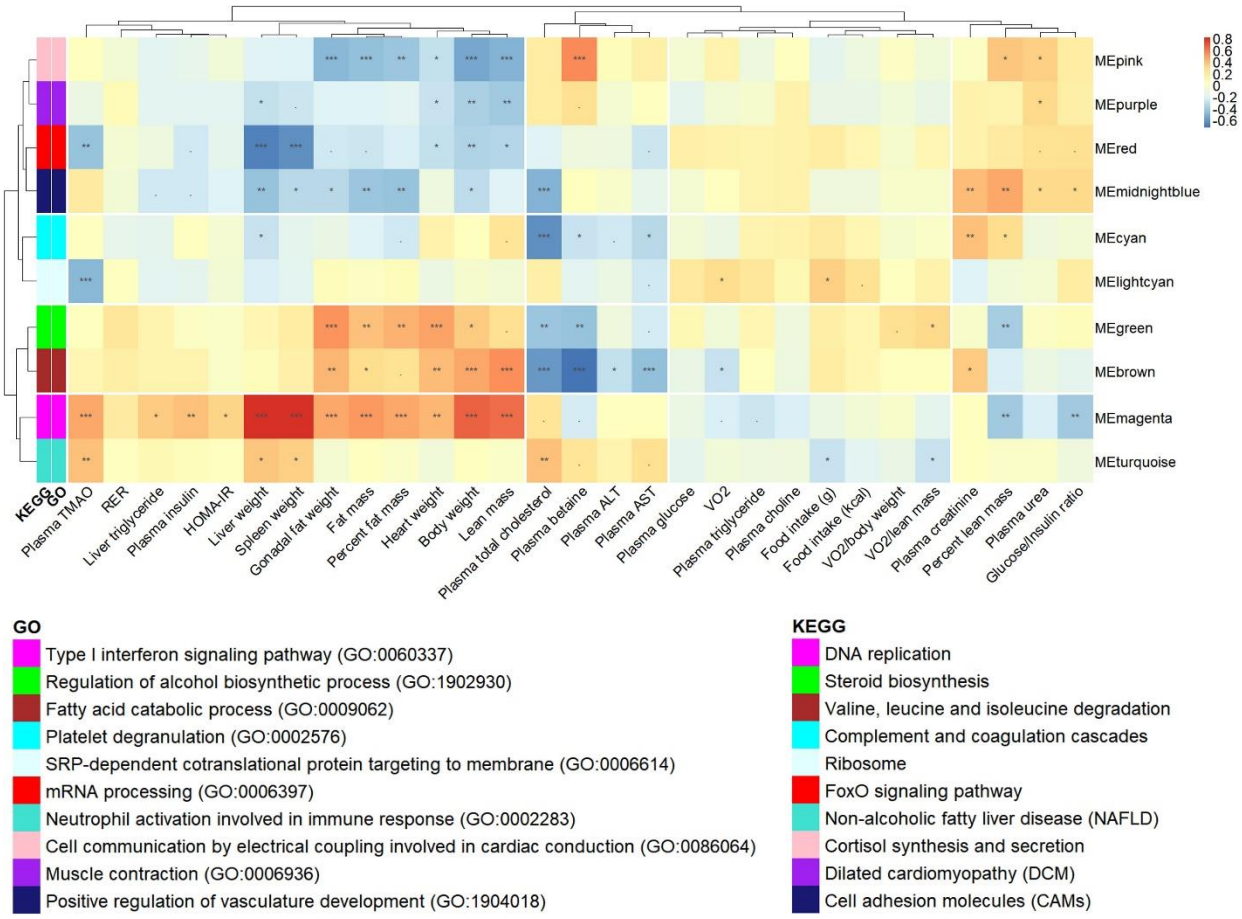


Figure 3.10. Association of hepatic co-expression gene modules with metabolic traits in the eight CC founder strains in female mice.

Spearman correlation between liver gene modules and metabolic traits in all mice. Module names were shown along the right axis, and top-enriched GO and KEGG terms in the legend. "***"P < 0.001, "**"P < 0.01, "*"P < 0.05, "."P < 0.10.

The magenta module is associated with a number of physiological traits including plasma TMAO concentration, plasma insulin, body composition, tissue weights such as liver, spleen, gonadal fat, and heart, and liver TG (**Figure 3.10**). Notably, the magenta module is enriched for " Type 1 interferon signaling pathway (GO: 0060337)" (GO term: -logP > 11) (**Tables 3.8-3.10**). Transcripts in the magenta module included various interferon regulatory factors, interleukin receptors, and chemokine (C-C motif) receptors and were up-regulated in the liver of the B6 and NZO strains but down-regulated in the liver from CAST (**Figures 3.7E and Table 3.11**). A recent study revealed that activation of type I interferon in $CD8^+$ T cells is associated with insulin resistance in fatty liver patients and obese mice (Ghazarian et al., 2017). Increased type I IFN signaling stimulates macrophage recruitment to lesions, promoting atherosclerosis in mice (Goossens et al., 2010). In our study, various genes related to type I IFN signaling were up-regulated in HFCA diet-fed mice, especially in B6 and NZO mice which are susceptible to MetSyn (**Figure 3.7E and Table 3.11)**. This suggests that the increased hepatic lipid content and TMAO concentrations in B6 and NZO may be due to an up-regulation or down-regulation of transcripts involved in the type 1 interferon signaling pathway.

We further investigated the relationship between strain and diet in the magenta module. The ME in the magenta module was higher in HFCA diet-fed mice than in AIN-93M diet-fed mice independent of strain ($P < 0.01$; **Figure 3.7E and Table 3.11**), and this trend was also observed in five individual strain analyses ($P < 0.01$ in B6 strain, $P < 0.05$ for 129 strain, $P < 0.05$ for NOD strain, $P < 0.01$ in NZO strain, and $P < 0.01$ in WSB strain). Furthermore, the magenta module had a higher proportion of genes upregulated by HFCA-diet (29.7%) than that of genes upregulated by AIN-93M diet (13.6%) (**Figure 3.7G and Tables 3.12 and 3.13**). The ME (PC1) of the magenta module was the highest in NZO mice and the lowest in CAST mice (**Figure 3.7E**

and Table 3.11), similar to the metabolic traits pattern by genetic backgrounds (**Table 3.14**). In this regard, we speculated that the genes clustered in the magenta module may increase their expression by the HFCA diet challenge and regulate metabolic traits.

Among the 441 genes included in the magenta module, we first identified 30 genes with >2-fold increased expression in HFCA versus control groups by DEG analysis (**Table 3.15**). Most of these 30 genes showed higher transcript abundance in laboratory-inbred strains (A/J, B6, 129, NOD, NZO, and WSB) than wild-derived strains (CAST and PWK). For example, the 30 genes that generally have the highest transcript abundance in NZO, A/J, and B6 strains also have the lowest transcript abundance in the CAST strain. Following a similar pattern, a number of clinically relevant MetSyn traits such as body composition, tissue weights, liver TG, and insulin resistance were highest in B6 and NZO mice and lowest in CAST strain (**Table 3.14 and 3.15**). In terms of diet-by-strain interaction for the 30 genes, laboratory-inbred strains (A/J, B6, 129, NOD, NZO, and WSB) generally showed diet-sensitive abundance changes, whereas wild-derived strains (CAST and PWK) showed a diet-resistant response (**Table 3.15**).

Table 3.15. List of 30 magenta module genes that were differentially expressed by diets (log2 fold change > 1) in eight CC founder strains. Significant diet effect and strain effect of 30 genes were displayed. n = 48 (6 mice per eight strains and 24 mice per two diets)

		Diet effect (HFCA vs AIN-93M)		Strain effect (Eight founder strains)		
Gene	**Name**	**Log2FC (HFCA/AIN93)**	**Adj.p-value for DEG analysis**	**p-value for comparison in strains (Kruskal-Wallis test)**	**Strain that has the highest value**	**Strain that has the lowest value**
Lilrb4a	leukocyte immunoglobulin-like receptor, subfamily B, member 4A	1.89	5.94E-06	5.30E-03	**A/J**	**CAST**
Nox4	NADPH oxidase 4	1.76	4.11E-04	1.40E-03	**B6**	**CAST**
Ccr5	chemokine (C-C motif) receptor 5	1.64	2.43E-06	0.015	**NZO**	**CAST**
Mup20	major urinary protein 20	1.60	4.37E-03	1.50E-04	**B6**	**CAST**
Adgre1	adhesion G protein-coupled receptor E1	1.60	8.86E-05	1.80E-03	**B6**	**CAST**
Clec4n	C-type lectin domain family 4, member n	1.58	2.57E-05	6.30E-04	**A/J**	**CAST**
Cd300ld	CD300 molecule like family member d	1.54	4.56E-03	3.90E-05	**B6**	**CAST**
Cxcl10	chemokine (C-X-C motif) ligand 10	1.48	1.73E-04	8.80E-03	**NZO**	**CAST**
Siglec1	sialic acid binding Ig-like lectin 1, sialoadhesin	1.41	1.07E-04	0.01	**NZO**	**CAST**
Abcg3	ATP binding cassette subfamily G member 3	1.40	3.53E-06	0.025	**B6**	**CAST**
Kcnj10	potassium inwardly-rectifying channel, subfamily J, member 10	1.38	1.80E-04	2.60E-03	**A/J**	**CAST**
Cd72	CD72 antigen	1.38	2.47E-05	9.10E-04	**B6**	**CAST**
Apol9a	apolipoprotein L 9a	1.34	8.48E-03	9.50E-04	**NZO**	**CAST**
Mcm2	minichromosome maintenance complex component 2	1.31	2.06E-08	NS	**NZO**	**CAST**
Mlkl	mixed lineage kinase domain-like	1.28	7.18E-06	3.60E-03	**A/J**	**CAST**
Oasl2	2'-5' oligoadenylate synthetase-like 2	1.27	1.36E-04	8.30E-04	**PWK**	**CAST**
Slfn9	schlafen 9	1.24	5.67E-06	0.011	**B6**	**CAST**
Fabp7	fatty acid binding protein 7, brain	1.22	1.86E-04	4.10E-03	**NZO**	**CAST**
Trim30a	tripartite motif-containing 30A	1.21	4.19E-07	0.024	**B6**	**CAST**
Ripor2	RHO family interacting cell polarization regulator 2	1.17	5.80E-05	8.00E-04	**NZO**	**CAST**
Slfn8	schlafen 8	1.13	1.82E-03	1.80E-05	**B6**	**CAST**

Gbp8	guanylate-binding protein 8	1.09	3.39E-03	1.80E-04	**A/J**	**CAST**
Lilra5	leukocyte immunoglobulin-like receptor, subfamily A (with TM domain), member 5	1.09	6.74E-05	6.30E-03	**NZO**	**CAST**
Mcm5	minichromosome maintenance complex component 5	1.07	1.51E-04	0.036	**NZO**	**CAST**
Ly6e	lymphocyte antigen 6 complex, locus E	1.07	9.86E-04	1.50E-03	**NZO**	**CAST**
Ifi44	interferon-induced protein 44	1.05	2.07E-02	2.10E-04	**NZO**	**CAST**
Ifit2	interferon-induced protein with tetratricopeptide repeats 2	1.02	1.62E-04	3.40E-03	**NZO**	**CAST**
Serpina3f	serine (or cysteine) peptidase inhibitor, clade A, member 3F	1.02	4.38E-03	1.90E-03	**NZO**	**CAST**
Oasl1	2'-5' oligoadenylate synthetase-like 1	1.02	8.69E-03	7.00E-05	**NZO**	**CAST**
Oas2	2'-5' oligoadenylate synthetase 2	1.02	2.56E-03	8.60E-04	**B6**	**CAST**

To investigate the association of the top 30 HFCA-specific DEGs identified in the magenta module with metabolic traits, we performed Spearman correlation and hierarchical clustering between the genes and traits (**Figure 3.11**). In general, most of the genes showed positive correlations with body composition, liver weight, spleen weight, liver TG, insulin resistance, and plasma TMAO. Among them, genes showing significant correlations with body weight, plasma TMAO, and liver TG were *Apol9a* (apolipoprotein L 9a), *Mcm2* (minichromosome maintenance complex component 2), *Siglec1* (sialic acid binding Ig-like lectin 1), *Ly6e* (lymphocyte antigen 6 complex, locus E), *Ripor2* (RHO family interacting cell polarization regulator 2), *Slfn8* (schlafen 8), and *Nox4* (NADPH oxidase 4). Among these genes, *Apol9a* and *Slfn8* showed a significant correlation with the clinical indexes of insulin resistance, such as plasma insulin, HOMA-IR, and glucose/insulin ratio (Wallace et al., 2004). Finally, *Nox4* had a number of clinically relevant correlations which we describe below.

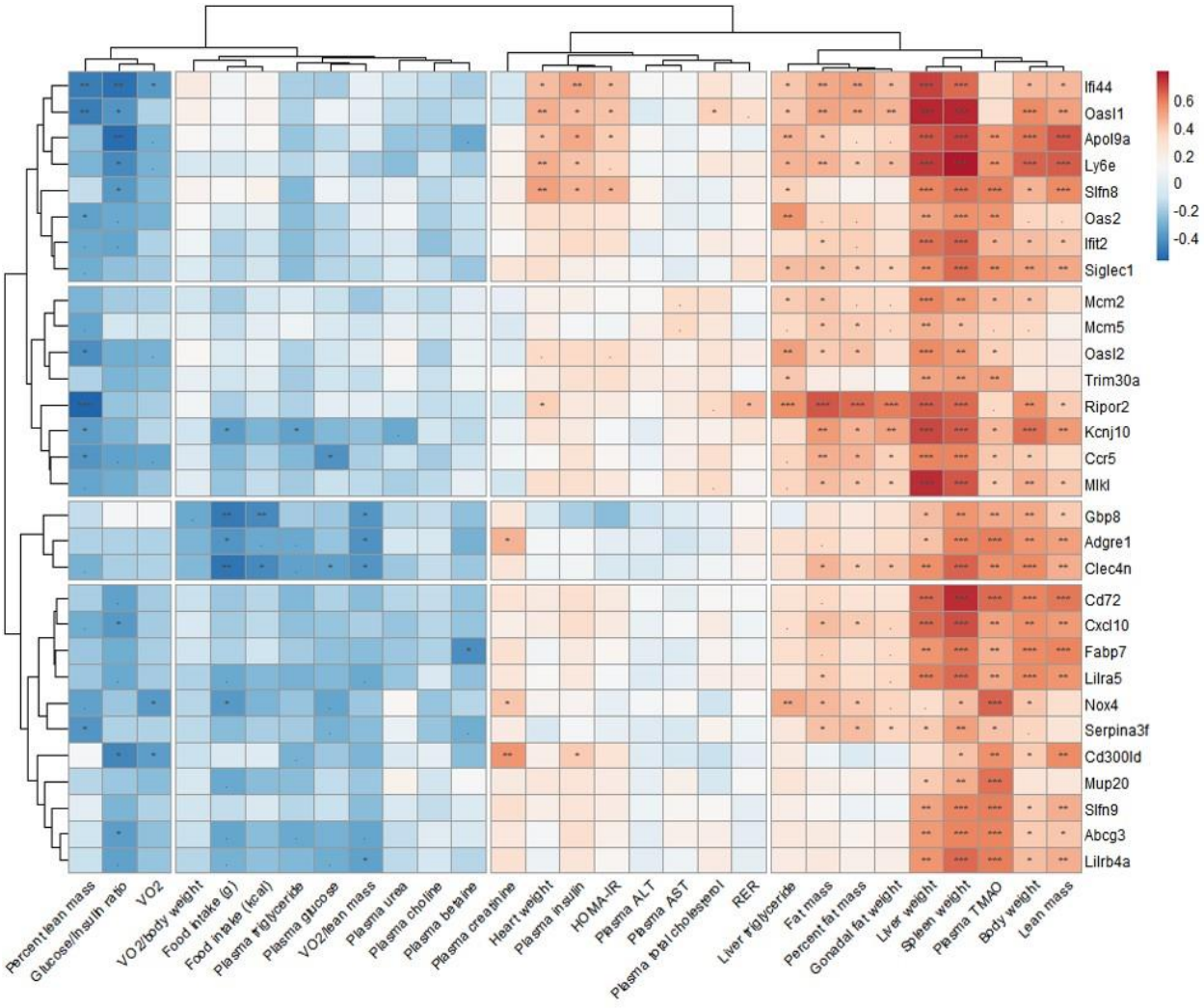


Figure 3.11. Association of hepatic magenta module genes with metabolic traits in the eight CC founder strains in female mice.

Spearman correlation between the top 30 HFCA-specific DEGs identified in the magenta module and metabolic traits in all mice. The p-values were adjusted using the BH FDR procedure. "***"P < 0.001, "**"P < 0.01, "*"P < 0.05, "."P < 0.10.

3.5.6. Association of Hepatic Gene Modules with Metabolic Traits Point to *Nox4*-associated Plasma TMAO and Liver TG Production

By referring to the literature and assessing the correlation with metabolic traits, we searched for putative genes involved in the regulation of MetSyn among the top 30 HFCA-specific DEGs identified in the magenta module. Interestingly, consistent with the WGCNA, DEG analysis, and strain effect analysis for magenta module genes, *Nox4* found to be a critical member of the magenta module. Among the genes identified in the magenta module (**Table 3.15),** *Nox4,* a hydrogen peroxide NADPH oxidase isoform and the primary source of inflammation-induced oxidative stress, was a significant DEG by diet (log2fold change (HFCA/AIN93) = 1.76, adjusted p-value = 4×10^{-4}) and strain (log2fold change (6 strains [A/J, B6, 129, NOD, NZO, and WSB] / 2 strains [CAST and PWK]) = 1.5, adjusted p-value = 0.02). In our study, the *Nox4* gene showed the highest associations with plasma TMAO (adjusted p-value = 6×10^{-5}; **Figure 3.11**) and with liver TG (adjusted p-value = 0.006; **Figure 3.11**). *Nox4* transcript levels were highest in B6 mice and lowest in CAST mice. Another key gene regulating TMAO levels is flavin-containing monooxygenase 3 (*Fmo3*) which converts TMA produced from dietary precursors such as choline and carnitine into TMAO by NADPH-dependent oxygenation in the liver (Lang et al., 1998). Considering that oxidative stress is caused by a deficiency in detoxification mechanisms, we observed the effects of genetic backgrounds on plasma TMAO production and gene abundance of *Fmo3* and *Nox4*. In terms of strain-dependent difference, hepatic *Nox4* abundance and plasma TMAO were highest in the B6 strain (**Figures 3.12A, 3.12D, and Table 3.16**), and hepatic *Fmo3* and *Nox4* abundance and plasma TMAO were lowest in the CAST strains (**Figures 3.12A, 3.12C, 3.12D, and Table 3.16**). In addition, plasma TMAO and *Nox4* abundance showed an increased pattern in HFCA diet-fed mice except for A/J

and CAST strains (**Figures 3.12A, 3.12D, and Table 3.16**), and the only significant diet-dependent difference was observed in the WSB strain. To investigate the relationship between plasma TMAO and *Fmo3* or *Nox4*, we performed the Spearman correlation between these transcripts and plasma TMAO concentration. Surprisingly, *Fmo3* transcript abundance was not significantly correlated with plasma TMAO in either diets (**Figure 3.12E**, AIN-93M diet: r = 0.19 and p = 0.39; HFCA diet: r = 0.39 and p = 0.066). Conversely, *Nox4* transcript abundance was highly correlated with plasma TMAO (r = 0.84 and p = $1.7x10^{-6}$) in AIN-93M diet-fed mice and was the strongest correlation between TMAO and any of 12,502 expressed transcripts (**Figure 3.12F**).

We next assessed how the 2 transcripts levels of these two genes were associated with hepatic lipid content another MetSyn trait that was highly associated with the magenta module. In previous studies, FMO3 overexpression increased hepatic and plasma lipids in atherosclerosis susceptible mice (Shih et al., 2015) and NOX4 deficient mouse model was shown to have reduced fibrosis due to inhibition of TGF-β-induced apoptosis in epithelial cells (Carnesecchi et al., 2011). In our study, the liver TG and MEs for the brown module containing the *Fmo3* gene and for the magenta module containing *Nox4* gene were similarly high in NZO mice and low in CAST mice (**Tables 3.11, 3.14, and 3.17**). Liver TG and *Nox4* abundance showed an increased pattern in HFCA diet-fed mice except for A/J and CAST strains (**Figures 3.12B, 3.12D, and Table 3.17**). Furthermore, *Fmo3* expression was significantly correlated with liver TG in both diets (**Figure 3.12G**, AIN-93M diet: r = 0.54 and p = 0.0066; HFCA diet: r = 0.65 and p = 0.0011), and *Nox4* expression was highly correlated with liver TG in HFCA diet-fed mice (r = 0.67 and p = 0.00057; **Figure 3.12H**).

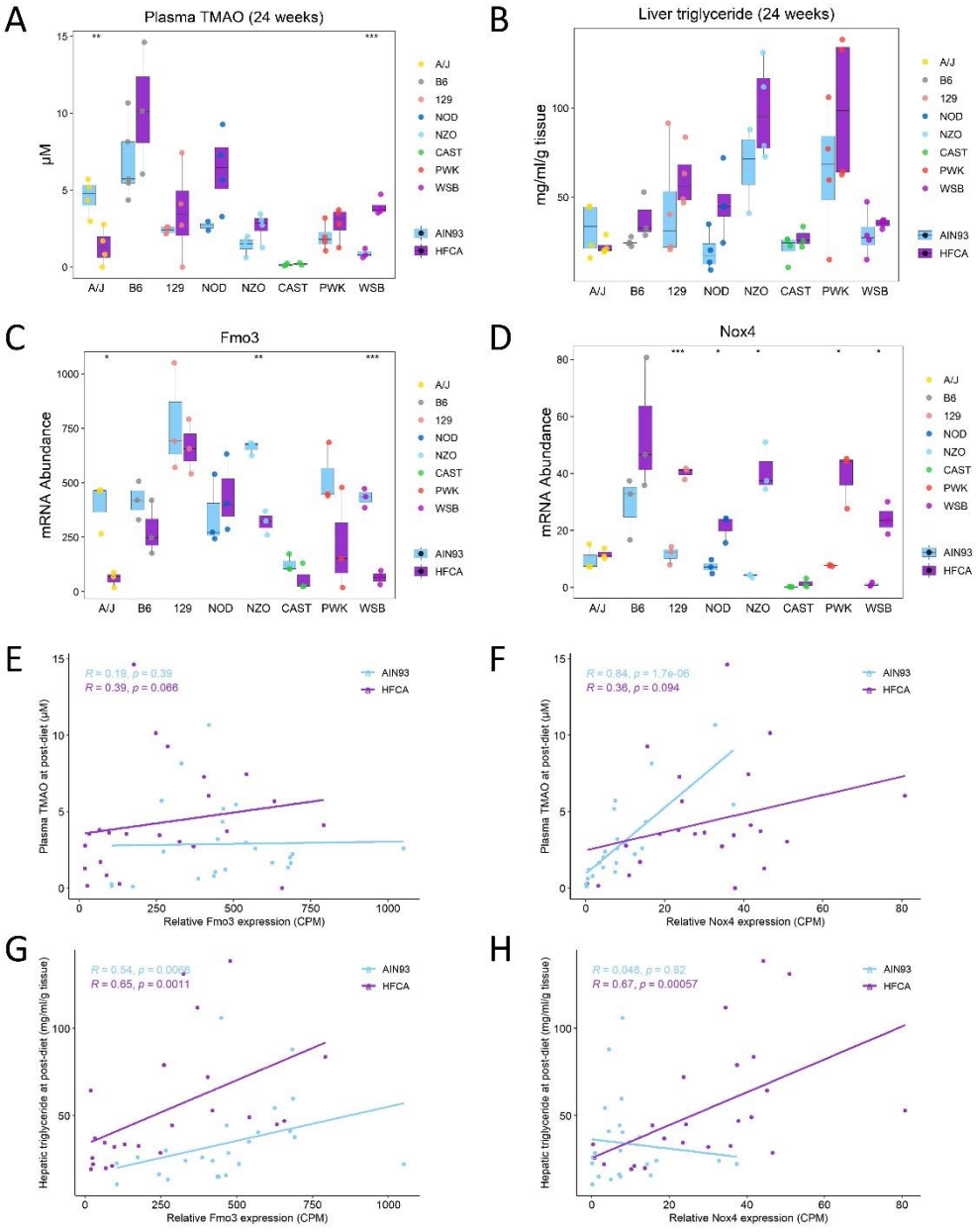


Figure 3.12. Association of hepatic gene modules with metabolic traits point to *Nox4*-associated plasma TMAO and creatinine production.

Effect of diet or genetic background on plasma TMAO (A) and liver TG (B), and hepatic expression of *Fmo3* (C) and *Nox4* (D). Spearman correlation between plasma TMAO and *Fmo3* (E) or *Nox4* (F) abundance. Spearman correlation between liver TG and *Fmo3* (G) or *Nox4* (H) abundance. "***"P < 0.001, "**"P < 0.01, "*"P < 0.05. N = 6 mice for each founder, 3 for AIN-93M diet (lightblue color) and 3 for HFCA diet-fed mice (purple color). A/J (yellow), B6 (gray), C57BL/6J; 129 (pink), 129S1/SvlmJ; NOD (blue), NOD/ShiLtJ; NZO (lightblue), NZO/HILtJ; CAST (green), CAST/EiJ; PWJ (red), PWK/PhJ; WSB (purple), WSB/EiJ.

Table 3.16. Diet-, strain-, and diet-by strain-dependent differences of plasma TMAO using Tukey's multiple comparison test in eight CC founder strains. n = 64 (33 mice for AIN-93M diet and 31 mice for HFCA diet).

Diet		Δ means	CI_lo	CI_hi	Adj.p_value
HFCA	AIN93	1.15	0.21	2.08	0.017

Strain		Δ means	CI_lo	CI_hi	Adj.p_value
CAST	B6	-7.97	-11.17	-4.77	0.00E+00
NZO	B6	-6.13	-8.94	-3.33	4.00E-07
PWK	B6	-5.75	-8.56	-2.94	1.80E-06
WSB	B6	-5.53	-8.44	-2.63	8.30E-06
B6	A/J	5.21	2.40	8.01	1.38E-05
129	B6	-5.17	-7.97	-2.36	1.59E-05
CAST	NOD	-4.97	-8.37	-1.57	7.70E-04
NZO	NOD	-3.13	-6.16	-0.10	0.039

Diet x strain				Δ means	CI_lo	CI_hi	Adj.p_value
B6	HFCA	CAST	AIN93	10.12	4.88	15.35	1.40E-06
B6	HFCA	A/J	HFCA	8.94	4.04	13.83	5.10E-06
B6	HFCA	NZO	AIN93	8.86	3.96	13.75	6.10E-06
B6	HFCA	WSB	AIN93	9.39	4.16	14.62	7.30E-06
CAST	HFCA	B6	HFCA	-10.06	-15.91	-4.21	1.80E-05
B6	HFCA	PWK	AIN93	8.30	3.41	13.20	2.36E-05
B6	HFCA	129	AIN93	7.87	2.98	12.76	6.71E-05
NZO	HFCA	B6	HFCA	-7.64	-12.54	-2.75	1.16E-04
PWK	HFCA	B6	HFCA	-7.44	-12.33	-2.54	1.90E-04
CAST	AIN93	B6	AIN93	-6.73	-11.40	-2.05	4.92E-04
129	HFCA	B6	HFCA	-6.70	-11.59	-1.81	1.07E-03
B6	HFCA	NOD	AIN93	7.59	1.74	13.44	2.38E-03
WSB	HFCA	B6	HFCA	-6.34	-11.23	-1.45	2.43E-03
A/J	HFCA	B6	AIN93	-5.55	-9.85	-1.25	2.56E-03
WSB	AIN93	B6	AIN93	-6.00	-10.68	-1.32	2.79E-03
NZO	AIN93	B6	AIN93	-5.47	-9.77	-1.17	3.14E-03
NOD	HFCA	CAST	AIN93	6.22	1.33	11.12	3.15E-03
CAST	HFCA	B6	AIN93	-6.67	-12.03	-1.31	4.27E-03
B6	HFCA	A/J	AIN93	5.71	0.81	10.60	9.72E-03
PWK	AIN93	B6	AIN93	-4.91	-9.21	-0.62	0.012
NOD	HFCA	WSB	AIN93	5.50	0.61	10.40	0.015
NOD	HFCA	A/J	HFCA	5.05	0.52	9.58	0.017
CAST	HFCA	NOD	HFCA	-6.17	-11.72	-0.62	0.017
NOD	HFCA	NZO	AIN93	4.97	0.44	9.50	0.020
129	AIN93	B6	AIN93	-4.48	-8.78	-0.18	0.034

Table 3.17. Diet-, strain-, and diet-by strain-dependent differences of liver TG using Tukey's multiple comparison test in eight CC founder strains. n = 64 (33 mice for AIN-93M diet and 31 mice for HFCA diet).

Diet		Δ means	CI_lo	CI_hi	Adj.p_value
HFCA	AIN93	15.75	5.09	26.40	4.65E-03

Strain		Δ means	CI_lo	CI_hi	Adj.p_value
CAST	NZO	-59.43	-93.90	-24.97	4.42E-05
PWK	CAST	58.15	23.69	92.62	6.59E-05
NZO	A/J	56.09	22.80	89.39	6.77E-05
PWK	A/J	54.81	21.52	88.11	1.02E-04
NZO	B6	53.64	20.34	86.94	1.49E-04
PWK	B6	52.36	19.06	85.66	2.23E-04
WSB	NZO	-51.09	-84.39	-17.80	3.31E-04
NZO	NOD	50.36	17.06	83.65	4.17E-04
WSB	PWK	-49.81	-83.11	-16.52	4.94E-04
PWK	NOD	49.08	15.78	82.37	6.20E-04

Diet x strain				Δ means	CI_lo	CI_hi	Adj.p_value
PWK	HFCA	NOD	AIN93	80.15	26.44	133.85	2.23E-04
NZO	HFCA	NOD	AIN93	79.24	25.54	132.95	2.72E-04
PWK	HFCA	B6	AIN93	74.99	24.04	125.93	2.85E-04
PWK	HFCA	CAST	AIN93	78.18	24.48	131.88	3.46E-04
NZO	HFCA	B6	AIN93	74.08	23.14	125.03	3.52E-04
NZO	HFCA	CAST	AIN93	77.28	23.57	130.98	4.22E-04
PWK	HFCA	A/J	HFCA	77.22	23.51	130.92	4.28E-04
NZO	HFCA	A/J	HFCA	76.32	22.61	130.02	5.22E-04
PWK	HFCA	WSB	AIN93	70.35	16.64	124.05	1.91E-03
NZO	HFCA	WSB	AIN93	69.45	15.74	123.15	2.32E-03
PWK	HFCA	A/J	AIN93	67.56	13.86	121.27	3.44E-03
PWK	HFCA	CAST	HFCA	72.46	14.45	130.46	3.80E-03
NZO	HFCA	A/J	AIN93	66.66	12.96	120.36	4.15E-03
CAST	HFCA	NZO	HFCA	-71.55	-129.56	-13.55	4.52E-03
WSB	HFCA	PWK	HFCA	-64.43	-118.13	-10.73	6.55E-03
WSB	HFCA	NZO	HFCA	-63.53	-117.23	-9.83	7.86E-03
PWK	HFCA	B6	HFCA	61.52	3.52	119.53	0.028
NZO	HFCA	B6	HFCA	60.62	2.61	118.63	0.033
PWK	HFCA	129	AIN93	55.87	2.17	109.58	0.034
NZO	HFCA	129	AIN93	54.97	1.27	108.68	0.040

Lastly, we investigated genome variation and haplotypes for the two genes to identify if there are unique haplotypes in the founder strains that could explain the strain variation in plasma TMAO and liver TG. We specifically assessed if the missense variants of the two genes could be functional variants that might explain the expression differences. We identified the total number of SNPs in the *Fmo3* and *Nox4* genes, calculated the similarity in the SNPs of these genes between the reference genome B6 strain and other strains, and obtained the number of missense variants and functional variants that may affect amino acid substitution using PROVEAN and SIFT *in silico* analysis. For example, we compared the similarity of each strain to the B6 strain in 658 SNPs in *Fmo3* and 2,641 SNPs in *Nox4*. As expected, the wild-derived strains CAST and PWK contained the most genetic diversity at these loci and have considerable genetic variation as compared to the B6 strain (**Table 3.18**). In particular, *Nox4* showed the greatest diversity (37.37% similarity) in SNPs between B6 and CAST strains, and both liver *Nox4* transcript abundance and plasma TMAO concentration had the greatest difference between B6 and CAST strains suggesting that the candidate gene *Nox4* has shared haplotypes in the strains with shared phenotypes (**Figure 3.12A, 3.12D, Table 3.16**). We also identified 9 SNP missense variants from *Fmo3* SNPs and 6 missense SNP variants from *Nox4* SNPs, of which two variants from *Fmo3* (rs37325482 at 162,967,784 bp and rs50797400 at 162,968,776 bp) and two variants from *Nox4* (rs250243260 at 87,246,820 bp and rs217947741 at 87,246,834 bp) were predicted to have a known deleterious structural consequence by *in silico* analysis (**Table 3.19**). Furthermore, we confirmed both *Nox4* functional variants were caused by SNPs in the CAST strain (**Table 3.19**).

Table 3.18. Similarity of each strain to the B6 strain among eight CC founder strains in total 658 SNPs in *Fmo3* and 2,641 SNPs in *Nox4*.

Strain	Fmo3 SNPs			Nox4 SNPs		
	% of SNP similarity	#SNPs in strain	#Total SNPs	% of SNP similarity	#SNPs in strain	#Total SNPs
A/J	90.73%	61	658	74.90%	663	2641
129	90.27%	64	658	81.29%	494	2641
NOD	94.53%	36	658	73.65%	696	2641
NZO	89.51%	69	658	81.33%	493	2641
CAST	49.09%	335	658	37.37%	1654	2641
PWK	24.47%	497	658	51.46%	1282	2641
WSB	66.11%	223	658	81.18%	497	2641

Table 3.19. Missense variants and predicted functional variants by PROVEAN and SIFT analysis of *Fmo3* and *Nox4* genes.

Chr	Position (bp)	Gene	dbSNP	B6 haplotype (Ref)	A/J	129	NOD	NZO	CAST	PWK	WSB	Ref amino acid	Altered amino acid	Prediction by PROVEAN	Prediction by SIFT
1	162,954,207	Fmo3	rs37584253	G	-	-	-	-	T	T	T	L	I	Neutral	Tolerated
1	162,955,216	Fmo3	rs244712117	C	-	-	-	-	-	-	T	E	K	Neutral	Tolerated
1	162,958,458	Fmo3	rs36964758	G	-	-	-	-	T	-	-	A	D	Neutral	Tolerated
1	162,958,468	Fmo3	rs32758001	T	-	-	-	-	-	C	-	M	V	Neutral	Tolerated
1	162,966,899	Fmo3	rs50521266	C	-	-	-	-	-	A	-	A	S	Neutral	Tolerated
1	162,967,784	Fmo3	**rs37325482**	G	-	-	-	-	-	-	**A**	P	S	**Deleterious**	**Damaging**
1	162,967,913	Fmo3	rs36935260	T	-	-	-	-	C	C	C	N	D	Neutral	Tolerated
1	162,968,776	Fmo3	**rs50797400**	C	-	-	-	-	-	**T**	-	D	N	**Deleterious**	**Damaging**
1	162,968,824	Fmo3	rs246719414	T	-	-	-	-	C	-	-	T	A	Neutral	Tolerated
7	87,246,768	Nox4	rs240455081	G	-	-	-	-	T	-	-	D	Y	Neutral	NA
7	87,246,786	Nox4	rs32414336	C	A	-	A	-	-	-	-	L	M	Neutral	Tolerated
7	87,246,820	Nox4	**rs250243260**	G	-	-	-	-	**A**	-	-	G	E	Neutral	**Damaging**
7	87,246,834	Nox4	**rs217947741**	C	-	-	-	-	**T**	-	-	R	C	Neutral	**Damaging**
7	87,321,631	Nox4	rs37089859	A	-	-	-	-	G	-	-	I	V	Neutral	Tolerated
7	87,395,824	Nox4	rs38453194	A	-	-	-	-	G	-	-	I	V	Neutral	NA

3.6. Discussion

In this study, we utilized the eight genetically diverse CC founder mouse strains fed AIN-93M or HFCA diet to assess the effect of diet or genetic backgrounds on MetSyn. We found that host genetics had a strong influence on the metabolic traits, including body composition, tissue weights, lipid panels, TMAO analytes, markers related to liver and kidney function, glucose metabolism, and energy expenditure in response to a range of metabolic stimuli. Consistent with previous studies, we found that of the eight strains, CAST had the lowest body adiposity, liver TG, and plasma TMAO levels, and the NZO mice were the most obese and insulin-resistant (Kreznar et al., 2017; Mitok et al., 2018). These results highlight the effects of genetic background on the liver transcriptome and phenotypic responses to diet, and also demonstrate the effectiveness of the CC founder strains in nutrigenomics studies.

The CC founder strains have been individually studied in several tissues, and Chick et al. (Chick et al., 2016) conducted liver proteome profiling across the eight CC founder strains. Our study represents an in-depth analysis of the effects of diet on the hepatic transcriptional network. To identify liver transcripts that underlie the diet- and strain-dependent differences in metabolic traits, we conducted transcriptomic analysis on liver collected from each strain-fed two different diets and identified and quantified 12,502 transcripts. Hierarchical clustering of the liver transcriptome indicated that mouse liver transcripts were clustered according to diet and strain. In particular, the wild-derived strains PWK and CAST, have altered transcript levels compared to other the other CC progenitor strains. This suggests that genetic variation has a profound impact on the liver transcriptome, which may be associated with differences in metabolic traits. Gene modules were determined using WGCNA which clustered co-expressed transcripts and these modules were enriched for specific functional terms and correlated with metabolic traits.

In this study, we also sought to find core modules with diet-responsive genes by elucidating the link between diet-specific DEGs and gene networks. We first identified 6,411 genes that were differentially regulated by diet independent of strain, and 9,228 genes responded by diet perturbation in at least one strain. Next, we identified 20 modules in gene network analysis and found that the top five modules (turquoise, brown, green, red, and magenta) contained 85.4% of 6,411 DEGs independent of strain and 67.5% of 9,228 DEGs that responded in at least one strain. These modules were enriched for functional annotations, associated with traits related to MetSyn, showed significant diet effects in at least 5 strains. These results show that diet responsive genes identified by all-strain or individual strain DEG analysis are segregated into specific gene networks that are enriched for functions associated with MetSyn phenotypes.

Correlations can be a method to build hypotheses to test causality. For example, the magenta module is enriched for type I interferon signaling pathway and positively correlated with body composition, tissue weights, insulin resistance (plasma insulin and HOMA-IR), liver TG, and plasma TMAO. In addition, the ME and the HFCA diet-specific DEG abundance in the magenta module were highest in the B6 and NZO strains that are susceptible to obesity and metabolic dysregulation associated with MetSyn. Conversely, the ME was the lowest in the CAST strain which is resistant to MetSyn (Joost and Schurmann, 2014; Mitok et al., 2018; O'Connor et al., 2014). The type I interferon signaling is involved in metabolic regulation, tissue inflammation, and MetSyn which are related to atherosclerosis, obesity, fatty liver, and type 1 diabetes (Alsaggar et al., 2017; Chen et al., 2020; Ghazarian et al., 2017; Marro et al., 2019; Somers et al., 2012; Wieser et al., 2018). Therefore, the role of type I interferon is critical in adipose tissue, hepatocytes, immune cells, and endothelial cells. Transcripts in the magenta module include interferon-induced protein 44 (*Ifi44*), interferon-induced protein with tetratricopeptide repeats 1 (*Ifit1*), C-X-C motif

chemokine ligand 10 (*Cxcl10*), NADPH oxidase 4 (*Nox4*), and 2'-5'-oligoadenylate synthase-like protein 1 (*Oasl1*), which are known to be up-regulated in the liver under high-fat diet feeding conditions or stimulation of type I interferon (Ghazarian et al., 2017; Ivashkiv and Donlin, 2014; Paik et al., 2014; Zhai et al., 2008). Recently IRF7, a master regulator of type I interferon response, showed increased expression in liver and adipose tissue from obese mice compared to controls, indicating increased type I interferon signaling (Wang et al., 2013). IRF7 knockout mice were resistant to diet-induced obesity, insulin resistance, and inflammation. The IRF7, together with our results, is associated with type I interferon signaling and modulates diet-induced obesity, insulin resistance, and their metabolic consequences. Other novel transcripts in the magenta module may also be important for type I interferon signaling and/or MetSyn.

Driven by our preliminary finding that plasma TMAO production can be modulated by FMO3 enzyme activity in the liver (**7**), we asked whether *Fmo3* gene abundance was differentially abundant across the eight CC founder strains. Liver transcriptomic analysis revealed that hepatic *Fmo3* expression is higher in B6, 129, and NZO mice and the lowest in CAST mice, consistent with the FMO3 protein abundance previously reported (Chick et al., 2016). FMO3 is a catalytic enzyme that converts TMA, produced from dietary precursors such as choline and carnitine by gut microbiota, into TMAO by host-dependent hepatic N-oxygenation in the liver (Lang et al., 1998). TMAO is mechanistically linked with MetSyn and other diseases, including atherosclerosis, obesity, fatty liver, type 2 diabetes mellitus, and chronic kidney disease (Barrea et al., 2018; Chen et al., 2016; Koeth et al., 2013; Lent-Schochet et al., 2018; Schugar et al., 2017; Stubbs et al., 2016; Wang et al., 2011), suggesting that the TMAO pathway may also be linked to the pathogenesis of MetSyn.

The fact that there is a positive correlation between *Fmo3* expression and plasma TMAO production among mouse strains susceptible to atherosclerosis highlights the

importance of FMO3 in the development of cardiovascular disease (Bennett et al., 2013). However, in our study using mice with highly divergent genetic backgrounds, no significant correlation was observed between *Fmo3* expression and plasma TMAO in either diet. For this reason, we searched for genes that are functionally similar to FMO enzymes and have an association with TMAO and focused on *Nox4*, which showed the highest correlation with plasma TMAO and was identified in the magenta module.

NOX4 is a hydrogen peroxide NADPH oxidase isoform and major producers of reactive oxidative species by transferring electrons from NADPH to molecular oxygen. NOX4 is found in various cardiovascular cells and tissues, which is involved in conditions related to MetSyn such as atherosclerosis (Lozhkin et al., 2017; Schurmann et al., 2015), hypertension (Bouabout et al., 2018), fatty liver (Leon-Mimila et al., 2020; Rabelo et al., 2018), insulin resistance (Den Hartigh et al., 2017), obesity (Jiang et al., 2011), and kidney injury (Jeong et al., 2018). Similar to NOX4, FMOs catalyze the NADPH-dependent oxidative metabolism of a variety of foreign chemicals, including dietary compounds, drugs, and environmental pollutants (Krueger and Williams, 2005). Genetic analysis of patients with trimethylaminuria shows that the lack of FMO3 enzymatic activity often occurs in mutations that affect the binding of necessary cofactors FAD (Zhang et al., 2003) or NADPH (Fujieda et al., 2003) highlighting the fundamental importance of these cofactors in the function of FMOs.

Our studies identified *Nox4*, as a potentially important gene related to TMAO concentrations. We note that this relationship is based on a transcriptional network and does not necessarily indicate a specific direct (i.e.- enzymatic) action of *Nox4* leading to TMAO production. Rather our association highlights the complex transcriptional network relating TMAO concentrations. For example, toll-like receptor 4 (TLR4), which is activated by the type I interferon signaling pathway, regulates the expression of *Fmo3* (Zhang et al., 2009),

and TLR4 and *Nox4* show direct interaction in several cell lines including hepatocytes (Park et al., 2004; Patel et al., 2006; Singh et al., 2017), thereby affecting the production of inflammatory mediators that are encoded by genes identified in the magenta module. In addition, ablation of *Nox4* lowers plasma homocysteine and betaine levels in mice, and NOX4 protects against acetaminophen-induced hepatotoxicity (Murray et al., 2015). FMO3 has also been shown to protect the liver from acetaminophen-induced hepatotoxicity (Rudraiah et al., 2014), and TMAO precursors choline and betaine affect homocysteine levels as methyl donors in one-carbon metabolism (Craciunescu et al., 2010). Our liver transcriptomic analysis revealed that *Nox4*, the only NADPH oxidase identified in the magenta module highly associated with MetSyn, has the highest correlation with plasma TMAO, as well as the highest expression in B6 and the lowest in CAST strain.

Our studies utilize 8 strains that are genetically diverse and thus we assessed the genetic variation at the *Nox4* locus. The *Nox4* allele from the CAST strain was contained a number of SNPs divergent from B6, and this difference is consistent with strain effects identified in liver *Nox4* transcript and TMAO. More interestingly, *Nox4* was revealed to be one of the genes that showed the highest correlation with liver TG among the genes identified in the magenta module, which is consistent with previous studies that demonstrated the association between *Nox4* and fatty liver in the mouse and humans (Bettaieb et al., 2015; Carnesecchi et al., 2011; Leon-Mimila et al., 2020; Rabelo et al., 2018). Both NOX4 and TMAO induce PKR/PERK activation (Bettaieb et al., 2015; Chen et al., 2019), leading to propagation of ER stress, which may contribute to the pathogenesis of fatty liver and MetSyn. These studies are further supported by associations between SNPs in NOX4 and metabolic syndrome in humans (He et al., 2018; Siqueira et al., 2015). In our *in-silico* analysis of the mouse variants suggests that functional variants of the *Nox4* gene may cause deleterious consequence in the protein structure and affect metabolic traits. These observations point to a

genetic underpinning of NOX4 in MetSyn which is similar to the findings of the current study.

Here we show that among the eight CC founder strains, B6 mice express a high level of *Nox4* and have high production of TMAO. Why would B6 liver synthesize the most TMAO, which contributes to the development of MetSyn, and CAST liver synthesize the least? One possible explanation is that the B6 strain has an impaired mechanism to regulate TMAO production or renal excretion. B6 is one of the strains with the highest plasma TMAO production among other inbred strains (Gregory et al., 2015), and it was reported to be more sensitive to damage caused by renal ischemia-reperfusion injury than the 129/Sv inbred strain (Lu et al., 2012). In addition, the CAST strain has the lowest hepatic expression of *Fmo3* of the CC founder strains which could explain reduced FMO3 enzyme activity and TMAO production. Further, studies utilizing the CC or the related Diversity Outbred population may shed further light on these strain differences.

In conclusion, our study provides a strong indication that host genetics affects the liver transcriptome under the intake of control or atherogenic diet. We also demonstrate that diet-by-strain interaction effects on the liver transcriptome are related to metabolic traits, suggesting that liver gene networks may underlie diet- or strain-dependent differences in MetSyn. The phenotype differences between mouse strains motivate us to find comparable phenotypic variations across the human population. Changes in plasma TMAO and liver TG similar to those seen in human MetSyn can be induced in B6 and NZO strains, but it should be taken into account that not all mouse strains developed MetSyn in this study. Understanding the changes in the liver transcriptome in response to diet and genetic background will be important to highlight the potential of precision nutrition and to understand interpersonal variability in disease risk. In this study, we saw dramatic strain variation in *Nox4* expression in the mouse liver that could determine the ability to generate

plasma TMAO and liver TG. Our results suggest that human genetic variations and variations in plasma TMAO and liver TG may contribute to the regulation of MetSyn.

3.7. Acknowledgments

We acknowledge Nikhil Joshi (University of California, Davis Bioinformatics Core) for assistance with RNA-Seq mapping analysis. This research was supported in part by R01HL128572 (BJB), a UNC Nutrition Obesity Research Consortium NORC pilot award from NIH P30DK056350 (BJB), and by USDA/ARS/Western Human Nutrition Research Center project funds 2032-51000-022-00D (BJB). The USDA is an equal opportunity employer.

3.8. Declaration of interests

The authors declare no conflicts of interest.

3.9. References

Alsaggar, M., Mills, M., and Liu, D. (2017). Interferon beta overexpression attenuates adipose tissue inflammation and high-fat diet-induced obesity and maintains glucose homeostasis. Gene Ther *24*, 60-66.

AlSiraj, Y., Chen, X., Thatcher, S.E., Temel, R.E., Cai, L., Blalock, E., Katz, W., Ali, H.M., Petriello, M., Deng, P., *et al.* (2019). XX sex chromosome complement promotes atherosclerosis in mice. Nat Commun *10*, 2631.

Altshuler, D., Hirschhorn, J.N., Klannemark, M., Lindgren, C.M., Vohl, M.C., Nemesh, J., Lane, C.R., Schaffner, S.F., Bolk, S., Brewer, C., *et al.* (2000). The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet *26*, 76-80.

Attie, A.D., Churchill, G.A., and Nadeau, J.H. (2017). How mice are indispensable for understanding obesity and diabetes genetics. Curr Opin Endocrinol Diabetes Obes *24*, 83-91.

Barrea, L., Annunziata, G., Muscogiuri, G., Di Somma, C., Laudisio, D., Maisto, M., de Alteriis, G., Tenore, G.C., Colao, A., and Savastano, S. (2018). Trimethylamine-N-oxide (TMAO) as Novel Potential Biomarker of Early Predictors of Metabolic Syndrome. Nutrients *10*, 1971.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - A Practical and Powerful Approach to Multiple Testing. J R Statst Soc B *57*, 289-300.

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

Bennett, B.J., de Aguiar Vallim, T.Q., Wang, Z., Shih, D.M., Meng, Y., Gregory, J., Allayee, H., Lee, R., Graham, M., Crooke, R., *et al.* (2013). Trimethylamine-N-oxide, a metabolite associated with atherosclerosis, exhibits complex genetic and dietary regulation. Cell Metab *17*, 49-60.

Bettaieb, A., Jiang, J.X., Sasaki, Y., Chao, T.I., Kiss, Z., Chen, X., Tian, J., Katsuyama, M., Yabe-Nishimura, C., Xi, Y., *et al.* (2015). Hepatocyte Nicotinamide Adenine Dinucleotide Phosphate Reduced Oxidase 4 Regulates Stress Signaling, Fibrosis, and Insulin Sensitivity During Development of Steatohepatitis in Mice. Gastroenterology *149*, 468-480 e410.

Bisschop, P.H., Bandsma, R.H., Stellaard, F., ter Harmsel, A., Meijer, A.J., Sauerwein, H.P., Kuipers, F., and Romijn, J.A. (2004). Low-fat, high-carbohydrate and high-fat, low-carbohydrate diets decrease primary bile acid synthesis in humans. Am J Clin Nutr *79*, 570-576.

Bouabout, G., Ayme-Dietrich, E., Jacob, H., Champy, M.F., Birling, M.C., Pavlovic, G., Madeira, L., Fertak, L.E., Petit-Demouliere, B., Sorg, T., *et al.* (2018). Nox4 genetic inhibition in experimental hypertension and metabolic syndrome. Arch Cardiovasc Dis *111*, 41-52.

Carnesecchi, S., Deffert, C., Donati, Y., Basset, O., Hinz, B., Preynat-Seauve, O., Guichard, C., Arbiser, J.L., Banfi, B., Pache, J.C., *et al.* (2011). A key role for NOX4 in epithelial cell death during development of lung fibrosis. Antioxid Redox Signal *15*, 607-619.

Cazanave, S., Podtelezhnikov, A., Jensen, K., Seneshaw, M., Kumar, D.P., Min, H.K., Santhekadur, P.K., Banini, B., Mauro, A.G., A, M.O., *et al.* (2017). The Transcriptomic Signature Of Disease Development And Progression Of Nonalcoholic Fatty Liver Disease. Sci Rep *7*, 17193.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr - interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *15*, 128.

Chen, H.J., Tas, S.W., and de Winther, M.P.J. (2020). Type-I interferons in atherosclerosis. J Exp Med *217*, e20190459.

Chen, S., Henderson, A., Petriello, M.C., Romano, K.A., Gearing, M., Miao, J., Schell, M., Sandoval-Espinola, W.J., Tao, J., Sha, B., *et al.* (2019). Trimethylamine N-Oxide Binds and Activates PERK to Promote Metabolic Dysfunction. Cell Metab *30*, 1141-1151 e1145.

Chen, Y.M., Liu, Y., Zhou, R.F., Chen, X.L., Wang, C., Tan, X.Y., Wang, L.J., Zheng, R.D., Zhang, H.W., Ling, W.H., *et al.* (2016). Associations of gut-flora-dependent metabolite trimethylamine-N-oxide, betaine and choline with non-alcoholic fatty liver disease in adults. Sci Rep *6*, 19076.

Cheng, A.A., Li, W., and Hernandez, L.L. (2018). Effect of high-fat diet feeding and associated transcriptome changes in the peak lactation mammary gland in C57BL/6 dams. Physiol Genomics *50*, 1059-1070.

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

Cho, C.E., Taesuwan, S., Malysheva, O.V., Bender, E., Tulchinsky, N.F., Yan, J., Sutter, J.L., and Caudill, M.A. (2017). Trimethylamine-N-oxide (TMAO) response to animal source foods varies among healthy young men and is influenced by their gut microbiota composition: A randomized controlled trial. Mol Nutr Food Res *61*, 1600324.

Choi, Y., and Chan, A.P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics *31*, 2745-2747.

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

Coffey, A.R., Kanke, M., Smallwood, T.L., Albright, J., Pitman, W., Gharaibeh, R.Z., Hua, K., Gertz, E., Biddinger, S.B., Temel, R.E., *et al.* (2019). microRNA-146a-5p association with the cardiometabolic disease risk factor TMAO. Physiol Genomics *51*, 59-71.

Coffey, A.R., Smallwood, T.L., Albright, J., Hua, K., Kanke, M., Pomp, D., Bennett, B.J., and Sethupathy, P. (2017). Systems genetics identifies a co-regulated module of liver microRNAs associated with plasma LDL cholesterol in murine diet-induced dyslipidemia. Physiol Genomics *49*, 618-629.

Cook, J.R., Matsumoto, M., Banks, A.S., Kitamura, T., Tsuchiya, K., and Accili, D. (2015). A mutant allele encoding DNA binding-deficient FoxO1 differentially regulates hepatic glucose and lipid metabolism. Diabetes *64*, 1951-1965.

Craciunescu, C.N., Johnson, A.R., and Zeisel, S.H. (2010). Dietary choline reverses some, but not all, effects of folate deficiency on neurogenesis and apoptosis in fetal mouse brain. J Nutr *140*, 1162-1166.

Daugherty, A. (2002). Mouse models of atherosclerosis. Am J Med Sci *323*, 3-10.

Den Hartigh, L.J., Omer, M., Goodspeed, L., Wang, S., Wietecha, T., O'Brien, K.D., and Han, C.Y. (2017). Adipocyte-Specific Deficiency of NADPH Oxidase 4 Delays the Onset of Insulin Resistance and Attenuates Adipose Tissue Inflammation in Obesity. Arterioscler Thromb Vasc Biol *37*, 466-475.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Fujieda, M., Yamazaki, H., Togashi, M., Saito, T., and Kamataki, T. (2003). Two novel single nucleotide polymorphisms (SNPs) of the FMO3 gene in Japanese. Drug Metab Pharmacokinet *18*, 333-335.

Getz, G.S., and Reardon, C.A. (2006). Diet and murine atherosclerosis. Arterioscler Thromb Vasc Biol *26*, 242-249.

Ghazarian, M., Revelo, X.S., Nøhr, M.K., Luck, H., Zeng, K., Lei, H., Tsai, S., Schroer, S.A., Park, Y.J., Chng, M.H.Y., *et al.* (2017). Type I Interferon Responses Drive Intrahepatic T cells to Promote Metabolic Syndrome. Sci Immunol *2*, 7616.

Goossens, P., Gijbels, M.J., Zernecke, A., Eijgelaar, W., Vergouwe, M.N., van der Made, I., Vanderlocht, J., Beckers, L., Buurman, W.A., Daemen, M.J., *et al.* (2010). Myeloid type I interferon signaling promotes atherosclerosis by stimulating macrophage recruitment to lesions. Cell Metab *12*, 142-153.

Grant, S.F., Thorleifsson, G., Reynisdottir, I., Benediktsson, R., Manolescu, A., Sainz, J., Helgason, A., Stefansson, H., Emilsson, V., Helgadottir, A., *et al.* (2006). Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet *38*, 320-323.

Gregory, J.C., Buffa, J.A., Org, E., Wang, Z., Levison, B.S., Zhu, W., Wagner, M.A., Bennett, B.J., Li, L., DiDonato, J.A., *et al.* (2015). Transmission of atherosclerosis susceptibility with gut microbial transplantation. J Biol Chem *290*, 5647-5660.

Hamaguchi, M., Kojima, T., Takeda, N., Nakagawa, T., Taniguchi, H., Fujii, K., Omatsu, T., Nakajima, T., Sarui, H., Shimazaki, M., *et al.* (2005). The metabolic syndrome as a predictor of nonalcoholic fatty liver disease. Ann Intern Med *143*, 722-728.

Hani, E.H., Boutin, P., Durand, E., Inoue, H., Permutt, M.A., Velho, G., and Froguel, P. (1998). Missense mutations in the pancreatic islet beta cell inwardly rectifying K+ channel gene (KIR6.2/BIR): a meta-analysis suggests a role in the polygenic basis of Type II diabetes mellitus in Caucasians. Diabetologia *41*, 1511-1515.

Hartiala, J., Bennett, B.J., Tang, W.H., Wang, Z., Stewart, A.F., Roberts, R., McPherson, R., Lusis, A.J., Hazen, S.L., Allayee, H., *et al.* (2014). Comparative genome-wide association studies in mice and humans for trimethylamine N-oxide, a proatherogenic metabolite of choline and L-carnitine. Arterioscler Thromb Vasc Biol *34*, 1307-1313.

He, W., Wang, Q., Gu, L., Zhong, L., and Liu, D. (2018). NOX4 rs11018628 polymorphism associates with a decreased risk and better short-term recovery of ischemic stroke. Exp Ther Med *16*, 5258-5264.

Hsu, J., and Smith, J.D. (2013). Genetic-genomic replication to identify candidate mouse atherosclerosis modifier genes. J Am Heart Assoc *2*, e005421.

Huang, P.L. (2009). A comprehensive definition for metabolic syndrome. Dis Model Mech *2*, 231-237.

Huda, M.N., VerHague, M., Albright, J., Smallwood, T., Bell, T.A., Que, E., Miller, D.R., Roshanravan, B., Allayee, H., Pardo-Manuel de Villena, F., *et al.* (2020). Dissecting the Genetic Architecture of Cystatin C in Diversity Outbred Mice. G3 (Bethesda).

Ivashkiv, L.B., and Donlin, L.T. (2014). Regulation of type I interferon responses. Nat Rev Immunol *14*, 36-49.

Jeong, B.Y., Lee, H.Y., Park, C.G., Kang, J., Yu, S.L., Choi, D.R., Han, S.Y., Park, M.H., Cho, S., Lee, S.Y., *et al.* (2018). Oxidative stress caused by activation of NADPH oxidase 4 promotes contrast-induced acute kidney injury. PLoS One *13*, e0191034.

Jiang, F., Lim, H.K., Morris, M.J., Prior, L., Velkoska, E., Wu, X., and Dusting, G.J. (2011). Systemic upregulation of NADPH oxidase in diet-induced obesity in rats. Redox Rep *16*, 223-229.

Joost, H.G., and Schurmann, A. (2014). The genetic basis of obesity-associated type 2 diabetes (diabesity) in polygenic mouse models. Mamm Genome *25*, 401-412.

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

Keller, M.P., Rabaglia, M.E., Schueler, K.L., Stapleton, D.S., Gatti, D.M., Vincent, M., Mitok, K.A., Wang, Z., Ishimura, T., Simonett, S.P., *et al.* (2019). Gene loci associated with insulin secretion in islets from nondiabetic mice. Journal of Clinical Investigation *129*, 4419-4432.

Kemis, J.H., Linke, V., Barrett, K.L., Boehm, F.J., Traeger, L.L., Keller, M.P., Rabaglia, M.E., Schueler, K.L., Stapleton, D.S., Gatti, D.M., *et al.* (2019). Genetic determinants of gut microbiota composition and bile acid profiles in mice. PLoS Genet *15*, e1008073.

Koeth, R.A., Wang, Z., Levison, B.S., Buffa, J.A., Org, E., Sheehy, B.T., Britt, E.B., Fu, X., Wu, Y., Li, L., *et al.* (2013). Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis. Nature Medicine *19*, 576-585.

Kohjima, M., Enjoji, M., Higuchi, N., Kato, M., Kotoh, K., Yoshimoto, T., Fujino, T., Yada, M., Yada, R., Harada, N., *et al.* (2007). Re-evaluation of fatty acid metabolism-related gene expression in nonalcoholic fatty liver disease. Int J Mol Med *20*, 351-358.

Kolde, R., Franzosa, E.A., Rahnavard, G., Hall, A.B., Vlamakis, H., Stevens, C., Daly, M.J., Xavier, R.J., and Huttenhower, C. (2018). Host genetic variation and its microbiome interactions within the Human Microbiome Project. Genome Med *10*, 6.

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

Krueger, S.K., and Williams, D.E. (2005). Mammalian flavin-containing monooxygenases: structure/function, genetic polymorphisms and role in drug metabolism. Pharmacol Ther *106*, 357-387.

Lambert, D.M., Mamer, O.A., Akerman, B.R., Choiniere, L., Gaudet, D., Hamet, P., and Treacy, E.P. (2001). In vivo variability of TMA oxidation is partially mediated by polymorphisms of the FMO3 gene. Mol Genet Metab *73*, 224-229.

Lan, X., Li, D., Zhong, B., Ren, J., Wang, X., Sun, Q., Li, Y., Liu, L., Liu, L., and Lu, S. (2015). Identification of differentially expressed genes related to metabolic syndrome induced with high-fat diet in E3 rats. Exp Biol Med (Maywood) *240*, 235-241.

Lang, D.H., Yeung, C.K., Peter, R.M., Ibarra, C., Gasser, R., Itagaki, K., Philpot, R.M., and Rettie, A.E. (1998). Isoform Specificity of Trimethylamine N-oxygenation by Human Flavin-Containing Monooxygenase (FMO) and P450 Enzymes: Selective Catalysis by FMO3. Biochem Pharmacol *56*, 1005-1012.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Lent-Schochet, D., Silva, R., McLaughlin, M., Huet, B., and Jialal, I. (2018). Changes to trimethylamine-N-oxide and its precursors in nascent metabolic syndrome. Horm Mol Biol Clin Investig *35*.

Leon-Mimila, P., Villamil-Ramirez, H., Li, X.S., Shih, D.M., Hui, S.T., Ocampo-Medina, E., Lopez-Contreras, B., Moran-Ramos, S., Olivares-Arevalo, M., Grandini-Rosales, P., *et al.* (2020). Trimethylamine N-oxide levels are associated with NASH in obese subjects with type 2 diabetes. Diabetes Metab, 101183.

Lozhkin, A., Vendrov, A.E., Pan, H., Wickline, S.A., Madamanchi, N.R., and Runge, M.S. (2017). NADPH oxidase 4 regulates vascular inflammation in aging and atherosclerosis. J Mol Cell Cardiol *102*, 10-21.

Lu, X., Li, N., Shushakova, N., Schmitt, R., Menne, J., Susnik, N., Meier, M., Leitges, M., Haller, H., Gueler, F., *et al.* (2012). C57BL/6 and 129/Sv mice: genetic difference to renal ischemia-reperfusion. J Nephrol *25*, 738-743.

Marchesini, G., Bugianesi, E., Forlani, G., Cerrelli, F., Lenzi, M., Manini, R., Natale, S., Vanni, E., Villanova, N., Melchionda, N., *et al.* (2003). Nonalcoholic fatty liver, steatohepatitis, and the metabolic syndrome. Hepatology *37*, 917-923.

Marro, B.S., Legrain, S., Ware, B.C., and Oldstone, M.B. (2019). Macrophage IFN-I signaling promotes autoreactive T cell infiltration into islets in type 1 diabetes model. JCI Insight *4*, e125067.

Matsuzawa, N., Takamura, T., Kurita, S., Misu, H., Ota, T., Ando, H., Yokoyama, M., Honda, M., Zen, Y., Nakanuma, Y., *et al.* (2007). Lipid-induced oxidative stress causes steatohepatitis in mice fed an atherogenic diet. Hepatology *46*, 1392-1403.

McGettigan, B., McMahan, R., Orlicky, D., Burchill, M., Danhorn, T., Francis, P., Cheng, L.L., Golden-Mason, L., Jakubzick, C.V., and Rosen, H.R. (2019). Dietary Lipids Differentially Shape Nonalcoholic Steatohepatitis Progression and the Transcriptome of Kupffer Cells and Infiltrating Macrophages. Hepatology *70*, 67-83.

Meng, H., Vera, I., Che, N., Wang, X., Wang, S.S., Ingram-Drake, L., Schadt, E.E., Drake, T.A., and Lusis, A.J. (2007). Identification of Abcc6 as the major causal gene for dystrophic cardiac calcification in mice through integrative genomics. Proc Natl Acad Sci U S A *104*, 4530-4535.

Mitok, K.A., Freiberger, E.C., Schueler, K.L., Rabaglia, M.E., Stapleton, D.S., Kwiecien, N.W., Malec, P.A., Hebert, A.S., Broman, A.T., Kennedy, R.T., *et al.* (2018). Islet proteomics reveals genetic variation in dopamine production resulting in altered insulin secretion. J Biol Chem *293*, 5860-5877.

Murray, T.V., Dong, X., Sawyer, G.J., Caldwell, A., Halket, J., Sherwood, R., Quaglia, A., Dew, T., Anilkumar, N., Burr, S., *et al.* (2015). NADPH oxidase 4 regulates homocysteine metabolism and protects against acetaminophen-induced liver damage in mice. Free Radic Biol Med *89*, 918-930.

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

Paik, Y.H., Kim, J., Aoyama, T., De Minicis, S., Bataller, R., and Brenner, D.A. (2014). Role of NADPH oxidases in liver fibrosis. Antioxid Redox Signal *20*, 2854-2872.

Park, H.S., Jung, H.Y., Park, E.Y., Kim, J., Lee, W.J., and Bae, Y.S. (2004). Cutting edge: direct interaction of TLR4 with NAD(P)H oxidase 4 isozyme is essential for lipopolysaccharide-induced production of reactive oxygen species and activation of NF-kappa B. J Immunol *173*, 3589-3593.

Patel, D.N., Bailey, S.R., Gresham, J.K., Schuchman, D.B., Shelhamer, J.H., Goldstein, B.J., Foxwell, B.M., Stemerman, M.B., Maranchie, J.K., Valente, A.J., *et al.* (2006). TLR4-NOX4-AP-1 signaling mediates lipopolysaccharide-induced CXCR6 expression in human aortic smooth muscle cells. Biochem Biophys Res Commun *347*, 1113-1120.

Pletscher-Frankild, S., Palleja, A., Tsafou, K., Binder, J.X., and Jensen, L.J. (2015). DISEASES: text mining and data integration of disease-gene associations. Methods *74*, 83-89.

Que, E., James, K.L., Coffey, A.R., Smallwood, T.L., Albright, J., Huda, M.N., Pomp, D., Sethupathy, P., and Bennett, B.J. (2020). Genetic Architecture Modulates Diet-Induced Hepatic mRNA and miRNA Expression Profiles in Diversity Outbred Mice. Genetics *216*, 241-259.

R Core Team (2016). R: A Language and Environment for Statistical Computing.

Rabelo, F., Stefano, J.T., Cavaleiro, A.M., Lima, R.V.C., de Campos Mazo, D.F., Carrilho, F.J., Correa-Giannella, M.L., and Oliveira, C.P. (2018). Association between the CYBA and NOX4 genes of NADPH oxidase and its relationship with metabolic syndrome in non-alcoholic fatty liver disease in Brazilian population. Hepatobiliary Pancreat Dis Int *17*, 330-335.

Remmerie, A., Martens, L., and Scott, C.L. (2020). Macrophage Subsets in Obesity, Aligning the Liver and Adipose Tissue. Front Endocrinol (Lausanne) *11*, 259.

Renaud, H.J., Cui, J.Y., Lu, H., and Klaassen, C.D. (2014). Effect of diet on expression of genes involved in lipid metabolism, oxidative stress, and inflammation in mouse liver-insights into mechanisms of hepatic steatosis. PLoS One *9*, e88584.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47.

Rudraiah, S., Rohrer, P.R., Gurevich, I., Goedken, M.J., Rasmussen, T., Hines, R.N., and Manautou, J.E. (2014). Tolerance to acetaminophen hepatotoxicity in the mouse model of autoprotection is associated with induction of flavin-containing monooxygenase-3 (FMO3) in hepatocytes. Toxicol Sci *141*, 263-277.

Schugar, R.C., Shih, D.M., Warrier, M., Helsley, R.N., Burrows, A., Ferguson, D., Brown, A.L., Gromovsky, A.D., Heine, M., Chatterjee, A., *et al.* (2017). The TMAO-Producing Enzyme Flavin-Containing Monooxygenase 3 Regulates Obesity and the Beiging of White Adipose Tissue. Cell Rep *19*, 2451-2461.

Schurmann, C., Rezende, F., Kruse, C., Yasar, Y., Lowe, O., Fork, C., van de Sluis, B., Bremer, R., Weissmann, N., Shah, A.M., *et al.* (2015). The NADPH oxidase Nox4 has anti-atherosclerotic functions. Eur Heart J *36*, 3447-3456.

Shih, D.M., Wang, Z., Lee, R., Meng, Y., Che, N., Charugundla, S., Qi, H., Wu, J., Pan, C., Brown, J.M., *et al.* (2015). Flavin containing monooxygenase 3 exerts broad effects on glucose and lipid metabolism and atherosclerosis. J Lipid Res *56*, 22-37.

Singh, A., Koduru, B., Carlisle, C., Akhter, H., Liu, R.M., Schroder, K., Brandes, R.P., and Ojcius, D.M. (2017). NADPH oxidase 4 modulates hepatic responses to lipopolysaccharide mediated by Toll-like receptor-4. Sci Rep *7*, 14346.

Siqueira, E.R., Pereira, L.B., Stefano, J.T., Patente, T., Cavaleiro, A.M., Silva Vasconcelos, L.R., Carmo, R.F., Moreira Beltrao Pereira, L.M., Carrilho, F.J., Correa-Giannella, M.L., *et al.* (2015). Association of a variant in the regulatory region of NADPH oxidase 4 gene and metabolic syndrome in patients with chronic hepatitis C. Eur J Med Res *20*, 45.

Smallwood, T.L., Gatti, D.M., Quizon, P., Weinstock, G.M., Jung, K.C., Zhao, L., Hua, K., Pomp, D., and Bennett, B.J. (2014). High-resolution genetic mapping in the diversity outbred mouse population identifies Apobec1 as a candidate gene for atherosclerosis. G3 (Bethesda) *4*, 2353-2363.

Somers, E.C., Zhao, W., Lewis, E.E., Wang, L., Wing, J.J., Sundaram, B., Kazerooni, E.A., McCune, W.J., and Kaplan, M.J. (2012). Type I interferons are associated with subclinical markers of cardiovascular disease in a cohort of systemic lupus erythematosus patients. PLoS One *7*, e37000.

Stubbs, J.R., House, J.A., Ocque, A.J., Zhang, S., Johnson, C., Kimber, C., Schmidt, K., Gupta, A., Wetmore, J.B., Nolin, T.D., *et al.* (2016). Serum Trimethylamine-N-Oxide is Elevated in CKD and Correlates with Coronary Atherosclerosis Burden. J Am Soc Nephrol *27*, 305-313.

Tyler, A.L., Ji, B., Gatti, D.M., Munger, S.C., Churchill, G.A., Svenson, K.L., and Carter, G.W. (2017). Epistatic Networks Jointly Influence Phenotypes Related to Metabolic Disease and Gene Expression in Diversity Outbred Mice. Genetics *206*, 621-639.

Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M., and Ng, P.C. (2016). SIFT missense predictions for genomes. Nat Protoc *11*, 1-9.

Wallace, T.M., Levy, J.C., and Matthews, D.R. (2004). Use and Abuse of HOMA Modeling. Diabetes Care *27*, 1487-1495.

Wang, X.A., Zhang, R., Zhang, S., Deng, S., Jiang, D., Zhong, J., Yang, L., Wang, T., Hong, S., Guo, S., *et al.* (2013). Interferon regulatory factor 7 deficiency prevents diet-induced obesity and insulin resistance. Am J Physiol Endocrinol Metab *305*, E485-495.

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

Warrier, M., Shih, D.M., Burrows, A.C., Ferguson, D., Gromovsky, A.D., Brown, A.L., Marshall, S., McDaniel, A., Schugar, R.C., Wang, Z., *et al.* (2015). The TMAO-Generating Enzyme Flavin Monooxygenase 3 Is a Central Regulator of Cholesterol Balance. Cell Rep *10*, 326-338.

Wieser, V., Adolph, T.E., Grander, C., Grabherr, F., Enrich, B., Moser, P., Moschen, A.R., Kaser, S., and Tilg, H. (2018). Adipose type I interferon signalling protects against metabolic dysfunction. Gut *67*, 157-165.

Winter, J.M., Gildea, D.E., Andreas, J.P., Gatti, D.M., Williams, K.A., Lee, M., Hu, Y., Zhang, S., Program, N.C.S., Mullikin, J.C., *et al.* (2017). Mapping Complex Traits in a Diversity Outbred F1 Mouse Population Identifies Germline Modifiers of Metastasis in Human Prostate Cancer. Cell Syst *4*, 31-45 e36.

Yoon, G., Cho, K.A., Song, J., and Kim, Y.K. (2019). Transcriptomic Analysis of High Fat Diet Fed Mouse Brain Cortex. Front Genet *10*, 83.

Zhai, Y., Qiao, B., Gao, F., Shen, X., Vardanian, A., Busuttil, R.W., and Kupiec-Weglinski, J.W. (2008). Type I, but not type II, interferon is critical in liver injury induced after ischemia and reperfusion. Hepatology *47*, 199-206.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol *4*, Article17.

Zhang, J., Chaluvadi, M.R., Reddy, R., Motika, M.S., Richardson, T.A., Cashman, J.R., and Morgan, E.T. (2009). Hepatic flavin-containing monooxygenase gene regulation in different mouse inflammation models. Drug Metab Dispos *37*, 462-468.

Zhang, J., Tran, Q., Lattard, V., and Cashman, J.R. (2003). Deleterious mutations in the flavin-containing monooxygenase 3 (FMO3) gene causing trimethylaminuria. Pharmacogenetics *13*, 495-500.

Zhang, W., Bu, S.Y., Mashek, M.T., I, O.S., Sibai, Z., Khan, S.A., Ilkayeva, O., Newgard, C.B., Mashek, D.G., and Unterman, T.G. (2016). Integrated Regulation of Hepatic Lipid and Glucose Metabolism by Adipose Triacylglycerol Lipase and FoxO Proteins. Cell Rep *15*, 349-359.

Zhou, Z., Zhao, X., Chen, L., Li, Y., Chen, Z., Wang, Y., Zhou, Z., and Chu, X. (2020). Integrated analysis of differentially expressed long noncoding RNAs and mRNAs associated with high-fat diet-induced hepatic insulin resistance in mice. Nutr Metab (Lond) *17*, 45.

CHAPTER 4.

Sexual Dimorphism of Atherosclerosis-associated Liver Transcriptome and Genetic Determinants in Hyperlipidemic Diversity Outbred-F1 Mice

4.1. Author Contributions

This manuscript in under review in the Cardiovascular Research journal. B.J.B designed all the experiments. M.K., M.N.H., and E.R. performed the experiments. M.K., M.N.H., and E.Q. analyzed raw data. M.K. and B.J.B. wrote the manuscript, which was reviewed by all authors.

4.2. Abstract

Sexual dimorphism in the incidence and complications of atherosclerosis is well known in human and rodent models; however, the underlying mechanisms by which sex as a biological variable affects atherosclerosis remains unclear. In order to examine the role of sex and gene-by-sex interactions affecting atherosclerosis and atherosclerosis-related traits, we generated a high-resolution genetic panel of mice from a cross between Diversity Outbred (DO) mice and a hyperlipidemic strain. We generated a high resolution genetic panel of mice susceptible to atherosclerosis by crossing atherosclerosis-susceptible male C57BL/6J mice, transgenic for both human apolipoprotein E-Leiden and cholesterol ester transfer protein genes, with female DO mice. We examined atherosclerosis, cardiometabolic traits and the liver transcriptome after 16 weeks of a high-fat/cholesterol diet. Our results demonstrate the tremendous effects of sex on cardiometabolic traits and hepatic gene expression with genetic loci associated with traits and transcripts frequently showing sex specificity. We revealed sex-specific candidate genes that were mapped to the quantitative trait loci for aortic lesion area and whose expression was locally regulated via global liver transcriptome. Several sexually dimorphic

transcripts including *Pten* are identified as candidates for sex-specific QTL for atherosclerosis. Finally, global analysis of gene expression identified a sex-specific regulation of the liver transcription factor LXRα (*Nr1h3*) which affected expression of target genes and gene-trait correlations.

Collectively, this study provides a rich resource to investigate the sex-differentiated pathogenesis of atherosclerosis, and shows that DO mice in conjunction with integrative genetics approaches can identify genes and genetic variants that contribute to atherosclerosis.

4.3. Introduction

Despite advanced innovations in medical and cardiovascular treatment, cardiovascular disease (CVD) remains the leading cause of death in men and women in most ethnic groups (Man et al., 2020). There is a complicated association between biological sex and CVD risk (Rodgers et al., 2019). Men are at increased risk of CVD as compared to women, however risk in women dramatically changes with the onset of menopause (Maas and Appelman, 2010). Ischemic CVD, including myocardial infarction and stroke, are often caused by atherosclerosis, a process by which lipid-laden plaques are formed in the vasculature.

Sex differences in CVD susceptibility and sexual dimorphism in atherosclerosis have been reported in humans and animals (Bubb et al., 2012; Isensee et al., 2008; Yang et al., 2006). Atherosclerosis in women is also closely associated with hypertension and autoimmune diseases such as systemic lupus erythematosus and rheumatoid arthritis (Fairweather et al., 2012), suggesting that the immune mechanisms that cause atherosclerosis in men may be different from those in women. Sex hormones alter the immune response in the key processes of atherosclerosis (Fairweather, 2014) and play an important role in regulating gene expression related to metabolic disorders including atherosclerosis (AlSiraj et al., 2019; Kukurba et al., 2016). Besides sex hormones, sexual dimorphism also arises due to

effects of sex chromosomes (epigenetic effects, genetic dosage, and different dosage compensation) (Charlesworth, 1996; Disteche, 2012; Hager et al., 2008; Winham et al., 2015). In addition, genotype-by-sex interactions can also affect atherosclerosis susceptibility (Wang et al., 2007). However, the exact mechanisms responsible for the sexually dimorphic regulation of atherosclerosis susceptibility remain incomplete.

Understanding how sex-biased gene expression contributes to the sexual dimorphism of cardiometabolic traits is critical to understanding how biological sex affects susceptibility to CVD (Parsch and Ellegren, 2013). Global hepatic gene expression studies in rodent models have identified more than 1,000 sex-biased transcripts, which collectively impact liver metabolism, inflammatory responses, and disease susceptibility (Roy and Chatterjee, 1983; Yang et al., 2006; Zhang et al., 2011). However, to date, less is known about gene-by-sex interactions on atherosclerosis. Several studies show the potential to explain sex differences associated with CVD risk in human liver pathophysiology, including sex differences in circulating lipid profiles (Blum and Blum, 2009; Nedungadi and Clegg, 2009).

In this study, we investigated the genetic regulation of atherosclerosis and known risk factors, in Diversity Outbred F1 (DO-F1) which were generated from a cross of Diversity Outbred (DO) females with male inbred C57BL/6J mice harboring transgenes for two hyperlipidemia-inducing mutations: human cholesteryl ester transfer protein (CETP) and apolipoprotein E-Leiden (APOE-Leiden). The CETP transgene reduces the concentration of HDL, and the APOE-Leiden transgene reduces the clearance of triglyceride-rich lipoproteins (Westerterp et al., 2006). Thus as compared to wild-type mice, the APOE-Leiden and CETP transgenic mice display a lipoprotein cholesterol profile similar to humans (Westerterp et al., 2006). These DO-F1 mice are highly diverse and contain a random assortment of DNA from eight founder strains: A/J (A/J), C57BL6/J (B6), 1291/SvImJ (129), NOD/ShiLtJ (NOD), NZO/HiLtJ (NZO), CAST/EiJ (CAST), PWK/PhJ (PWK), and WSB/EiJ (WSB) (Churchill et

al., 2012). Thus, the DO-F1 mice have high phenotypic variability due to genetic heterogeneity (Vorobyev et al., 2019).

To the best of our knowledge, this study is reporting the first example of sexual dimorphism and quantitative trait loci (QTL) mapping of atherosclerosis in DO-F1 mice. We demonstrated sex differences in cardiometabolic traits, gene expression levels, and genetic effects of cardiometabolic traits and gene expression (eQTLs). By incorporating aortic lesion area QTLs with eQTLs, we report multiple sex-specific genetic effects on atherosclerosis that colocalize with *cis*-acting eQTLs. Finally, with sex-specific gene expression and genetic regulation of transcription factors expressed in the liver, this study emphasizes the importance of understanding sex as a biological variable at the molecular level.

4.4. Methods

4.4.1. Ethics Statement

We followed all NIH animal welfare guidelines and the animal care and study protocols were approved by the University of California Davis Animal Care and Use Committee.

4.4.2. Animals: Hyperlipidemic Eight DO Founder Strains-F1 Mice and DO-F1 Mice

Animal care and study protocols were approved by the University of California Davis Animal Care and Use Committee. 6-week-old CETP/ApoE3 Leiden males, hemizygous to the CETP and ApoE3 transgenes (Tg), were kindly provided by Dr. Lusis (Bennett et al., 2015). Three females from each of the eight DO founder inbred strains (five classical laboratory inbred strains [A/J, B6, 129, NOD and NZO] and three wild-derived inbred strains [CAST, PWK, and WSB]) and DO females (JAX stock number 009376, outbreeding generation # 26,28) were obtained from The Jackson Laboratory (Bar Harbor, ME) and maintained in the Mouse Biology Program vivarium at the University of California Davis. For the strain survey experiment, we crossed CETP/ApoE3 Leiden males to one of eight strains to generate eight

different F1 strains of mice (AJ-F1, B6-F1, 129-F1, NOD-F1, NZO-F1, CAST-F1, PWK-F1, and WSB-F1, respectively) and quantified atherosclerotic traits in F1 female and male mice. In addition, a total of 200 F0 J:DO females were crossed with CETP/ApoE3 Leiden males to breed 238 (CETP/ApoE3 Leiden × J:DO) F1 females and 234 (CETP/ApoE3 Leiden × J:DO) F1 males (**Figure 4.1**). Female and male progeny were genotyped to confirm the presence of CETP and ApoE3-Leiden transgenes and maintained on a synthetic diet, AIN-76A (D10001, Research Diets, New Brunswick, NJ) until 8 weeks of age. At the age of about 8 weeks, all mice were fed with a synthetic high-fat and high-cholesterol (33 kcal % fat from cocoa butter and 1.25% cholesterol) diet (Research Diets D121083) (see **Table 4.1**) *ad libitum*. Mice were euthanized for tissue collection after fed this diet for 16 weeks. Animals were maintained on a 12 h light and dark cycle under temperature- and humidity-controlled conditions. Euthanasia of all mice was performed by cervical dislocation after anesthesia with isoflurane. For eight different F1 strains of mice and (CETP/ApoE3 Leiden x J:DO) F1 mice, aorta, liver, subcutaneous fat, gonadal fat, cecum, and spleen were collected upon euthanasia.

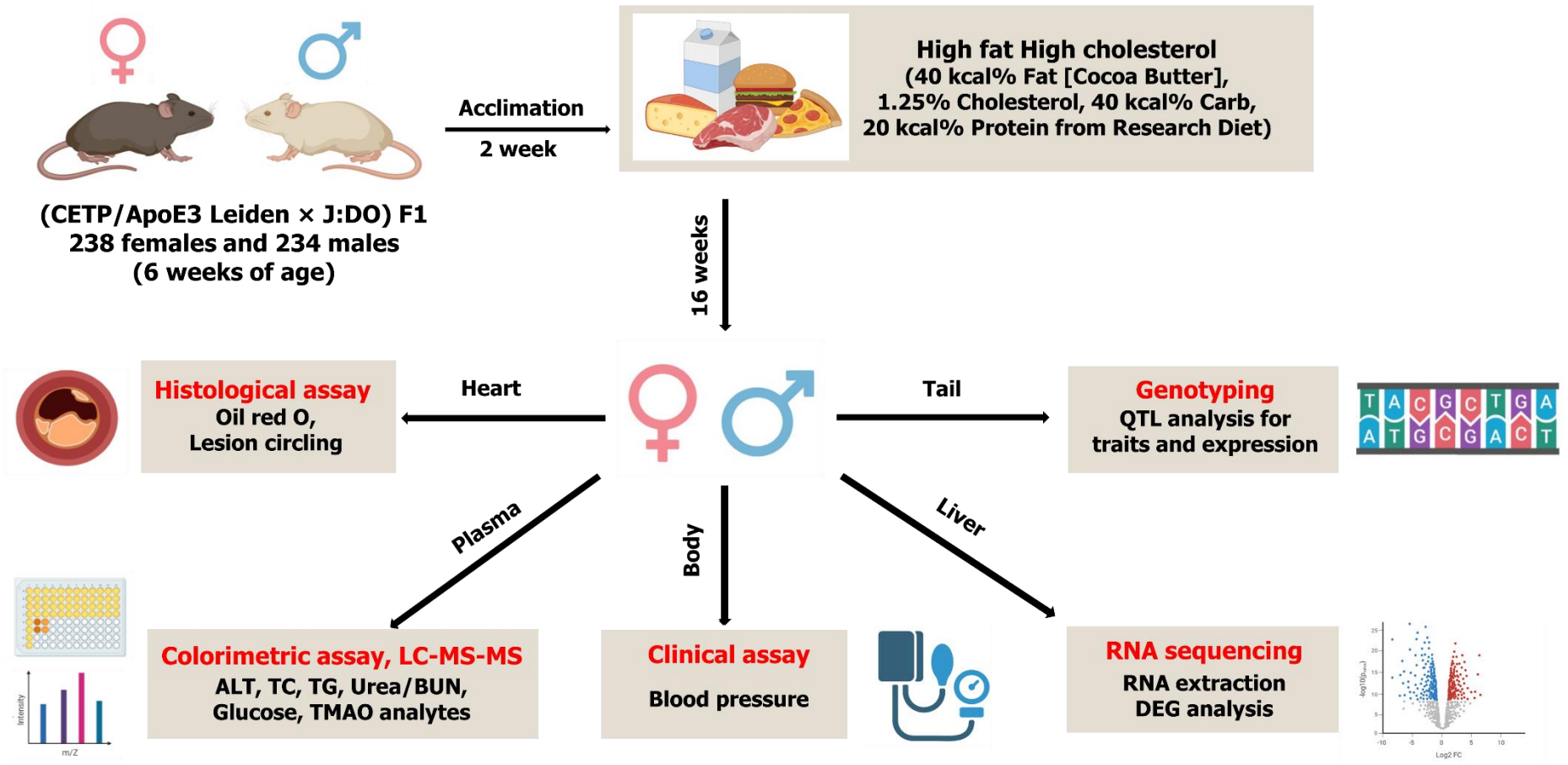


Figure 4.1. Study design and timeline for Diversity Outbred (DO)-F1 mice.
A total of 200 F0 J:DO females (JAX stock number 009376, outbreeding generation # 26,28) were crossed with CETP/ApoE3 Leiden males to breed 238 (CETP/ApoE3 Leiden × J:DO) F1 females and 234 (CETP/ApoE3 Leiden × J:DO) F1 males. Female and male progeny were genotyped to confirm the presence of CETP and ApoE3-Leiden transgenes and maintained on a synthetic diet, AIN-76A until 8 weeks of age. At the age of about 8 weeks, all mice were fed with a synthetic high-fat and high-cholesterol (33 kcal % fat from cocoa butter and 1.25% cholesterol) diet *ad libitum*. Mice were euthanized for tissue collection after fed this diet for 16 weeks.

Table 4.1. Nutrient constituents in AIN76 and high-fat and high-cholesterol (HFHC) diet.

Class description	Ingredient	AIN76 (Grams)	HFHC (Grams)
Protein	Casein, Lactic, 30 Mesh	200.00 g	200.00 g
Protein	Cystine, L	3.00 g	3.00 g
Carbohydrate	Starch, Corn	150.00 g	212.00 g
Carbohydrate	Sucrose, Fine Granulated	500.00 g	124.41 g
Carbohydrate	Lodex 10 (Maltodextrin)	0.00 g	71.00 g
Fiber	Solka Floc (Cellulose)	50.00 g	50.00 g
Fat	Cocoa Butter, Deodorized	0.00 g	155.00 g
Fat	Soybean Oil	50.00 g	25.00 g
Mineral	Potassium Citrate, Monohydrate		16.50 g
Mineral	Calcium Phosphate, Dibasic		13.00 g
Mineral	Calcium Carbonate, Light	37.00 g	5.50 g
Mineral	Mineral mixture		5.00 g
Mineral	Sodium Chloride		2.59 g
Vitamin	Choline Bitartrate	2.00 g	2.00 g
Vitamin	Vitamin mixture	10.00 g	1.00 g
Special	Cholesterol	0.0 g	11.25 g
Special	Sodium Cholate	0.0 g	0.0 g
	Total:	**1002.0 g**	**897.35 g**

The two diets used in this study were manufactured by Research Diets. AIN-76 was fed to the two studies population (Eight DO founder strains F1 mice, and DO-F1 mice) from 6-8 weeks of age in order to ensure that there were no spurious effects due to the potential variable composition of standard laboratory chow. DO founder strains F1 mice and DO-F1 mice were fed high-fat and high-cholesterol (HFHC) diet for 16 weeks from 8-24 weeks of age. HFHC diet is considered atherogenic and were intended to induce the formation of atherosclerosis in the DO founder strains or DO-F1 mice.

4.4.3. Plasma Collection

Mice at 24 weeks of age fasted at 4 h and plasma was collected from the retro-orbital plexus under isoflurane anesthesia immediately before euthanasia of mice. Blood was collected using heparinized glass capillaries into plasma collection tubes with EDTA (Becton Dickinson, Franklin Lakes, NJ). Blood was kept on ice and the separated plasmas by centrifugation were frozen at -80°C in aliquots prior to analysis. 4 h fasting plasma samples were used for measurements of ALB, ALT, TC, HDL-C, TG, glucose, urea nitrogen, TMAO, choline, betaine, and carnitine. In the eight different F1 strains of mice, only TC and TG were measured in plasma samples.

4.4.4. Plasma Clinical Cardiometabolic Traits

Plasma levels of albumin, ALT, TC, HDL-C, TG, glucose, urea nitrogen were quantitated at 24 weeks of age using a COBAS INTEGRA 400 plus Analyzer (Roche Diagnostics, Indianapolis, IN, USA). 50 μl plasma was diluted 3 times with 1 × phosphate-buffered saline (PBS) and processed using standard procedures according to the analyzer's instructions. HDL-C was subtracted from TC to determine VLDL-C/LDL-C levels.

4.4.5. Plasma Metabolite Analysis using LC-MS/MS

At 24 weeks of age, circulating plasma analytes such as TMAO, choline, betaine, creatinine, and carnitine were quantified using liquid chromatography–mass spectrometry (LC-MS) (Wang et al., 2014) with modifications. Undeuterated analytes ranging from 0-100 μM in methanol were analyzed to establish analyte standard curves. All standards were purchased from Sigma-Aldrich (St. Louis, MO) and all reagent solvents of mass spectrometry grade were purchased from Fisher Scientific (Waltham, MA).

Samples (20 μl plasma) were aliquoted into Eppendorf tubes and mixed with 80 μl of 10 μM surrogate standard. The samples were vortexed for 30 seconds and centrifuged at 18,000 g at 10°C for 10 min. The supernatants and surrogate standards were transferred to

150 µl glass inserts in high-performance liquid chromatography (HPLC) vials. The supernatants (5 µl) were injected into a silica column (150 by 2 mm, 3 µm silica; Cat. No. 00F-4162-B0, Phenomenex, Torrance, CA) at a flow rate of 0.8 ml/min using a Waters Acquity UPLC (Waters, Milford, MA) interfaced with an API 4000 Q-TRAP mass spectrometer (AB SCIEX, Framingham, MA). Analytes and isotope-labeled internal standards were monitored in multiple reaction monitoring (MRM) mode using characteristic precursor–product ion transitions. Parameters for the ion monitoring were optimized for each analyte. Calibration curves for quantification of the analytes were prepared by spiking the analytes of various concentrations into control plasma samples.

4.4.6. Measurement of Blood Pressure in Tail

CODA mouse tail-cuff system (Kent Scientific Corp, Torrington, CT) was used for measurement of mouse blood pressure in tail and this technique was verified by telemetry with a 99% correlation (Feng et al., 2008). In short, blood pressure was measured by placing a cuff on the tail of the mouse to block blood flow and placing a non-invasive blood pressure sensor over the occlusive cuff. The mice were first acclimated to restraints for 10-20 min/d for at least 3 days, and then 20 cycles of blood pressure measurements were performed. Multiple measurements of systolic blood pressure, diastolic blood pressure, heart rate (beat per minute), cardiac output (µl/min), and tail blood volume (µl) were performed and then averaged. Heating pad warmed the blood pressure measuring area and maintained a quiet and dark environment to ensure reliable measurements within the parameters of this technology.

4.4.7. Atherosclerotic Lesion Size

Hearts containing the proximal aorta were dissected in 24-week old mice, perfused with 1×PBS, and stored in 10% formalin at 4°C. The upper part of the heart was removed by a transverse cut parallel to the atria, which were then embedded in the optimal cutting

temperature compound and stored at -80°C. Consecutive sections (10 μm thick) from the aortic sinus were mounted on slides by the UNC Histology Research Core Facility. Sections were placed on 8-10 pieces of slides. Frozen sections (10 μm) were stained with Oil Red O by the UNC Histology Core Facility and lesion area was quantified from every third section to the proximal aorta. Stained slides were imaged using Zeiss AxioCam MR3 (Zeiss, Munich, Germany), and aortic lesion areas were quantified using Aperio's ImageScope (Vista, CA). Data were presented as the mean lesion area in μm^2.

4.4.8. RNA Library Preparation and Sequencing

Total RNA was extracted from snap-frozen liver using miRVana total RNA isolation kit (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol. The quality and amount of liver RNA were evaluated using a Bioanalyzer (Agilent, Inc., Santa Clara, CA). The average RNA-integrity score for 163 DO-F1 liver samples was 9.01 ± 0.4. RNA samples from 85 females and 77 males were submitted to the UC Davis DNA Technologies Core at the Genome Center. The RNA-seq libraries were constructed from 1 μg total RNA after poly-A library preparation. To minimize technical variability, all samples were assigned to each lane and the pooled libraries were sequenced on two lanes of the Illumina NovaSeq 6000 sequencing (Illumina Inc., San Diego, CA, USA) to achieve paired-end reads of at least 25 million 150 bp. Raw data were deposited at National Center for Biotechnology Information's Gene Expression Omnibus (GEO accession GSE179091).

4.4.9. Mouse Genotyping and Haplotype Reconstruction

Genotyping was performed in tail biopsies using the Mouse Universal Genotyping Array (GigaMUGA, 143,259 markers) by Neogen (Lincoln, NE) (Morgan et al., 2015). Identified genotypes were converted to founder strain–haplotype reconstructions using R/QTL2 software (Gatti et al., 2014). GigaMUGA markers were interpolated to an evenly spaced grid

at 0.02-cM intervals, and markers were added to fill the areas that physically represent the sparse.

4.4.10. RNA-Seq Mapping and Quantification

Raw read data was filtered using HTStream (version 1.1.0, https://github.com/ibest/HTStream), which included screening for contaminants (such as PhiX and rRNA), PCR deduplication readout, quality-based trimming, adapter trimming, and overlapping paired-end reads. Genetic variants such as SNPs and insertions/deletions (Sanger REL-1410) in eight founder strains were incorporated into the B6 reference strain genome (GRcm38/mm10) to generate individualized genomes. After obtaining transcript sequences of all annotated genes in each 8 founder strain genome, the allele sequences for each transcript were incorporated into one pooled transcriptome for read alignment. The pooled transcriptome was aligned to the individualized genomes of each strain, and then the expected read counts obtained from each transcript allele were quantified using an expectation maximization algorithm (EMASE, https://github.com/churchill-lab/emase) (Raghupathy et al., 2018). We filtered in 13,094 genes (381 X-linked, 7 Y-linked, 15 mitochondrial genes, and 12,691 autosomal genes) with mean TPM greater than 1 in 162 liver samples. This TPM filter was used to remove genes that were only expressed at low levels. We normalized the filtered genes by the upper quartile value to account for differences in library size and transformed them to rank normal scores using the 'rankZ' function in the DOQTL R package (Gatti et al., 2014) for the eQTL analysis.

4.4.11. Quantitative Trait Loci Mapping for Aortic Lesion Area and Transcripts

QTL mapping was analyzed using the R (v3.5.3) package R/QTL2 (v0.20). Marker genotypes from a GeneSeek final report were filtered for GigaMUGA-containing single nucleotide polymorphism (SNPs) and encoded into a hetero/homozygous set (e.g. AA, AB) (https://kbroman.org/qtl2/). These genotype codes were processed with GigaMUGA genotype

codes of the DO founder strains using a Hidden Markov model and haplotypes were defined as previously reported (Broman, 2012a, b). We estimated the posterior probability to have one of the eight possible genotype states at each SNP. We performed association mapping by fitting a linear mixed model at each SNP, where we regressed the traits on the diplotype probabilities. A kinship matrix was incorporated as a random effects term to adjust for the genetic relatedness between mice. Genotype probabilities were reduced to eight founder allele probabilities and were used to produce a kinship matrix (genetic relatedness) using the "leave-one-chromosome-out" method to minimize bias from same chromosome SNPs (Yang et al., 2014).

Genome scans for cardiometabolic traits were performed using three different models using the scan1 function in R/qtl2: 1) sex additive - sex and generation number were included as additive covariates, 2) female mice - generation number was included as an additive covariate, and 3) male mice - generation number was included as an additive covariate. Reported mapping statistics were logarithms of odds ratios (LODs) which describe the difference of log-scaled likelihood between full and null models. Confidence intervals for QTL were calculated as 95% Bayesian credible intervals (Sen and Churchill, 2001). Using a linear mixed model with allele probabilities as random effects, the association between the trait and each founder strain genotype in each QTL was determined using Best Linear Unbiased Predictor (BLUP). SNPs in the candidate genes found in QTL were identified based on the Wellcome Trust Sanger mouse genomes database (www.sanger.ac.uk), release 1303, based on genome assembly GRCm38 (Yalcin et al., 2011).

Permutation analysis was used for subsequent filtering to obtain significant QTL results. The significance threshold at $P<0.05$, the highly suggestive threshold at $P<0.1$, and the suggestive threshold at $P<0.63$ of all reported QTLs were empirically determined by permutation analysis, where rows of the genotyping data were randomized for each trait and a

maximum LOD score was generated (Doerge and Churchill, 1996). In this study, cardio-metabolic trait QTL (cQTL) and eQTL were filtered out at the 95th quantile of the 1,000 times null distributions. For cQTL mapping, traits with a Shapiro-Wilk W value ≥ 0.95 were considered as normalized data. Non-normal phenotypic traits were log2 transformed and the aortic lesion area was transformed to rank normal scores using the 'rankZ' function in the DOQTL R package. eQTL was defined as *cis*-eQTL when the SNP with the maximum LOD score was within ±4 Mb at the transcription start site, and *trans*-eQTL was defined when this condition was not met.

4.4.12. Heritability

To determine the extent to which phenotypic variation is affected by genotypic variation, a linear mixed-effect model was used to estimate the narrow-sense heritability scores of the cardio-metabolic traits and liver transcriptome. This was performed using the function est_herit in R/qtl2 by submitting a kinship matrix and each trait value.

4.4.13. Differential Expression Analysis

Sex-specific DEGs analysis was performed using the R package 'limma' version 3.6.1 (Ritchie et al., 2015) from TMM (trimmed mean of M values) normalized log2 transformed TPM values. Genes with BH-adjusted p values less than 0.05 were classified as DEGs. 369 genes (2.8%) from contrasts (|log2 fold change| > 3) were included in **Table 4.4** and a stringent threshold was used for the filtering to visualize a Volcano plot in **Figure 4.6C** (|log2 fold change| > 2 and –log adjusted p-value > 14).

We also used GTEx v8 data (Consortium, 2020) consisting of liver RNA-seq samples from 226 post-mortem donors to compare the extent to which sex-specific DEGs identified in this study overlap with sex-specific DEGs in human liver tissue. We filtered 286 human genes that were also identified DO-F1 liver transcripts from the top 500 sex-specific DEGs

(adjusted p-value <0.01) published using GTEx v8 data (Oliva et al., 2020), and compared overlaps between significant DO-F1 sex-specific DEGs and 286 genes.

4.4.14. Enrichment Analysis

Enrichment analyses for sex-specific DEGs, modules, and genes that were significantly correlated with aortic lesion area in each sex were performed using enrichR (Chen et al., 2013) to generate enrichment categories from the GO Biological Process 2018, KEGG 2019 Mouse, and Jenson Diseases (Pletscher-Frankild et al., 2015). This analysis identifies differential enrichment in the functional categories of transcripts. The GO Biological process 2018 contains 5,103 terms and 14,433 genes. While it is clear that individual GO terms can be found in related classes of ontology, GO terms do not occupy strictly fixed levels in a hierarchy. Each GO term identified is associated with a unique GO annotation number that relates to a specific function. Both the Gene Ontology website (http://geneontology.org/docs/faq/) and enrichR tool do not utilize a specific hierarchy thus all available terms are used in the analysis.

4.4.15. Other Statistical Analysis

All statistical analyses were performed in R (v.3.5.3) (R Core Team). Sex differences or genotype differences were assessed using Wilcoxon signed-rank test. Spearman's correlation was used to correlate the cardiometabolic traits and liver transcripts (log2TPM). The p-values were adjusted using the BH false discovery rate (FDR) procedure (Benjamini and Hochberg, 1995), and correlation coefficients and adjusted p-value were visualized using the ‘pheatmap’ package (Kolde et al., 2018). Significance was determined with a p-value < 0.05. Summary statistics were calculated to evaluate the magnitude of variability of the cardiometabolic traits by sex.

4.5. Results

4.5.1. Sexual Dimorphism on Atherosclerosis and Cardiometabolic Traits

We fed male and female DO-F1 mice a HFHC diet for 16 weeks to determine the genetic regulation of atherosclerosis and cardiometabolic risk factors including blood pressure and plasma lipids (**Figure 4.1**). First, we identified a significant effect of sex on atherosclerosis formation in the aortic root with female mice having increased atherosclerotic lesion area ($P \leq 2\times10^{-16}$) than males. The mean aortic lesion areas were 91,638 ± 83,287 μm^2 and 22,211 ± 30,113 μm^2, respectively in females and males (**Figure 4.2**). These results are consistent with eight DO founder strains-F1 mice (**Figure 4.3 and Table 4.2**; $P \leq 0.001$).

Next, we determined the sexual dimorphism in risk factors associated with atherosclerosis development. Males had increased plasma glucose (**Figure 4.4**) concentration compared to females (**Table 4.3**; $P \leq 0.001$) while plasma lipids and TMAO related analytes (TMAO, choline, and betaine) (**Figure 4.4**) were significantly higher in females than males (**Table 4.3**; $P \leq 0.001$). We note that not all traits were sexually dimorphic as plasma ALT, HDL-C, diastolic blood pressure, and systolic blood pressure were not different between sexes (**Table 4.3**).

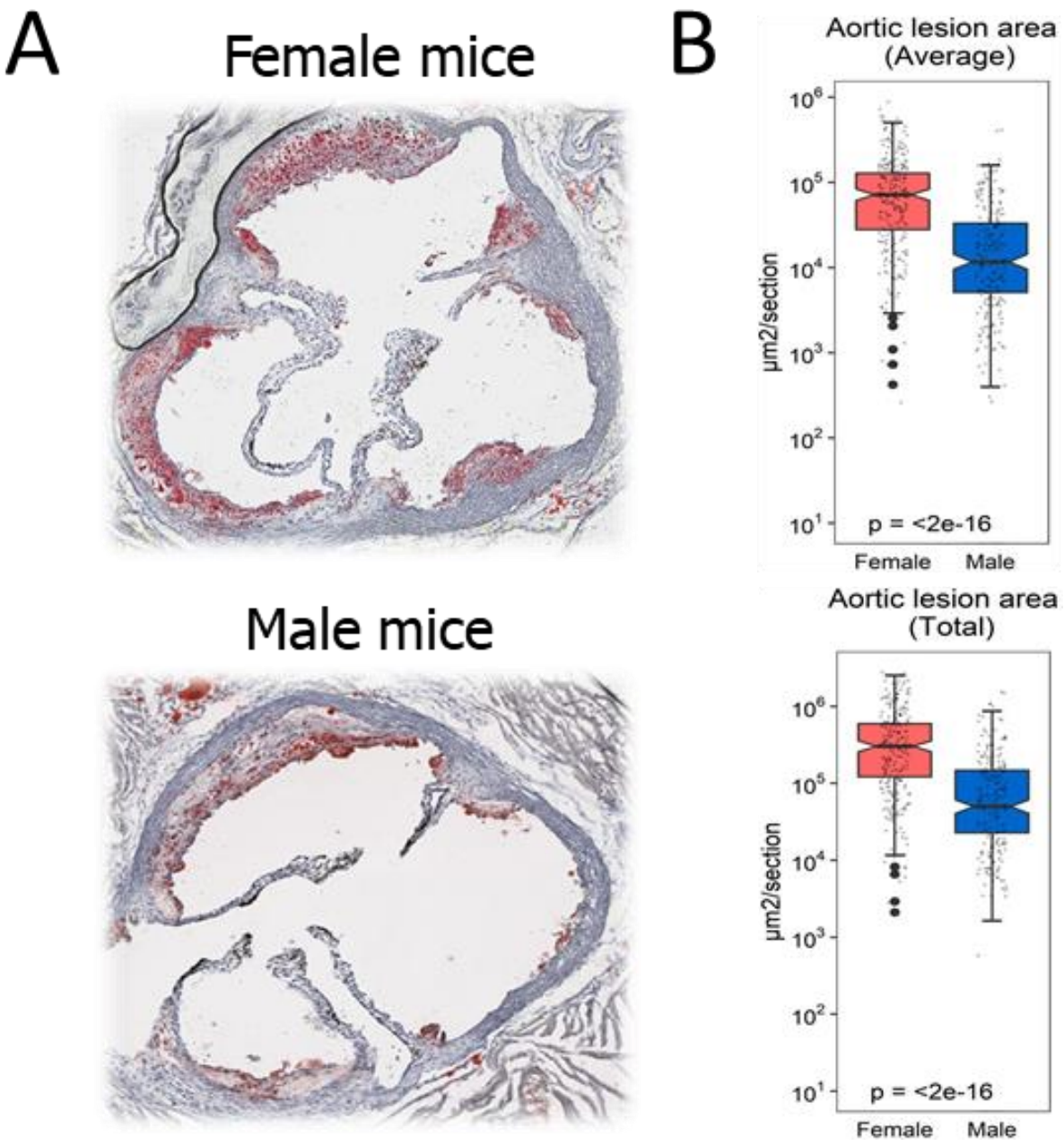


Figure 4.2. Sex differences in aortic lesion area in Diversity Outbred (DO)-F1 mice.
A total of 200 F0 J:DO females were crossed with CETP / ApoE3 Leiden males to breed 238 (CETP/ApoE3 Leiden × J:DO) F1 females and 234 (CETP/ApoE3 Leiden × J:DO) F1 males. (A) Oil red O staining of representative females and males at the 24 weeks. (B) Aortic lesion areas in females were significantly larger than males (mean ± SD). The black larget points are outliers.

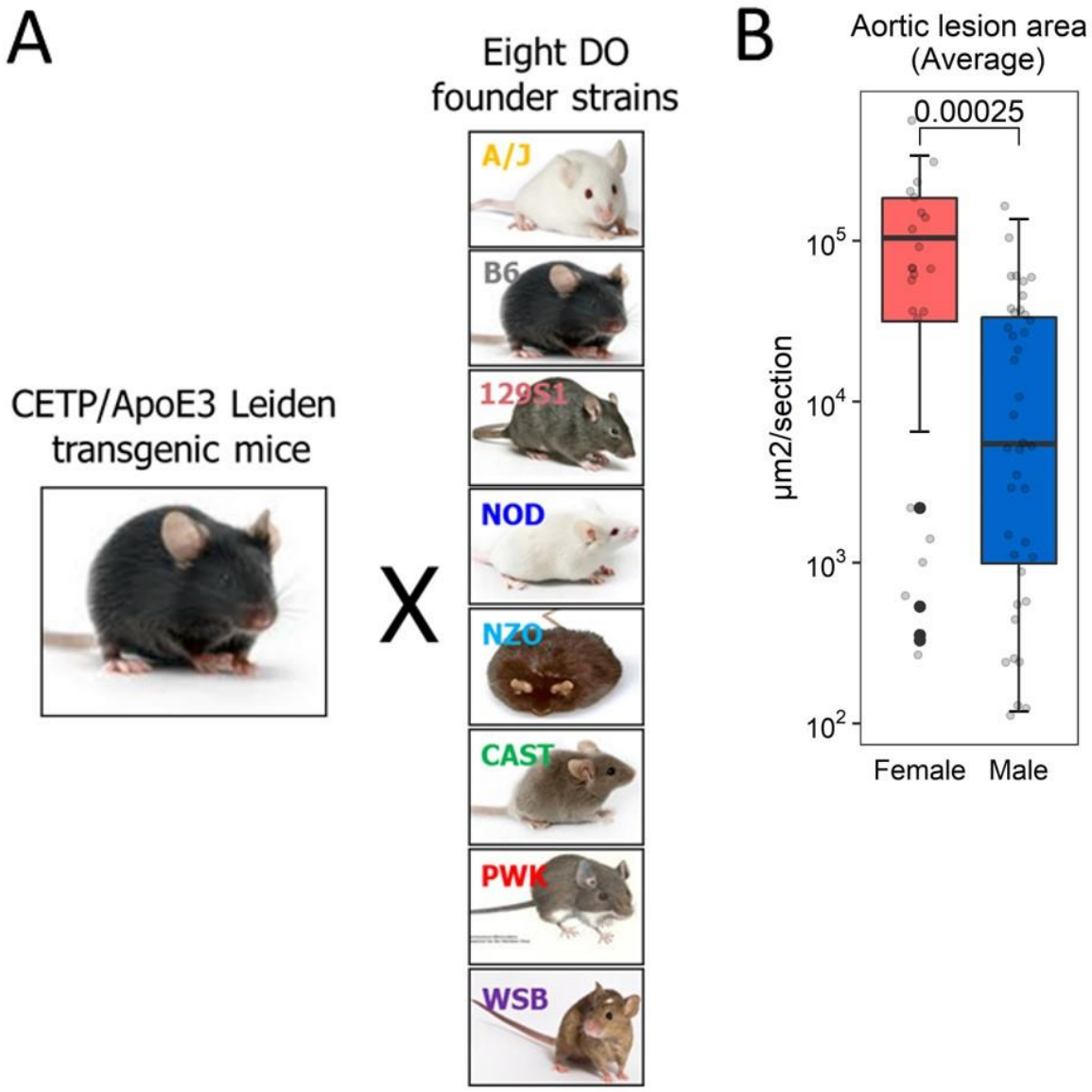


Figure 4.3. Sex differences in cardiometabolic traits in eight DO founder F1 mice. Female eight DO founder strains were crossed with CETP / ApoE3 Leiden males to breed 102 (CETP/ApoE3 Leiden × eight DO founder strains) F1 mice. At the age of about 8 weeks, all mice were fed with a synthetic high-fat and high-cholesterol (HFHC) diet for 16 weeks. (A) Schematic illustration showing breed of (CETP/ApoE3 Leiden × eight DO founder strains) F1 females and (CETP/ApoE3 Leiden × eight DO founder strains) F1 males. (B) Aortic lesion area is higher in females in eight DO founder strains-F1 mice. The p-values were Wilcoxon signed-rank test for aortic lesion area. The black larget points are outliers.

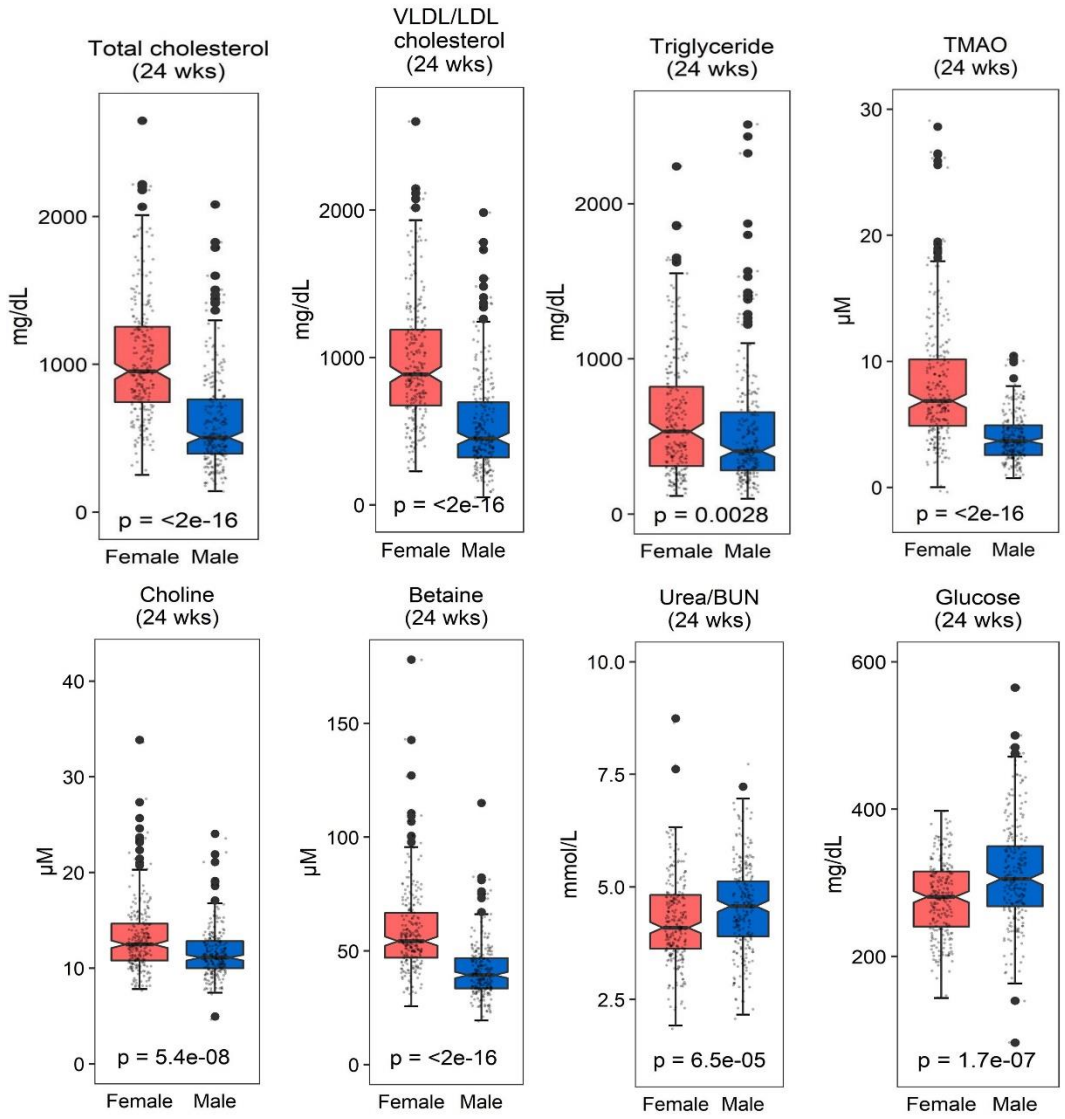


Figure 4.4. Sex differences in cardiometabolic traits in DO-F1 mice.
Plasma total cholesterol (TC), very low-density lipoprotein cholesterol/low-density lipoprotein cholesterol (VLDL-C/LDL-C), triglyceride (TG), trimethylamine (TMA) N-oxide (TMAO), choline, and betaine levels were higher in females and plasma urea nitrogen and glucose level were higher in males at 24 weeks. The p-values were Wilcoxon signed-rank test between sexes for each cardio-metabolic trait. The black larget points are outliers.

Table 4.2. Sex differences in cardiometabolic traits in DO founder strains-F1 mice at 24 weeks. n = 102 (37 females and 65 males).

Category	Trait	Unit	Weeks	Male mice	Female mice	p-value (<0.05)
				Mean ± SD	Mean ± SD	
Colorimetric Chemistry	Total Cholesterol	mg/dL	24	325.72 ± 240.80	354.46 ± 283.58	NS
Colorimetric Chemistry	Glucose	mg/dL	24	323.11 ± 83.56	296.73 ± 85.03	NS
Colorimetric Chemistry	Triglyceride	mg/dL	24	146.40 ± 115.03	205.15 ± 215.64	NS
Colorimetric Chemistry	Free glycerol	mg/dL	24	62.47 ± 48.35	92.93 ± 80.56	NS
Atherosclerosis	Aortic Lesion Area	µm2/section	24	24,416 ± 36,308	113,010 ± 97,437	1.40E-03

Table 4.3. Sex differences in cardiometabolic traits in DO-F1 mice at 24 weeks. n = 472 (238 females and 234 males).

Category	Trait	Unit	Weeks	Male mice	Female mice	p-value (<0.01)
				Mean ± SD	Mean ± SD	
Colorimetric Chemistry	Albumin	g/L	24	31.14 ± 3.22	31.25 ± 2.75	NS
Colorimetric Chemistry	ALT	U/L	24	123.37 ± 147.72	114.43 ± 97.54	NS
Colorimetric Chemistry	Total Cholesterol	mg/dL	24	607.88 ± 325.55	1033.04 ± 413.52	2.00E-16
Colorimetric Chemistry	VLDL/LDL-Cholesterol	mg/dL	24	540.82 ± 325.81	966.07 ± 414.16	2.00E-16
Colorimetric Chemistry	HDL-Cholesterol	mg/dL	24	67.06 ± 25.41	66.97 ± 27.68	NS
Colorimetric Chemistry	Glucose	mg/dL	24	313.01 ± 71.83	278.86 ± 51.49	1.70E-07
Colorimetric Chemistry	Triglyceride	mg/dL	24	533.51 ± 392.88	621.13 ± 390.96	2.80E-03
Colorimetric Chemistry	Urea/BUN	mmol/L	24	4.54 ± 0.95	4.22 ± 0.92	6.50E-05
Blood Pressure	Diastolic Blood Pressure	mmHG	24	84.11 ± 11.16	86.05 ± 12.78	NS
Blood Pressure	Systolic Blood Pressure	mmHG	24	111.71 ± 12.34	112.97 ± 13.37	NS
Blood Pressure	Mean Atrial Blood Pressure	mmHG	24	92.98 ± 11.19	94.69 ± 12.69	NS
Blood Pressure	Heart Rate	BPM	24	722.95 ± 54.06	726.83 ± 57.52	NS
Blood Pressure	Cardiac Output	µl/cycle	24	16 ± 5.35	16.55 ± 5.33	NS
Blood Pressure	Tail Blood Volume	µl	24	52.2 ± 17.19	53.46 ± 16.41	NS
LC/MS	TMAO	µM	24	3.97 ± 1.82	8.08 ± 4.86	2.00E-16
LC/MS	Choline	µM	24	11.56 ± 2.55	13.1 ± 3.56	5.40E-08
LC/MS	Betaine	µM	24	41.31 ± 12.04	58.37 ± 18.68	2.00E-16
LC/MS	Carnitine	µM	24	11.88 ± 3.16	11.87 ± 3.22	NS
Atherosclerosis	Aortic Lesion Area	µm2/section	24	22,211 ± 30,113	91,638 ± 83,287	2.00E-16
Atherosclerosis	Total aortic Lesion Area	µm2	24	97,978 ± 136,953	421,224 ± 394,929	2.00E-16

4.5.2. Sexually-biased Association between Atherosclerosis and Cardiometabolic Traits

To determine sex differences in the relationship between atherosclerosis and cardiometabolic traits, we performed Spearman correlation between aortic lesion area and cardiometabolic traits in females and males, respectively (**Figure 4.5A**). Plasma TC (Male: R=0.49, $P \leq 1.9\times10^{-15}$; Female: R=0.38, $P \leq 1.7\times10^{-9}$), plasma VLDL-C/LDL-C (Male: R=0.5, $P \leq 5.4\times10^{-16}$; Female: R=0.38, $P \leq 1.4\times10^{-9}$), and plasma TG (Male: R=0.32, $P \leq 4.5\times10^{-7}$; Female: R=0.13, $P \leq 0.042$) were positively correlated with the aortic lesion area and all three traits showed higher correlations with the aortic lesion area in males than in females (**Figure 4.5B-D**). Thus, the relationship among classic risk factors for atherosclerosis was stronger in male mice indicating that sex-based differences in lesion size may be due to complex genetic or transcript interactions.

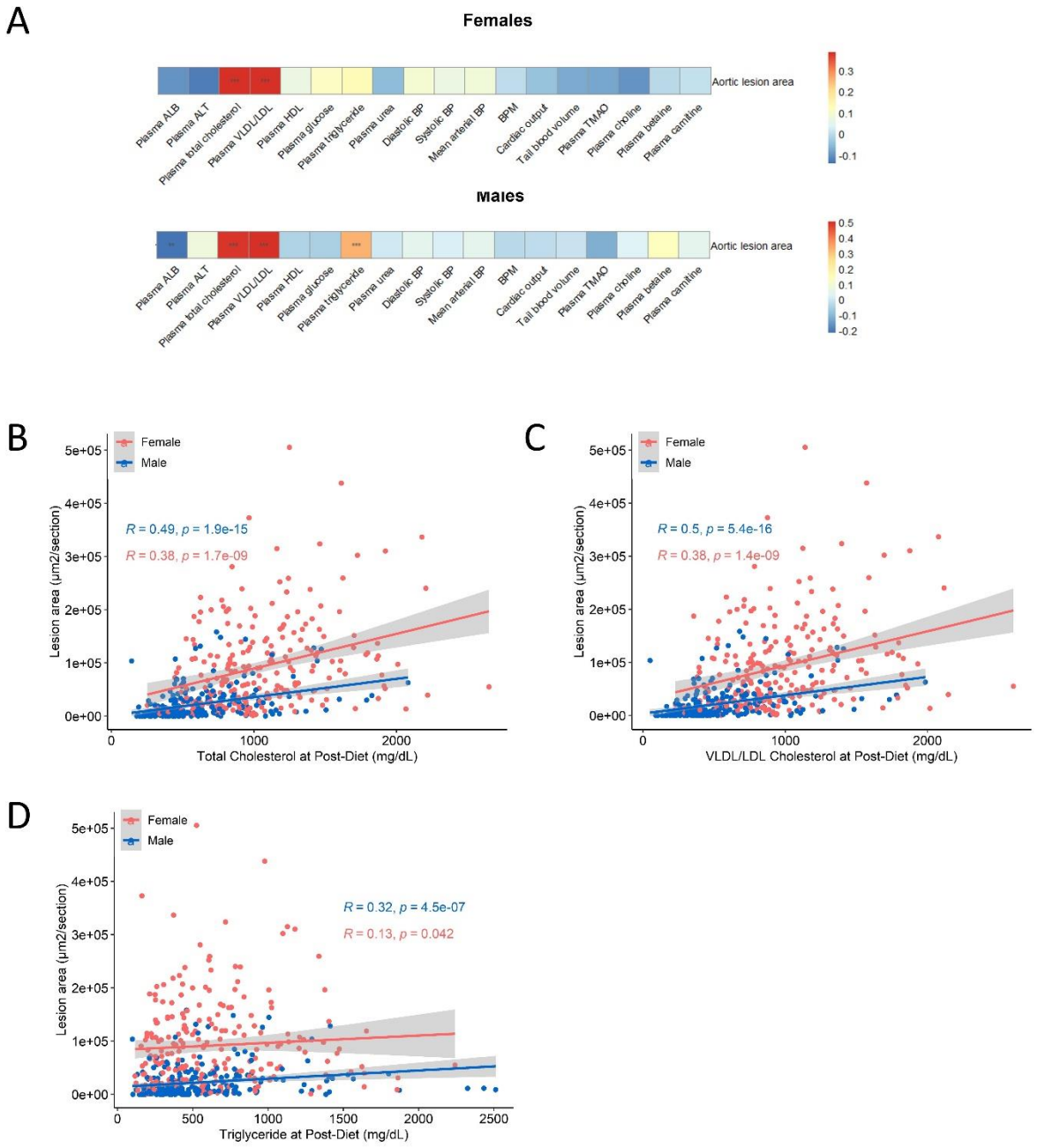


Figure 4.5. Identification of cardiometabolic traits associated with aortic lesion area in DO-F1 mice.

(A) Spearman correlation between aortic lesion area and cardiometabolic traits in females and males, respectively. The p-values were adjusted using the Benjamini-Hochberg (BH) FDR procedure. "***"P < 0.001, "**"P < 0.01, "*"P < 0.05, "."P < 0.10. (B-E) Spearman correlation between aortic lesion area and plasma total cholesterol (TC) (B), plasma very low-density lipoprotein cholesterol/low-density lipoprotein cholesterol (VLDL-C/LDL-C) (C), and plasma triglyceride (TG) (D) by sex.

4.5.3. Sex Influences Liver Gene Expression Profile

Previous studies have reported that thousands of genes are differentially expressed by sex (Kukurba et al., 2016; Rinn and Snyder, 2005; Wang et al., 2007; Yang et al., 2006). To better understand the underlying mechanism of the sexual dimorphism of atherosclerosis, we determined the liver gene expression profile using RNA-Seq in 85 females and 77 males of the DO-F1 mice. Samples were selected based on the aortic lesion area levels. Half of the mice with high aortic lesion size and half of the mice with low aortic lesion size were selected from each sex. A total of 13,094 genes including X-linked, Y-linked, mitochondrial (MT), and autosomal genes were determined to be robustly expressed. Principal component analysis demonstrates a global difference in gene expression between sexes (**Figure 4.6A**) with a total of 8,866 genes differentially expressed with 4,448 upregulated in females and 4,418 upregulated in males (adjusted p-value <0.05; **Table 4.4**). These 8,866 genes were defined as "sex-specific DEGs" and contain a number of previously reported sex chromosome genes (Female: *Xist*; Male: *Eif2s3y*, *Ddx3y*, *Uty*) and Cyp450 genes (Female: *Cyp2b9*, *Cyp2b13*, *Cyp2c40*; Male: *Cyp7b1*, *Cyp2u1*, *Cyp2f2*) (**Figure 4.6C and Table 4.4**) (Rinn and Snyder, 2005). When the number of sex-specific DEGs was counted for each chromosome, the proportions were generally similar between sexes except Y-linked and MT genes (**Figure 4.6B**).

To elucidate the differences in sex-specific biological pathways, we performed GO Biological process and KEGG pathway enrichment analysis of genes identified by the sex-specific DEGs. The enrichment analysis showed that the GO terms and KEGG pathways identified in each sex are sex-specific with no overlap at the FDR<0.001 level (**Figure 4.6D**) and upregulation of immune response in the females and upregulation of mitochondrial function in the males. Specifically, GO Biological process and KEGG pathways that were highly enriched in females included ''Neutrophil mediated immunity (GO: 0002446) and

Degranulation (GO: 0043312) (biological process)'' (-logP > 12) and ''Lysosome and Osteoclast differentiation (pathways)'' (-logP > 11). GO Biological process and KEGG pathways that were highly enriched in males included ''Respiratory electron transport chain (biological process)'' (-logP > 23) and ''Thermogenesis, Oxidative phosphorylation, and Non-alcoholic fatty liver disease (pathways)'' (-logP > 18) (**Figures 4.6E and 4.6F and Table 4.5 and 4.6**).

Next, we sought to understand if sex-specific DEGs have also been identified as sexually dimorphic in humans. The GTEx v8 data (Oliva et al., 2020), generated from 226 post-mortem donors, is among the few datasets available with human liver transcriptome analysis. A subset of 465 sex-specific liver DEGs (adjusted p-value<0.01) published using GTEx v8 are publicly available (Oliva et al., 2020), and of these 266 genes (224 genes in autosome and 42 genes in chromosome X) were among our 13,094 transcripts expressed in the DO-F1 livers. Among these 266 genes detected in humans, 187 transcripts (~70 %) were sexually dimorphic in DO-F1 livers, suggesting that transcripts in both human and mice liver tissue exhibited a high percentage of sex bias which may influence gender differences in phenotypes and disease risk.

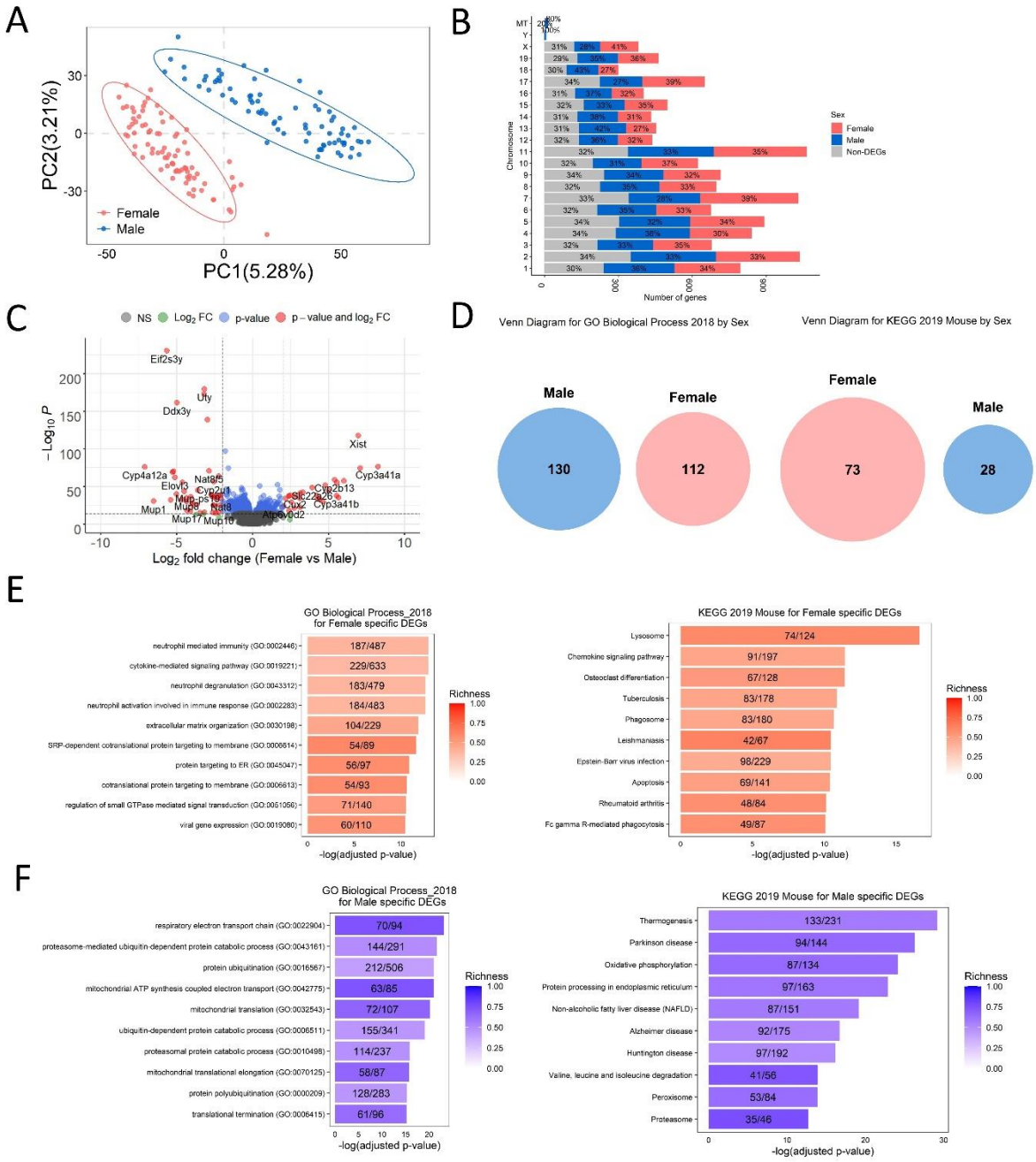


Figure 4.6. Sex differences in regulation of liver gene expression in DO-F1 mice. Sex-specific DEGs analysis was performed using the R package 'limma' from TMM (trimmed mean of M values) normalized log2 transformed TPM values. PCA analysis (A), the number of differential expression genes (DEGs) in each chromosome and mitochondria (MT) (B), and Volcano plot (C) of sex differences in liver gene expression in DO-F1 mice. (C) Horizontal dotted lines indicate adjusted p-value < 1×10^{-14}, vertical dotted gray lines indicate a 4-fold difference. (D) Venn diagram to identify sex-specific significant (adjusted p-value < 0.001) gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathway between female, male-specific DEGs, or non-DEGs in enrichment analysis. Top10 GO and KEGG pathway of female (E) and male (F) DEGs identified in enrichment analysis. Pathways are ordered from top to bottom by significance (highest to lowest) and colored by gene richness.

Table 4.4. List of differentially expressed genes (DEGs, |log2 fold change| > 3) between sexes in DO-F1 mice (adjusted p-value <0.05). n = 162 (85 females and 77 males)

Gene	Name	Chr	Start (Mbp)	logFC (Female vs Male)	Average expression (log2TPM)	Adj. P-value
Cyp3a41a	cytochrome P450, family 3, subfamily a, polypeptide 41A	5	145.69	8.26	5.22	3.05E-77
Sult2a2	sulfotransferase family 2A, dehydroepiandrosterone (DHEA)-preferring, member 2	7	13.73	7.07	4.02	5.03E-75
Xist	inactive X specific transcripts	X	103.46	6.95	4.10	2.52E-118
Cyp2c69	cytochrome P450, family 2, subfamily c, polypeptide 69	19	39.84	6.01	3.92	1.08E-57
Sult2a1	sulfotransferase family 2A, dehydroepiandrosterone (DHEA)-preferring, member 1	7	13.80	5.66	3.45	1.22E-35
Cyp2c40	cytochrome P450, family 2, subfamily c, polypeptide 40	19	39.77	5.55	4.78	5.87E-56
Cyp3a41b	cytochrome P450, family 3, subfamily a, polypeptide 41B	5	145.56	5.52	3.12	4.29E-38
Cyp2b13	cytochrome P450, family 2, subfamily b, polypeptide 13	7	26.06	5.40	3.68	1.46E-59
A1bg	alpha-1-B glycoprotein	15	60.92	5.00	2.92	6.95E-45
Fmo3	flavin containing monooxygenase 3	1	162.95	4.79	3.06	3.52E-52
Cyp2a4	cytochrome P450, family 2, subfamily a, polypeptide 4	7	26.31	4.65	6.42	1.16E-30
Cyp2b9	cytochrome P450, family 2, subfamily b, polypeptide 9	7	26.17	4.42	6.00	1.72E-33
Cyp3a16	cytochrome P450, family 3, subfamily a, polypeptide 16	5	145.44	4.29	2.34	1.52E-33
Cyp3a44	cytochrome P450, family 3, subfamily a, polypeptide 44	5	145.77	4.22	2.68	7.93E-39
Slc22a26	solute carrier family 22 (organic cation transporter), member 26	19	7.78	3.91	3.60	4.14E-49
Sult2a5	sulfotransferase family 2A, dehydroepiandrosterone (DHEA)-preferring, member 5	7	13.62	3.31	1.80	8.59E-43
Hao2	hydroxyacid oxidase 2	3	98.87	3.13	2.70	2.10E-41
Sult2a3	sulfotransferase family 2A, dehydroepiandrosterone (DHEA)-preferring, member 3	7	14.07	3.13	1.69	1.23E-22
Cyp2a22	cytochrome P450, family 2, subfamily a, polypeptide 22	7	26.93	3.10	5.04	8.26E-35
Uty	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked	Y	1.10	-3.16	1.50	5.15E-180
Mup16	major urinary protein 16	4	61.52	-3.17	5.56	3.22E-16
Kdm5d	lysine (K)-specific demethylase 5D	Y	0.90	-3.20	1.52	3.43E-173
Mup2	major urinary protein 2	4	60.14	-3.24	7.62	1.28E-13
Scara5	scavenger receptor class A, member 5	14	65.67	-3.24	2.12	2.33E-34
Mup19	major urinary protein 19	4	61.78	-3.44	9.75	1.93E-15
Mup-ps19	major urinary protein, pseudogene 19	4	61.96	-3.60	3.01	1.34E-45
Slco1a1	solute carrier organic anion transporter family, member 1a1	6	141.91	-3.65	5.41	1.65E-26
Mup15	major urinary protein 15	4	61.44	-3.73	5.59	2.66E-13
Slco1a1	solute carrier organic anion transporter family, member 1a1	6	141.91	-3.81	5.86	9.49E-26
Serpina1e	serine (or cysteine) peptidase inhibitor, clade A, member 1E	12	103.95	-3.95	7.01	3.98E-26
Hsd3b5	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 5	3	98.62	-3.96	2.51	7.85E-29

Serpina4-ps1	serine (or cysteine) peptidase inhibitor, clade A, member 4, pseudogene 1	12	104.08	-4.05	2.73	1.44E-36
Mup9	major urinary protein 9	4	60.42	-4.08	7.71	9.06E-18
Mup17	major urinary protein 17	4	61.59	-4.31	7.67	1.56E-18
Mup8	major urinary protein 8	4	60.22	-4.34	2.87	4.60E-35
Mup12	major urinary protein 12	4	60.74	-4.34	3.03	1.39E-30
Hsd3b5	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 5	3	98.62	-4.48	2.61	4.53E-17
Mup20	major urinary protein 20	4	62.05	-4.50	10.00	8.56E-44
Mup21	major urinary protein 21	4	62.15	-4.62	4.98	4.55E-56
Mup14	major urinary protein 14	4	61.30	-4.68	9.09	2.76E-33
Ddx3y	DEAD box helicase 3, Y-linked	Y	1.26	-4.99	2.37	3.23E-162
Cyp4a12b	cytochrome P450, family 4, subfamily a, polypeptide 12B	4	115.41	-5.00	3.74	1.61E-40
Elovl3	elongation of very long chain fatty acids (FEN1/Elo2, SUR4/Elo3, yeast)-like 3	19	46.13	-5.12	4.51	1.19E-62
Gm47283	predicted gene, 47283	Y	90.78	-5.22	2.48	1.32E-70
Cyp2d9	cytochrome P450, family 2, subfamily d, polypeptide 9	15	82.45	-5.26	7.06	4.26E-69
Mup7	major urinary protein 7	4	60.07	-5.40	9.88	8.39E-33
Elovl3	elongation of very long chain fatty acids (FEN1/Elo2, SUR4/Elo3, yeast)-like 3	19	46.13	-5.43	4.91	1.22E-57
Eif2s3y	eukaryotic translation initiation factor 2, subunit 3, structural gene Y-linked	Y	1.01	-5.65	2.69	3.22E-231
Mup1	major urinary protein 1	4	60.50	-6.52	8.36	3.62E-31
Cyp4a12a	cytochrome P450, family 4, subfamily a, polypeptide 12a	4	115.30	-7.11	5.17	2.44E-76

Table 4.5. Top 30 Gene Ontology results for sex-specific DEGs in liver in DO-F1 mice. n = 162 (85 females and 77 males).

Sex	Term	Overlap (gene count)	Adjusted. P.value	Rank
Female	neutrophil mediated immunity (GO:0002446)	187/487	1.17E-13	1
Female	cytokine-mediated signaling pathway (GO:0019221)	229/633	1.17E-13	2
Female	neutrophil degranulation (GO:0043312)	183/479	2.39E-13	3
Female	neutrophil activation involved in immune response (GO:0002283)	184/483	2.39E-13	4
Female	extracellular matrix organization (GO:0030198)	104/229	1.42E-12	5
Female	SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	54/89	2.24E-12	6
Female	protein targeting to ER (GO:0045047)	56/97	1.25E-11	7
Female	cotranslational protein targeting to membrane (GO:0006613)	54/93	2.30E-11	8
Female	regulation of small GTPase mediated signal transduction (GO:0051056)	71/140	2.84E-11	9
Female	viral gene expression (GO:0019080)	60/110	3.45E-11	10
Female	viral transcription (GO:0019083)	61/113	3.65E-11	11
Female	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (GO:0000184)	60/112	8.36E-11	12
Female	transmembrane receptor protein tyrosine kinase signaling pathway (GO:0007169)	148/396	3.82E-10	13
Female	positive regulation of intracellular signal transduction (GO:1902533)	168/479	5.00E-09	14
Female	cellular response to cytokine stimulus (GO:0071345)	161/456	6.89E-09	15
Female	plasma membrane bounded cell projection assembly (GO:0120031)	97/241	2.34E-08	16
Female	regulation of cell migration (GO:0030334)	117/316	1.32E-07	17
Female	positive regulation of apoptotic process (GO:0043065)	114/307	1.70E-07	18
Female	inflammatory response (GO:0006954)	97/252	3.42E-07	19
Female	regulation of apoptotic process (GO:0042981)	248/815	9.09E-07	20
Female	regulated exocytosis (GO:0045055)	64/148	1.04E-06	21
Female	cellular protein metabolic process (GO:0044267)	160/484	1.30E-06	22
Female	platelet degranulation (GO:0002576)	56/124	1.30E-06	23
Female	protein phosphorylation (GO:0006468)	156/470	1.36E-06	24
Female	toll-like receptor signaling pathway (GO:0002224)	43/86	1.61E-06	25
Female	response to lipopolysaccharide (GO:0032496)	65/155	2.77E-06	26
Female	peptidyl-tyrosine phosphorylation (GO:0018108)	40/79	3.11E-06	27
Female	positive regulation of programmed cell death (GO:0043068)	95/257	3.69E-06	28
Female	viral process (GO:0016032)	84/220	4.28E-06	29
Female	positive regulation of leukocyte chemotaxis (GO:0002690)	33/61	6.12E-06	30
Male	respiratory electron transport chain (GO:0022904)	70/94	7.03E-24	1
Male	proteasome-mediated ubiquitin-dependent protein catabolic process (GO:0043161)	144/291	2.50E-22	2
Male	protein ubiquitination (GO:0016567)	212/506	1.14E-21	3
Male	mitochondrial ATP synthesis coupled electron transport (GO:0042775)	63/85	1.14E-21	4

Male	mitochondrial translation (GO:0032543)	72/107	6.26E-21	5
Male	ubiquitin-dependent protein catabolic process (GO:0006511)	155/341	8.33E-20	6
Male	proteasomal protein catabolic process (GO:0010498)	114/237	1.44E-16	7
Male	mitochondrial translational elongation (GO:0070125)	58/87	1.85E-16	8
Male	protein polyubiquitination (GO:0000209)	128/283	4.86E-16	9
Male	translational termination (GO:0006415)	61/96	6.61E-16	10
Male	mitochondrial translational termination (GO:0070126)	58/89	7.03E-16	11
Male	mitochondrial respiratory chain complex assembly (GO:0033108)	61/97	1.17E-15	12
Male	regulation of primary metabolic process (GO:0080090)	77/139	1.92E-15	13
Male	fatty acid beta-oxidation (GO:0006635)	39/50	1.25E-14	14
Male	protein modification by small protein removal (GO:0070646)	117/261	2.07E-14	15
Male	protein deubiquitination (GO:0016579)	115/257	4.21E-14	16
Male	mitochondrial electron transport, NADH to ubiquinone (GO:0006120)	36/46	1.40E-13	17
Male	translational elongation (GO:0006414)	61/105	1.73E-13	18
Male	Golgi vesicle transport (GO:0048193)	118/271	1.73E-13	19
Male	regulation of cellular amine metabolic process (GO:0033238)	38/51	3.06E-13	20
Male	regulation of cellular amino acid metabolic process (GO:0006521)	38/51	3.06E-13	21
Male	protein modification by small protein conjugation (GO:0032446)	156/398	3.70E-13	22
Male	SCF-dependent proteasomal ubiquitin-dependent protein catabolic process (GO:0031146)	47/72	5.24E-13	23
Male	ER to Golgi vesicle-mediated transport (GO:0006888)	87/180	5.24E-13	24
Male	regulation of cellular ketone metabolic process (GO:0010565)	42/61	8.44E-13	25
Male	regulation of transcription from RNA polymerase II promoter in response to hypoxia (GO:0061418)	49/78	1.17E-12	26
Male	mitochondrion organization (GO:0007005)	81/167	3.27E-12	27
Male	fatty acid catabolic process (GO:0009062)	43/65	3.27E-12	28
Male	fatty acid oxidation (GO:0019395)	36/50	7.36E-12	29
Male	NADH dehydrogenase complex assembly (GO:0010257)	42/64	8.76E-12	30

Table 4.6. Top 30 KEGG pathway results for sex-specific DEGs in liver in DO-F1 mice. n = 162 (85 females and 77 males).

Sex	Term	Overlap (gene count)	Adjusted .P.value	Rank
Female	Lysosome	74/124	2.41E-17	1
Female	Chemokine signaling pathway	91/197	3.65E-12	2
Female	Osteoclast differentiation	67/128	3.72E-12	3
Female	Tuberculosis	83/178	1.30E-11	4
Female	Phagosome	83/180	2.24E-11	5
Female	Leishmaniasis	42/67	3.46E-11	6
Female	Epstein-Barr virus infection	98/229	3.66E-11	7
Female	Apoptosis	69/141	4.07E-11	8
Female	Rheumatoid arthritis	48/84	7.41E-11	9
Female	Fc gamma R-mediated phagocytosis	49/87	8.26E-11	10
Female	NF-kappa B signaling pathway	54/102	1.65E-10	11
Female	B cell receptor signaling pathway	42/72	4.93E-10	12
Female	Regulation of actin cytoskeleton	91/217	4.93E-10	13
Female	Focal adhesion	85/199	6.61E-10	14
Female	Pathways in cancer	182/535	7.83E-10	15
Female	Toxoplasmosis	53/108	7.24E-09	16
Female	Human T-cell leukemia virus 1 infection	94/245	5.14E-08	17
Female	C-type lectin receptor signaling pathway	52/112	1.10E-07	18
Female	Measles	62/144	1.50E-07	19
Female	Proteoglycans in cancer	80/203	1.59E-07	20
Female	Hematopoietic cell lineage	45/94	3.00E-07	21
Female	AGE-RAGE signaling pathway in diabetic complications	47/101	4.30E-07	22
Female	Influenza A	68/168	4.98E-07	23
Female	NOD-like receptor signaling pathway	79/205	5.12E-07	24
Female	Human papillomavirus infection	123/360	5.12E-07	25
Female	TNF signaling pathway	49/110	1.05E-06	26
Female	Herpes simplex virus 1 infection	141/433	1.40E-06	27
Female	MAPK signaling pathway	103/294	1.40E-06	28
Female	Chagas disease (American trypanosomiasis)	46/103	2.09E-06	29
Female	Endocytosis	95/269	2.60E-06	30
Male	Thermogenesis	133/231	6.51E-30	1
Male	Parkinson disease	94/144	5.15E-27	2
Male	Oxidative phosphorylation	87/134	7.28E-25	3

Male	Protein processing in endoplasmic reticulum	97/163	1.27E-23	4
Male	Non-alcoholic fatty liver disease (NAFLD)	87/151	6.67E-20	5
Male	Alzheimer disease	92/175	1.93E-17	6
Male	Huntington disease	97/192	6.70E-17	7
Male	Valine, leucine and isoleucine degradation	41/56	1.15E-14	8
Male	Peroxisome	53/84	1.36E-14	9
Male	Proteasome	35/46	1.91E-13	10
Male	Autophagy	66/130	9.21E-12	11
Male	Fatty acid degradation	35/50	1.04E-11	12
Male	Propanoate metabolism	24/31	1.52E-09	13
Male	Ubiquitin mediated proteolysis	64/138	2.55E-09	14
Male	Citrate cycle (TCA cycle)	24/32	4.16E-09	15
Male	PPAR signaling pathway	42/85	3.38E-07	16
Male	Pyruvate metabolism	24/38	8.36E-07	17
Male	Glyoxylate and dicarboxylate metabolism	21/31	9.09E-07	18
Male	RNA transport	67/167	1.03E-06	19
Male	Aminoacyl-tRNA biosynthesis	32/66	2.24E-05	20
Male	Retrograde endocannabinoid signaling	58/150	2.88E-05	21
Male	Butanoate metabolism	17/27	6.53E-05	22
Male	Terpenoid backbone biosynthesis	15/23	1.21E-04	23
Male	Protein export	17/28	1.21E-04	24
Male	Mitophagy	29/63	1.97E-04	25
Male	Biosynthesis of unsaturated fatty acids	18/32	2.67E-04	26
Male	Lysine degradation	27/59	4.03E-04	27
Male	Tryptophan metabolism	23/48	5.95E-04	28
Male	Fatty acid elongation	16/29	9.38E-04	29
Male	Porphyrin and chlorophyll metabolism	20/41	1.23E-03	30

4.5.4. Sex Effects on the Genetic Regulation of Liver Gene Expression

Sex-specific cardiometabolic traits can be partially derived from sex-specific genetic effects, which can also affect gene expression. We investigated the sexual dimorphism of the genetic regulation of gene expression by performing heritability and eQTL analysis in each sex independently. For a better understanding of the interaction between genetic polymorphism and sexually dimorphic development of atherosclerosis and cardiometabolic traits, we performed the heritability and QTL analysis for cardiometabolic traits in males and females using R package qtl2 (Broman et al., 2019). The narrow-sense heritability (h^2) was calculated using a linear mixed model upon accounting for genetic relatedness between the mice (kinship matrix) and DO generation. The heritability of the liver transcriptome was lower in females than in males, and this trend was the same for genes with *cis*-eQTL and *trans*-eQTL (**Figure 4.7A**).

To further investigate gene-by-sex interactions, we generated separate *cis*- and *trans*-eQTL for each sex, and assessed their relationship with sex-specific DEGs. We classified eQTLs with gene-by-sex interactions into the categories: female-specific eQTL, male-specific eQTL, overlapping eQTL with same direction, or overlapping eQTL with opposite direction (**Figure 4.7B**) (Yao et al., 2014). We separated eQTLs into these categories by fitting an independent linear model for each sex. First, we calculated the number of hepatic *cis*-eQTLs and *trans*-eQTLs at permutation-based p-value thresholds of 0.05 in both sexes in autosome and chromosome X (chrX). Strikingly, eQTLs in each sex on chrX were 4.6 times more frequent in males (1,041 eQTLs) than in females (225 eQTLs), all of which were *trans*-eQTLs. It is important to note that the genetic architecture of chrX in the DO-F1 mouse model is complex where males can only have a DO chrX while females have both a copy of the DO chrX and a paternal copy from their C57BL/6J sires. The frequency of eQTLs on chrX in the liver tissue of DO mice is highly male-biased (Melia and Waxman, 2020), we

focused on the autosomal QTLs consistent with previously reported DO-F1 study (Winter et al., 2017). We identified 1,408 *cis*-eQTLs and 980 *trans*-eQTL in females and 1,061 *cis*-eQTLs and 1,331 *trans*-eQTL in males (**Table 4.7)**. Of these only 751 *cis*-eQTLs and 183 *trans*-eQTL were significant (P<0.05) in both sexes (**Figures 4.7C**). The sex-specific proportions of eQTL were more pronounced for *trans*-eQTLs as 91.4% were sex-specific at 0.05 threshold (**Figure 4.7C**). We next examined if these sex-specific *trans*-eQTL form eQTL hotspots, where many expression traits colocalize on the same locus, and sex-specific hotspots were found on chromosome 16 in females and chromosome 10 in males (**Figure 4.7D and 4.7E**). In addition, we assessed the sex-specific DEGs by their eQTL status in each sex and found that sex-specific DEGs were less likely to have a significant eQTL (Fisher's exact test, $P \leq 2.2 \times 10^{-16}$ in females; Fisher's exact test, $P \leq 2.0 \times 10^{-9}$ in males) (**Table 4.8**). Among the 2,258 genes that have female-biased eQTLs, the proportion identified as genes upregulated in females was 26.4% (632 genes), and among the 2,250 genes that have male-biased eQTLs, the proportion identified as genes upregulated in males was 28.3% (676 genes) (**Figure 4.7F**). For example, *Mup* gene family, known as sexually dimorphic genes (Clodfelter et al., 2006) located at 60.1-61.9 Mbp on chromosome 4 were highly upregulated in males and have male-specific *cis*-eQTLs (**Figure 4.8 and 4.9)**.

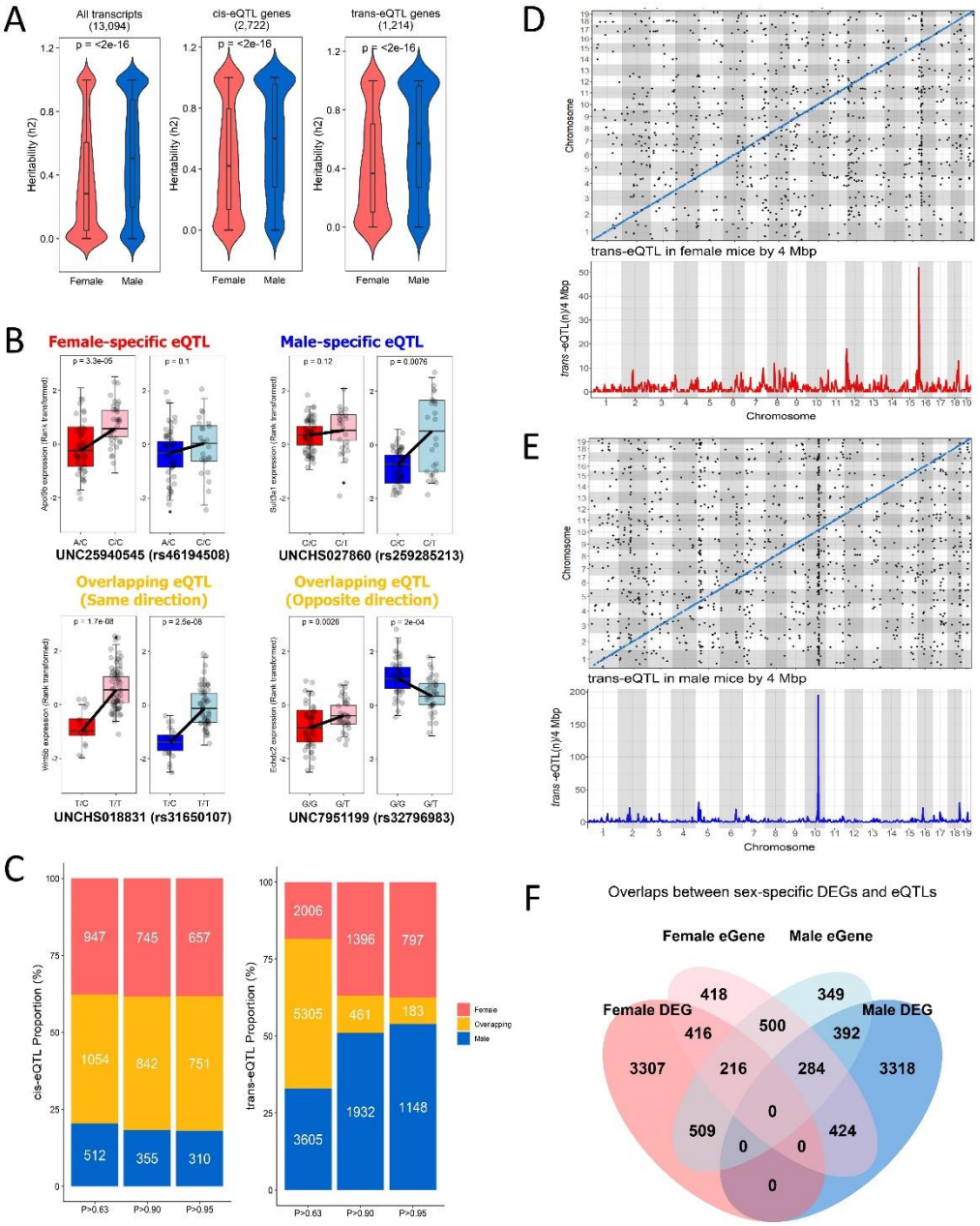


Figure 4.7. Sex-biased eQTLs and their relationship with sex-specific DEGs and narrow-sense heritability (h^2) in liver transcriptome in the DO-F1 mice.
(A) Narrow sense heritability of autosomal transcripts, *cis*-eQTL genes, and *trans*-eQTL genes between sexes. (B) Gene-by-sex interactions into four categories: female-specific eQTL, male-specific eQTL, overlapping eQTL between sexes with same direction or opposite direction. (C) Proportion of the sex-specific *cis*-eQTLs or *trans*-eQTLs at multiple genome-wide P-value thresholds from 0.05 to 0.63 between sexes. The white number in the bar graph is the number of sex-specific eQTLs or eQTLs in both sexes. Red bar contains female-specific eQTLs, blue bar contains male-specific eQTLs, and yellow bar contains eQTLs in both sexes. (D,E) The number of significant (P<0.05) genome-wide *trans*-eQTLs in females (D) and males (E) occurring within a 4 Mbp genomic window. (F) Overlaps between sex-specific DEGs and genes that have sex-biased eQTLs (sb-eGenes). Red circle contains female-specific DEGs, blue circle contains male-specific DEGs, pink circle contains female-specific eGenes, and light blue circle contains male-specific eGenes.

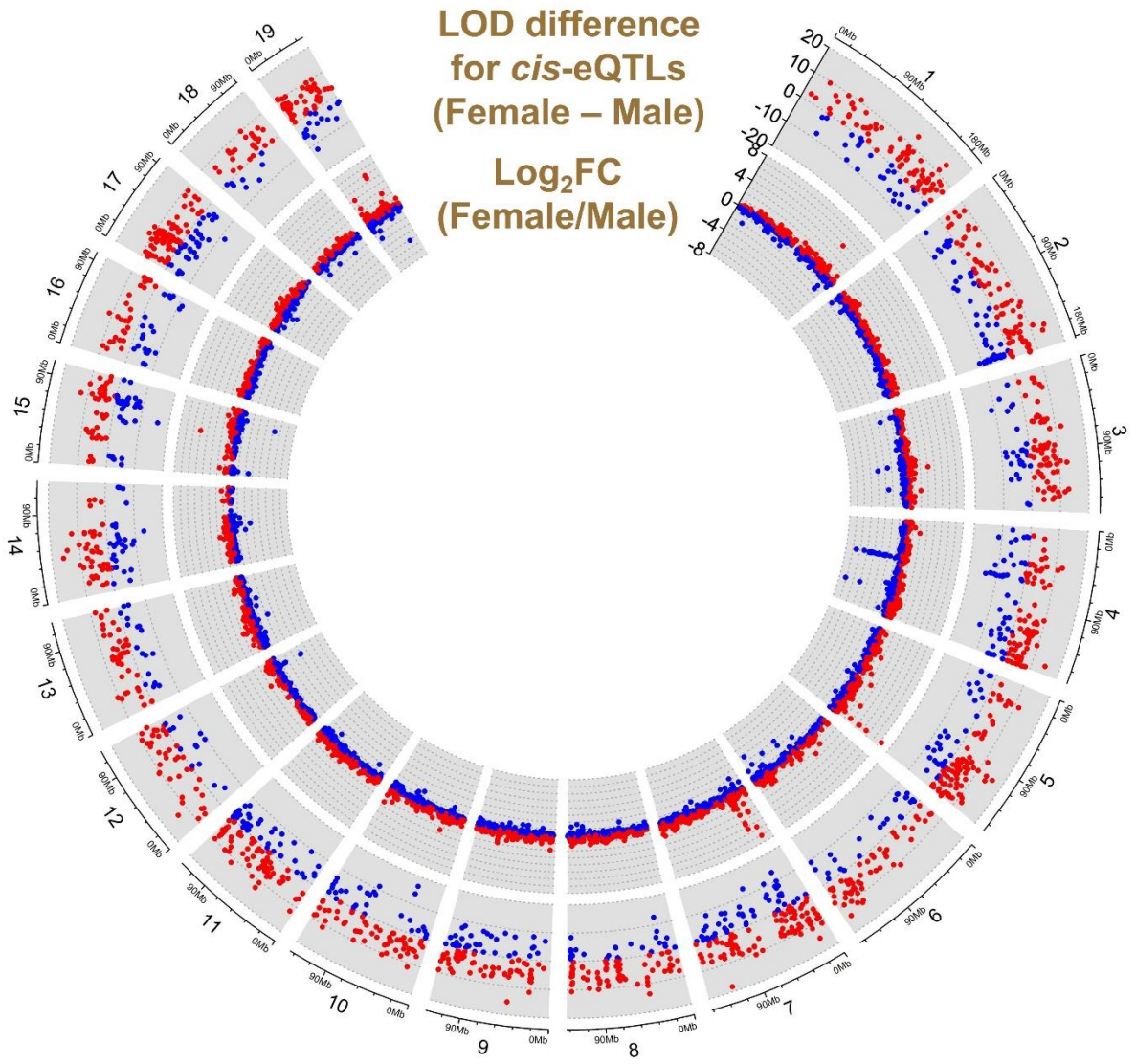


Figure 4.8. Genome-wide distribution of sex-biased genes with genetic regulations. Genomic position enrichment of sex-biased local expression QTL (*cis*-eQTL) and gene expression across autosome. The outer layer shows the chromosome location. The outer track represent difference of LOD scores for 1,718 *cis*-eQTLs (657 female-specific *cis*-eQTLs, 310 male-specific *cis*-eQTLs, and 751 overlapping *cis*-eQTLs) between females and males. Female-biased *cis*-eQTLs have positive LOD difference range (red) and male-biased *cis*-eQTLs have negative LOD difference range (blue). Each dot represents a sex-biased *cis*-eQTL on each chromosome of the mouse genome for a given transcript. The inner track shows the relative log2fold change between sexes for 13,094 liver gene expressions. Red and blue dots are upregulated genes in females and males, respectively.

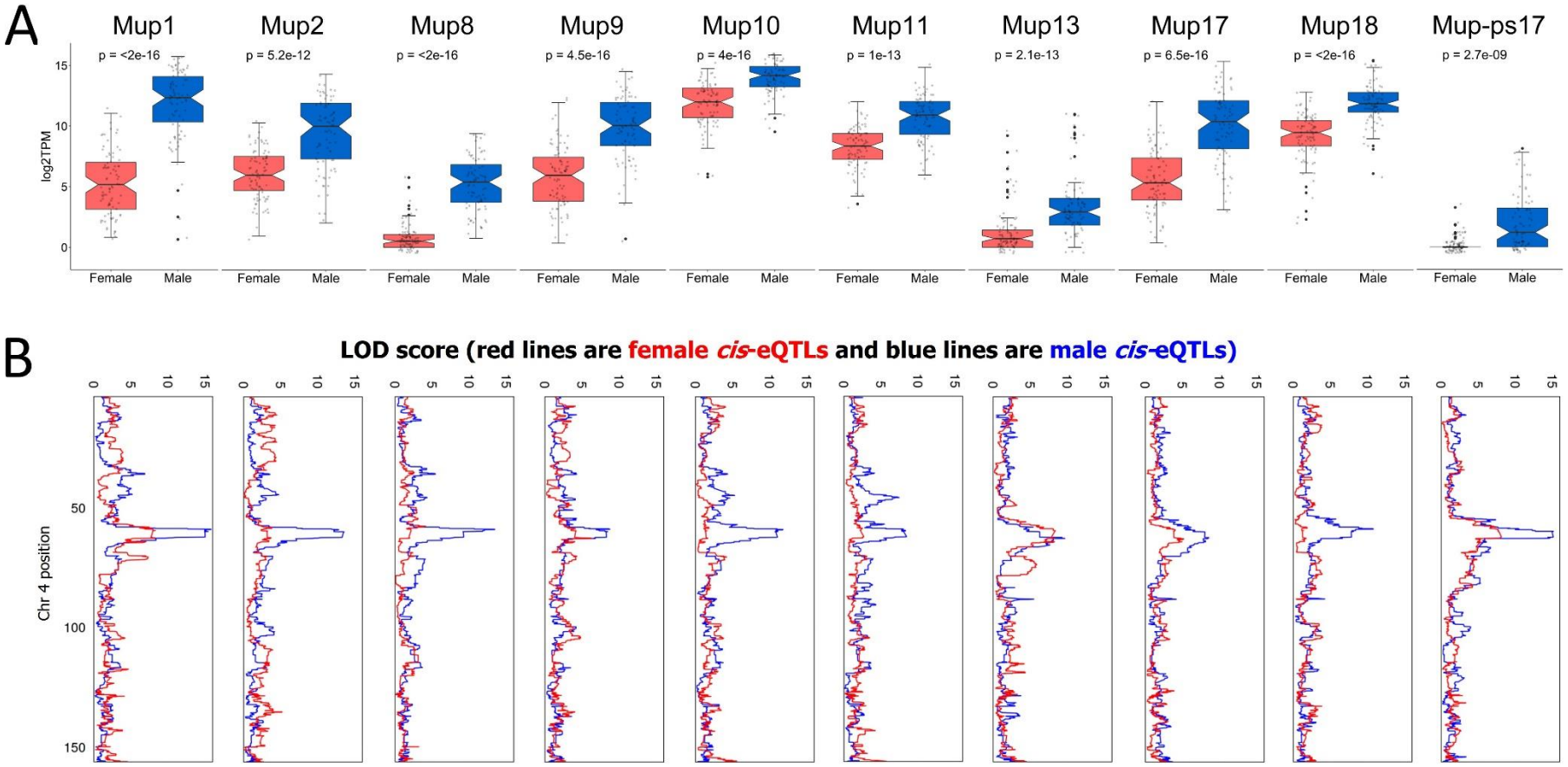


Figure 4.9. Enrichment of sex-biased genes with genetic regulations on chromosome 4.
(A) Expressions of 10 *Mup* genes such as *Mup1*, *Mup2*, *Mup8*, *Mup9*, *Mup10*, *Mup11*, *Mup13*, *Mup17*, *Mup18*, and *Mup-ps17* (log2TPM) between sexes, respectively. 10 *Mup* genes were significantly upregulated in males ($P < 0.001$). The p-values were confirmed by Wilcoxon signed-rank test. (B) Comparison of *cis*-QTLs for 10 *Mup* genes in between females and males on chromosome 4. LOD score in female mice is red line and LOD score in female mice is blue line.

Table 4.7. The number of cis-eQTLs and trans-eQTLs at multiple genome-wide P value thresholds from 0.05 to 0.63 in DO-F1 mice. n = 162

Total 13094 transcripts	**Permutation-based LOD threshold**		
***cis*-eQTL**	0.05	0.1	0.63
Female mice	1408	1587	2001
Male mice	1061	1197	1566
***trans*-eQTL**	0.05	0.1	0.63
Female mice	980	1857	7311
Male mice	1331	2393	8910
Total eQTL	0.05	0.1	0.63
Female mice	2388	3444	9312
Male mice	2392	3590	10476

Table 4.8. Fisher's exact test to evaluate the overlap between sex-specific DEGs with sex-biased eQTLs.

		# DEGs		
		Yes	No	Total
# female-biased eQTL	Yes	1340	918	2258
	No	7526	3310	10836
	Total	8866	4228	13094
Fisher's exact test in females: odds ratio 0.64, 95% CI 0.58–0.71, p-value = 2.2E-16)				
# male-biased eQTL	Yes	1401	849	2250
	No	7465	3379	10844
	Total	8866	4228	13094
Fisher's exact test in males: odds ratio 0.75, 95% CI 0.68–0.82, p-value = 2.0E-09)				

4.5.5. Sex-Specific Quantitative Trait Loci Mapping for Atherosclerosis and Cardio-metabolic Traits

We next assessed the heritability for clinical traits to examine the genetic contribution of our 20 cardio-metabolic traits. Similar to our studies of gene expression, the number of traits in females with a lower heritability than males was 11 out of 15 traits (73.3%) excluding traits with <0.1 h^2 differences. In both sexes, traits with h^2 less than 0.5 include plasma ALT, plasma HDL-C, and blood pressure (diastolic/systolic/mean arterial blood pressure, beat per minute, and tail blood volume) (**Figure 4.10 and Table 4.9**).

Based on our observation of sex-specific differences in cardiometabolic traits and gene expression, we next investigated genetic loci that were associated with traits in each sex. We first focused on the identified aortic lesion area QTLs. Specifically in the atherosclerosis mouse model, since females are more susceptible than males, and the effects on host genetics also perturb sexual dimorphism (AlSiraj et al., 2019; Su et al., 2006), we conducted QTL analysis in sex additive model, female mice, and male mice, respectively. A female-specific model identified one highly suggestive (P<0.1) QTL on chromosome 10 (**Figure 4.11A and 4.11C and Table 4.10**) and one significant (P<0.05) novel QTL was identified on chromosome 19 only in a male-specific model (**Figure 4.11B and 4.11D and Table 4.10**). We identified 33 and 58 genes in the confidence intervals for aortic lesion area QTLs on chromosomes 10 and 19, and the number of expressed genes in the liver were 28 and 58, respectively (**Table 4.11**). Next, we also performed QTL analysis on 20 cardiometabolic traits to explore the sex-specific QTLs for cardiometabolic traits associated with atherosclerosis. **Figure 4.10** shows 8 out of the 20 cardio-metabolic traits illustrate the substantial effect of host genetics on sexual dimorphism (**Table 4.10**). Specifically, we identified QTLs for plasma ALT and TG in females only, and QTLs for plasma glucose, urea nitrogen, and betaine only in males.

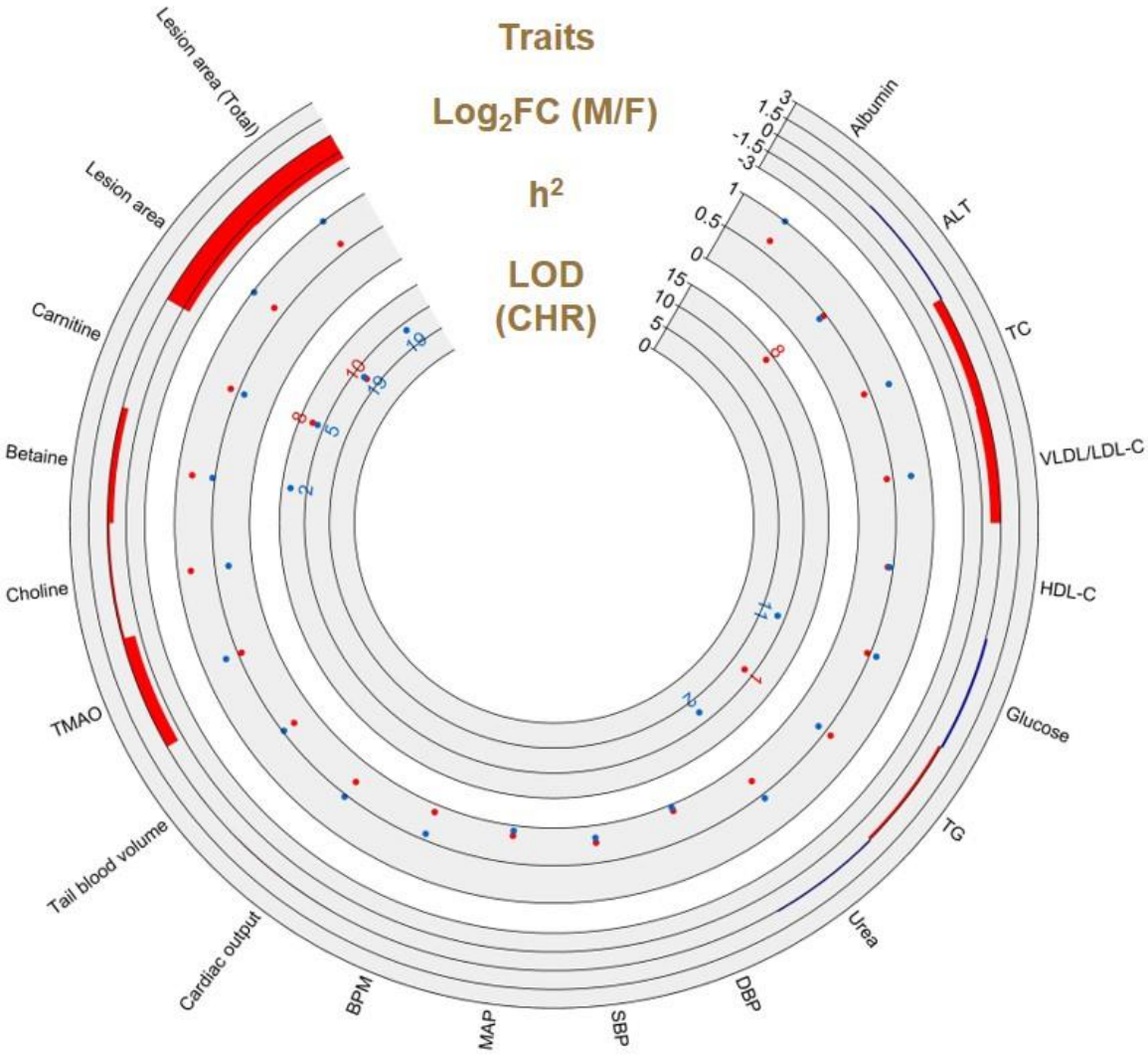


Figure 4.10. Genetic architecture of quantitative trait loci (QTL) for cardiometabolic traits.

Circos plot showing an overview of genetic regulation of cardio-metabolic traits at 24 weeks. The outermost track (blue bars for males and red bars for females) show the relative log2fold change (FC) (male versus female). The middle track represents narrow-sense heritability (h^2) estimates for the cardio-metabolic traits in females (red) and males (blue). The innermost track represents logarithm of the odds (LOD) scores of significant quantitative trait loci (QTL) in females (red) and males (blue), with their respective chromosomes indicated below (male) or above (female) dots.

Table 4.9. Narrow sense heritability for cardio-metabolic traits in DO-F1 mice. n = 461 (235 females and 226 males).

Category	Trait	Unit	Weeks	Narrow sense heritability		
				All mice	Female mice	Male mice
Colorimetric Chemistry	Albumin	g/L	24	0.64	0.67	1.00
Colorimetric Chemistry	ALT	U/L	24	0.19	0.47	0.40
Colorimetric Chemistry	Total Cholesterol	mg/dL	24	0.37	0.42	0.77
Colorimetric Chemistry	VLDL/LDL-Cholesterol	mg/dL	24	0.35	0.42	0.74
Colorimetric Chemistry	HDL-Cholesterol	mg/dL	24	0.31	0.43	0.45
Colorimetric Chemistry	Glucose	mg/dL	24	0.38	0.46	0.59
Colorimetric Chemistry	Triglyceride	mg/dL	24	0.29	0.59	0.38
Colorimetric Chemistry	Urea/BUN	mmol/L	24	0.22	0.27	0.56
Blood Pressure	Diastolic Blood Pressure	mmHG	24	0.10	0.09	0.04
Blood Pressure	Systolic Blood Pressure	mmHG	24	0.12	0.23	0.17
Blood Pressure	Mean Atrial Blood Pressure	mmHG	24	0.11	0.14	0.07
Blood Pressure	Heart Rate	BPM	24	0.22	0.10	0.42
Blood Pressure	Cardiac Output	µl/cycle	24	0.32	0.28	0.53
Blood Pressure	Tail Blood Volume	µl	24	0.28	0.31	0.48
LC/MS	TMAO	µM	24	0.23	0.45	0.67
LC/MS	Choline	µM	24	0.53	0.83	0.32
LC/MS	Betaine	µM	24	0.41	0.81	0.53
LC/MS	Carnitine	µM	24	0.43	0.61	0.41
Atherosclerosis	Aortic Lesion Area	µm2/section	24	0.50	0.64	0.98
Atherosclerosis	Total aortic Lesion Area	µm2	24	0.46	0.62	1.00

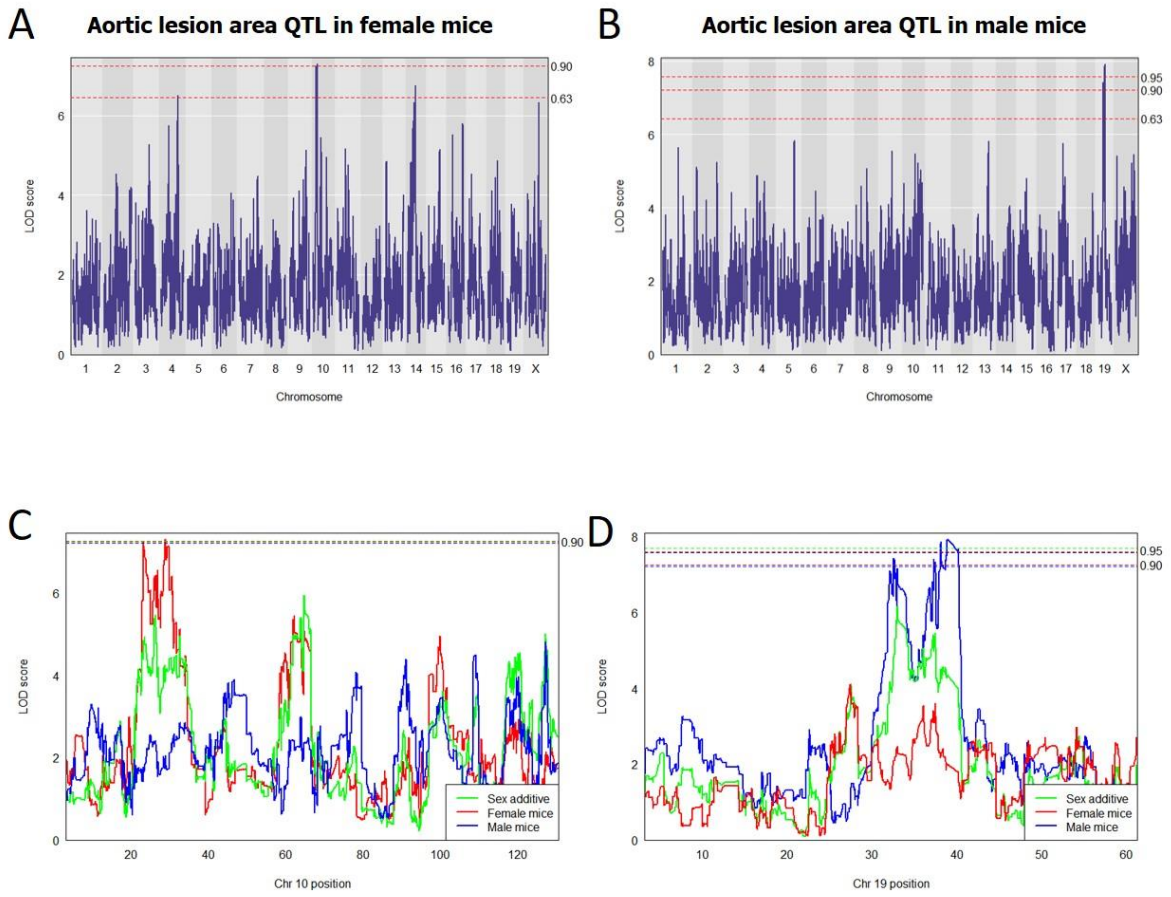


Figure 4.11. Aortic lesion area chromosomal QTL graphs in four models of DO-F1 mice.

(A,B) Genome-wide aortic lesion area QTLs on chromosome 10 in female mice (A) and chr19 in male mice (B) fed a synthetic high-fat and high-cholesterol (HFHC) diet for 16 weeks. Dashed lines correspond to $P < 0.05$ (significant), $P < 0.1$ (highly suggestive) or $P < 0.63$ (suggestive) thresholds. (C,D) Comparison of aortic lesion area QTLs in three models on chromosome 10 in female mice (C) and chromosome 19 in male mice (D). Black, males and females sex–genotype interaction model; red, sex additive model; green, female mice; blue, male mice. Dashed lines correspond to $P < 0.05$ (significant) and $P < 0.1$ (highly suggestive)

Table 4.10. Significant QTL results for cardiometabolic traits in four models and strain difference in regression coefficient of the association between each trait and marker SNP. n = 461 (235 females and 226 males).

Model	Traits	Chr	Position (Mbp)	LOD	CI (low)[a]	CI(hi)[a]	LOD threshold ($p < 0.05$)	#Genes	Marker
Female mice	Plasma triglyceride	1	36.50	8.11	35.81	38.31	8.08	31	UNC454349
Female mice	Plasma carnitine	8	105.06	12.35	104.65	105.93	7.78	56	UNC15429900
Female mice	Plasma ALT	8	113.40	13.55	112.14	113.83	12.33	2	UNCHS024248
Female mice	Aortic lesion area (Average)	4	132.73	6.51	55.60	134.86	7.60	502	UNC8247956
Female mice	Aortic lesion area (Average)	10	28.85	7.30	22.90	30.75	7.60	33	JAX00286290
Female mice	Aortic lesion area (Average)	14	68.04	6.74	37.40	70.35	7.60	237	UNC24171938
Female mice	Aortic lesion area (Total)	10	28.85	6.71	22.78	64.75	7.68	177	JAX00286290
Male mice	Plasma betaine	2	83.51	13.28	83.51	83.53	12.60	1	ICR5037
Male mice	Plasma urea	2	134.38	7.74	133.94	136.06	7.70	6	UNCHS006856
Male mice	Plasma carnitine	5	45.32	11.28	44.71	46.59	10.02	8	UNCHS014223
Male mice	Plasma glucose	11	100.41	8.45	100.11	104.92	7.72	122	UNC20243587
Male mice	Aortic lesion area (Average)	19	38.17	7.92	32.00	40.23	7.56	58	UNCHS047709
Male mice	Aortic lesion area (Total)	19	38.17	8.57	32.10	40.23	7.69	57	UNCHS047709
Sex additive	Plasma carnitine	5	45.42	10.02	45.38	46.55	8.70	7	UNCHS014228
Sex additive	Plasma triglyceride	6	124.86	7.78	122.58	125.54	7.64	70	UNC12014770
Sex additive	Tail blood volume	15	76.31	8.01	76.11	79.72	7.66	108	UNCHS040869
Sex additive	Diastolic BP	17	82.93	8.20	81.63	83.56	7.90	4	UNC28476662
Sex additive	Mean arterial BP	17	82.93	8.14	81.63	83.56	7.86	4	UNC28476662
Sex additive	Aortic lesion area (Average)	19	33.00	6.55	32.09	39.09	7.69	54	UNCHS047622
Sex additive	Aortic lesion area (Total)	19	37.48	6.80	32.09	40.12	7.71	54	UNCHS047690
Sex interactive	Plasma albumin	3	93.19	11.90	92.85	97.99	10.81	96	UNCHS009409
Sex interactive	Plasma glucose	11	100.41	12.79	100.11	100.47	11.54	39	UNC20243587
Sex interactive	Aortic lesion area (Average)	4	131.37	9.94	55.60	132.70	10.68	379	JAX00566629
Sex interactive	Aortic lesion area (Average)	10	63.65	9.27	23.18	66.19	10.68	201	JAX00290960
Sex interactive	Aortic lesion area (Total)	4	131.37	9.83	62.10	132.70	10.68	406	JAX00566629

Regression coefficient of the association between trait and Marker SNP								Strain that have highest coefficient	Strain that have lowest coefficient
A/J	B6	129	NOD	NZO	CAST	PWK	WSB		
-100.40	-117.95	-48.13	17.25	-17.51	-27.11	397.94	-104.10	PWK	B6
-0.99	-0.21	-1.25	-1.18	-1.68	2.38	1.87	1.06	CAST	NZO
-46.79	-53.92	-37.66	-73.50	-115.38	502.82	-110.25	-65.32	CAST	NZO
0.23	0.24	0.07	-0.06	0.20	-0.46	0.16	-0.39	B6	CAST
-0.06	0.05	0.16	-0.13	0.40	-0.24	-0.38	0.21	NZO	PWK
0.09	0.33	-0.06	0.20	-0.41	-0.03	-0.23	0.11	B6	NZO
-0.06	0.11	0.19	-0.13	0.36	-0.23	-0.40	0.16	NZO	PWK
-1.15	1.09	1.39	-0.63	-0.72	-0.35	0.36	0.00	129	A/J
0.17	-0.25	-0.40	-0.53	0.51	0.56	0.11	-0.16	CAST	NOD
-1.94	-1.94	-0.72	11.62	-2.66	-2.61	-0.16	-1.59	NOD	NZO
-18.09	-22.64	-17.97	41.52	15.04	33.43	-4.55	-26.74	NOD	WSB
0.15	0.00	0.27	-0.50	-0.15	0.19	0.34	-0.30	PWK	NOD
0.14	0.00	0.27	-0.47	-0.16	0.15	0.33	-0.27	PWK	NOD
-1.62	-1.17	-0.50	9.61	-1.96	-2.23	-0.28	-1.84	NOD	CAST
-0.16	0.34	0.08	0.00	0.25	-0.22	-0.28	0.00	B6	CAST
-2.37	-2.75	4.21	-1.99	0.06	9.20	0.33	-6.68	CAST	WSB
0.67	-5.93	-0.03	-0.57	-3.74	-1.04	11.01	-0.37	PWK	B6
0.57	-5.87	-0.05	-0.23	-3.76	-1.16	10.77	-0.26	PWK	B6
0.28	-0.06	-0.12	-0.23	-0.11	0.13	0.17	-0.06	A/J	NOD
0.26	-0.07	0.15	-0.20	-0.21	-0.09	0.26	-0.11	PWK	NOD
1.30	-0.18	0.60	0.02	-0.73	-0.63	-0.59	0.21	A/J	NZO
-7.34	-12.35	-1.40	10.98	6.32	21.85	-8.18	-9.88	CAST	B6
0.18	0.09	-0.08	-0.21	0.16	-0.18	-0.03	0.07	A/J	NOD
0.07	-0.07	0.19	0.01	0.24	-0.15	-0.25	-0.04	NZO	PWK
0.18	0.09	-0.08	-0.21	0.16	-0.18	-0.03	0.07	A/J	NOD

Table 4.11. 28 genes in female aortic lesion area QTL on chromosome 10 and 58 genes in male aortic lesion area QTL on chromosome 19 and their DEG status and correlation with aortic lesion area.

Model	Gene	Name	Chr	Start (Mbp)	End (Mbp)	DEG	Correlation with aortic lesion area			
							r (Female)	adj. p-value (Female)	r (Male)	adj. p-value (Male)
Female mice	Tbpl1	TATA box binding protein-like 1	10	22.70	22.73	Female	-0.215	0.095	0.064	0.657
Female mice	Rps12	ribosomal protein S12	10	23.79	23.79	No	**-0.267**	**0.033**	-0.101	0.471
Female mice	Slc18b1	solute carrier family 18, subfamily B, member 1	10	23.80	23.83	Male	-0.066	0.640	0.035	0.804
Female mice	Vnn3	vanin 3	10	23.85	23.87	Female	-0.036	0.806	-0.096	0.473
Female mice	Vnn1	vanin 1	10	23.89	23.91	Male	-0.082	0.555	-0.083	0.538
Female mice	Stx7	syntaxin 7	10	24.15	24.19	Female	-0.088	0.522	0.190	0.137
Female mice	Moxd1	monooxygenase, DBH-like 1	10	24.22	24.30	Male	**-0.253**	**0.044**	0.068	0.636
Female mice	Ccn2	cellular communication network factor 2	10	24.60	24.60	No	-0.097	0.479	0.015	0.916
Female mice	Enpp1	ectonucleotide pyrophosphatase/phosphodiesterase 1	10	24.64	24.71	Female	-0.188	0.143	-0.382	0.001
Female mice	Enpp3	ectonucleotide pyrophosphatase/phosphodiesterase 3	10	24.77	24.84	Male	0.058	0.682	-0.095	0.480
Female mice	Med23	mediator complex subunit 23	10	24.87	24.91	Female	-0.217	0.087	0.072	0.599
Female mice	Arg1	arginase, liver	10	24.92	24.93	No	-0.134	0.310	-0.244	0.052
Female mice	Akap7	A kinase (PRKA) anchor protein 7	10	25.17	25.30	Female	0.096	0.480	-0.084	0.534
Female mice	Epb41l2	erythrocyte membrane protein band 4.1 like 2	10	25.36	25.52	Female	-0.182	0.158	0.070	0.608
Female mice	Smlr1	small leucine-rich protein 1	10	25.53	25.54	Male	-0.041	0.791	0.077	0.589
Female mice	L3mbtl3	L3MBTL3 histone methyl-lysine binding protein	10	26.27	26.38	No	-0.118	0.380	0.145	0.267
Female mice	Arhgap18	Rho GTPase activating protein 18	10	26.77	26.92	Female	-0.214	0.092	0.023	0.873
Female mice	Lama2	laminin, alpha 2	10	26.98	27.62	Female	-0.230	0.068	-0.034	0.809
Female mice	Gm10145	predicted gene 10145	10	27.94	27.94	No	0.152	0.255	-0.018	0.905
Female mice	Ptprk	protein tyrosine phosphatase, receptor type, K	10	28.07	28.60	No	**-0.234**	**0.037**	-0.119	0.367
Female mice	Themis	thymocyte selection associated	10	28.67	28.88	Male	-0.119	0.390	0.125	0.363
Female mice	Echdc1	enoyl Coenzyme A hydratase domain containing 1	10	29.31	29.35	Male	-0.068	0.631	-0.399	0.001
Female mice	Rnf146	ring finger protein 146	10	29.34	29.36	Male	-0.019	0.899	-0.143	0.272
Female mice	Rspo3	R-spondin 3	10	29.45	29.54	No	-0.105	0.439	-0.101	0.448

Female mice	Gm10275	predicted pseudogene 10275	10	29.70	29.70	No	-0.213	0.097	0.168	0.209
Female mice	Trmt11	tRNA methyltransferase 11	10	30.53	30.60	Male	-0.152	0.246	0.009	0.951
Female mice	Hint3	histidine triad nucleotide binding protein 3	10	30.60	30.62	Male	-0.138	0.295	-0.077	0.572
Female mice	Ncoa7	nuclear receptor coactivator 7	10	30.65	30.81	Female	-0.093	0.497	0.102	0.445
Male mice	A1cf	APOBEC1 complementation factor	19	31.87	31.95	Male	-0.064	0.653	**-0.380**	**0.002**
Male mice	Asah2	N-acylsphingosine amidohydrolase 2	19	31.98	32.10	Male	-0.090	0.513	-0.175	0.173
Male mice	Sgms1	sphingomyelin synthase 1	19	32.12	32.39	Female	-0.232	0.065	0.237	0.059
Male mice	Rpl9-ps6	ribosomal protein L9, pseudogene 6	19	32.47	32.47	No	0.092	0.518	-0.003	0.987
Male mice	Minpp1	multiple inositol polyphosphate histidine phosphatase 1	19	32.49	32.52	No	**-0.326**	**0.007**	-0.062	0.654
Male mice	Papss2	3'-phosphoadenosine 5'-phosphosulfate synthase 2	19	32.60	32.67	Female	0.151	0.250	**-0.285**	**0.021**
Male mice	Atad1	ATPase family, AAA domain containing 1	19	32.67	32.74	Male	**-0.269**	**0.030**	-0.197	0.122
Male mice	Pten	phosphatase and tensin homolog	19	32.76	32.83	Male	-0.141	0.284	**-0.321**	**0.009**
Male mice	Rnls	renalase, FAD-dependent amine oxidase	19	33.14	33.39	No	-0.067	0.646	0.072	0.619
Male mice	Lipo3	lipase, member O3	19	33.56	33.59	Female	-0.082	0.553	0.146	0.264
Male mice	Stambpl1	STAM binding protein like 1	19	34.19	34.24	Female	0.128	0.335	0.174	0.177
Male mice	Acta2	actin, alpha 2, smooth muscle, aorta	19	34.24	34.26	No	-0.025	0.867	0.056	0.683
Male mice	Fas	Fas (TNF receptor superfamily member 6)	19	34.29	34.33	Female	0.146	0.266	-0.035	0.805
Male mice	Lipa	lysosomal acid lipase A	19	34.49	34.53	Female	0.058	0.681	-0.025	0.860
Male mice	Ifit2	interferon-induced protein with tetratricopeptide repeats 2	19	34.55	34.58	Female	0.067	0.637	0.192	0.133
Male mice	Ifit3	interferon-induced protein with tetratricopeptide repeats 3	19	34.58	34.59	Female	0.010	0.949	0.120	0.382
Male mice	Ifit3b	interferon-induced protein with tetratricopeptide repeats 3B	19	34.61	34.61	Female	0.123	0.369	0.212	0.104
Male mice	Ifit1	interferon-induced protein with tetratricopeptide repeats 1	19	34.64	34.65	Female	0.097	0.479	0.161	0.214
Male mice	Slc16a12	solute carrier family 16 (monocarboxylic acid transporters), member 12	19	34.67	34.75	Female	-0.003	0.984	-0.141	0.280
Male mice	Pank1	pantothenate kinase 1	19	34.81	34.88	Male	0.100	0.464	**-0.269**	**0.030**
Male mice	Rpp30	ribonuclease P/MRP 30 subunit	19	36.08	36.10	No	-0.034	0.818	0.170	0.188
Male mice	Ankrd1	ankyrin repeat domain 1 (cardiac muscle)	19	36.11	36.12	Female	-0.150	0.263	0.059	0.685
Male mice	Pcgf5	polycomb group ring finger 5	19	36.35	36.46	Male	-0.114	0.396	0.036	0.797
Male mice	Hectd2	HECT domain E3 ubiquitin protein ligase 2	19	36.55	36.62	Male	**-0.259**	**0.038**	-0.041	0.770
Male mice	Hectd2os	Hectd2, opposite strand	19	36.62	36.69	Male	**0.250**	**0.047**	0.013	0.936
Male mice	Ppp1r3c	protein phosphatase 1, regulatory subunit 3C	19	36.73	36.74	No	0.187	0.153	-0.023	0.882
Male mice	Tnks2	tankyrase, TRF1-interacting ankyrin-related ADP-ribose polymerase 2	19	36.83	36.89	No	-0.207	0.104	0.048	0.730

Male mice	Btaf1	B-TFIID TATA-box binding protein associated factor 1	19	36.93	37.01	Male	**-0.308**	**0.012**	0.139	0.290
Male mice	Cpeb3	cytoplasmic polyadenylation element binding protein 3	19	37.02	37.21	Male	-0.229	0.069	-0.223	0.077
Male mice	Marchf5	membrane associated ring-CH-type finger 5	19	37.21	37.22	Male	**-0.354**	**0.003**	-0.106	0.424
Male mice	Ide	insulin degrading enzyme	19	37.27	37.33	Male	**-0.269**	**0.030**	0.041	0.772
Male mice	Rpl10-ps6	ribosomal protein L10, pseudogene 6	19	37.30	37.30	No	0.052	0.729	0.228	0.078
Male mice	Kif11	kinesin family member 11	19	37.38	37.42	Female	-0.155	0.246	-0.032	0.832
Male mice	Gm38345	predicted gene, 38345	19	37.43	37.43	No	-0.092	0.518	-0.137	0.313
Male mice	Hhex	hematopoietically expressed homeobox	19	37.43	37.44	Male	-0.005	0.973	0.033	0.814
Male mice	Exoc6	exocyst complex component 6	19	37.54	37.68	No	**-0.329**	**0.007**	-0.122	0.357
Male mice	Cyp26a1	cytochrome P450, family 26, subfamily a, polypeptide 1	19	37.70	37.70	No	-0.037	0.799	-0.119	0.368
Male mice	Myof	myoferlin	19	37.90	38.04	Female	-0.059	0.676	0.187	0.144
Male mice	Rbp4	retinol binding protein 4, plasma	19	38.12	38.13	Male	-0.060	0.671	**-0.417**	**0.000**
Male mice	Fra10ac1	FRA10AC1 homolog (human)	19	38.19	38.22	Male	0.072	0.607	-0.037	0.791
Male mice	Slc35g1	solute carrier family 35, member G1	19	38.40	38.41	Male	**-0.263**	**0.034**	-0.091	0.498
Male mice	Plce1	phospholipase C, epsilon 1	19	38.48	38.79	No	**-0.283**	**0.022**	0.171	0.186
Male mice	Noc3l	NOC3 like DNA replication regulator	19	38.79	38.82	Male	**-0.275**	**0.026**	0.147	0.259
Male mice	Tbc1d12	TBC1D12: TBC1 domain family, member 12	19	38.84	38.92	Male	-0.160	0.219	-0.016	0.912
Male mice	Hells	helicase, lymphoid specific	19	38.93	38.97	No	**-0.352**	**0.004**	-0.016	0.913
Male mice	Cyp2c55	cytochrome P450, family 2, subfamily c, polypeptide 55	19	39.01	39.04	Female	-0.035	0.810	-0.002	0.990
Male mice	Cyp2c29	cytochrome P450, family 2, subfamily c, polypeptide 29	19	39.29	39.33	No	**0.311**	**0.011**	0.060	0.664
Male mice	Cyp2c38	cytochrome P450, family 2, subfamily c, polypeptide 38	19	39.39	39.46	Female	0.153	0.243	-0.194	0.130
Male mice	Cyp2c39	cytochrome P450, family 2, subfamily c, polypeptide 39	19	39.51	39.57	Female	0.231	0.066	-0.235	0.062
Male mice	Cyp2c67	cytochrome P450, family 2, subfamily c, polypeptide 67	19	39.61	39.65	Male	-0.223	0.082	**-0.364**	**0.003**
Male mice	Cyp2c68	cytochrome P450, family 2, subfamily c, polypeptide 68	19	39.69	39.74	Female	0.123	0.372	-0.171	0.200
Male mice	Cyp2c40	cytochrome P450, family 2, subfamily c, polypeptide 40	19	39.77	39.81	Female	0.225	0.075	-0.036	0.799
Male mice	Cyp2c69	cytochrome P450, family 2, subfamily c, polypeptide 69	19	39.84	39.89	Female	0.182	0.164	-0.080	0.573
Male mice	Cyp2c37	cytochrome P450, family 2. subfamily c, polypeptide 37	19	39.99	40.01	Female	0.200	0.117	-0.062	0.652
Male mice	Cyp2c54	cytochrome P450, family 2, subfamily c, polypeptide 54	19	40.04	40.07	No	0.200	0.123	-0.029	0.847
Male mice	Cyp2c50	cytochrome P450, family 2, subfamily c, polypeptide 50	19	40.09	40.11	No	0.229	0.068	0.026	0.854
Male mice	Cyp2c70	cytochrome P450, family 2, subfamily c, polypeptide 70	19	40.15	40.19	No	0.237	0.059	0.053	0.703
Male mice	Pdlim1	PDZ and LIM domain 1 (elfin)	19	40.22	40.27	Female	0.009	0.953	0.096	0.475

4.5.6. Colocalization between aortic lesion area QTLs and *cis*-eQTLs, and correlation approach reveal sex-specific candidate genes associated with atherosclerosis

Next, we sought *cis*-eQTLs colocalized with female or male aortic lesion area QTLs to prioritize the candidate genes associated with lesion development. We found 10 *cis*-eQTLs colocalized with female aortic lesion area QTL on chromosome 10 (LOD: 7.3, BCI: 22.90-30.75 Mbp), and 16 *cis*-eQTLs colocalized with male aortic lesion area QTL on chromosome 19 (LOD: 7.92, BCI: 32.00-40.23 Mbp) (**Table 4.12**). Of these 26 genes that have *cis*-eQTLs, 19 were sex-specific DEGs and *Pten* (phosphatase and tensin homolog) on chromosome 19 was the only gene that showed a significant upregulation ($P \leq 2.4 \times 10^{-6}$) and correlation (R=-0.32, $P \leq 0.0047$) with the aortic lesion area in males (**Figures. 4.12A and 4.12B; Table 4.12**). We next compared whether shared founder allele effects were observed between aortic lesion areas QTL and *Pten cis*-QTL. The CAST and PWK alleles were associated with increased lesion size, and the NOD, NZO, and WSB alleles were associated with decreased lesions on chromosome 19 QTL (**Figure 4.12C**). However, *cis*-eQTL for *Pten* showed the opposite allele effect pattern as the aortic lesion area QTL. *Pten cis*-eQTL also showed a dependence on the CAST and PWK allele, but in the opposite direction (**Figure 4.12C**). Only one SNP (rs30401869) in the *Pten* gene was associated with both aortic lesion area ($P<0.05$) and *Pten* gene expression ($P<0.001$ in all mice, females, or males) (**Figure 4.12D and 4.12E and Table 4.13**). The direction of these associations suggests that SNPs present in CAST and PWK lead to a decrease of *Pten* expression and an increase of aortic lesion size at the colocalized locus.

Table 4.12. Co-localization between significant aortic lesion area QTLs and liver cis-eQTLs, and a direct correlation between the mapped trait and transcript.

Aortic lesion area QTL						cis-eQTL[b]						Trait:Transcript correlation coefficient and p-value					
												Whole mice		Female mice		Male mice	
Model	Chr	Position (Mbp)	LOD	CI (low)[a]	CI (hi)[a]	Gene	Chr	Position (Mbp)	LOD	CI (low)[a]	CI (hi)[a]	DEG	r	adjusted p-value	r	adjusted p-value	r	adjusted p-value
Female mice	10	28.85	7.30	22.73	30.80	Rps12	10	24.18	38.93	23.11	24.63	No	-0.138	0.109	**-0.267**	**0.033**	-0.101	0.471
Female mice	10	28.85	7.30	22.73	30.80	Ptprk	10	27.86	9.20	26.23	28.90	No	-0.156	0.069	**-0.234**	**0.037**	-0.119	0.367
Female mice	10	28.85	7.30	22.73	30.80	Moxd1	10	23.90	16.38	23.43	24.63	Male	**-0.202**	**0.017**	**-0.253**	**0.044**	0.068	0.636
Female mice	10	28.85	7.30	22.73	30.80	Arhgap18	10	26.43	11.18	26.38	26.88	Female	-0.112	0.879	-0.214	0.092	0.023	0.873
Female mice	10	28.85	7.30	22.73	30.80	Gm10145	10	28.35	28.33	27.92	28.67	No	0.108	0.221	0.152	0.255	-0.018	0.905
Female mice	10	28.85	7.30	22.73	30.80	Hint3	10	30.41	21.80	29.75	31.01	Male	-0.123	0.203	-0.138	0.295	-0.077	0.572
Female mice	10	28.85	7.30	22.73	30.80	Akap7	10	25.61	14.64	24.17	25.66	Female	0.015	0.853	0.096	0.480	-0.084	0.534
Female mice	10	28.85	7.30	22.73	30.80	Slc18b1	10	23.56	8.84	23.11	25.06	Male	-0.018	0.159	-0.066	0.640	0.035	0.804
Female mice	10	28.85	7.30	22.73	30.80	Smlr1	10	25.66	9.74	24.58	25.80	Male	-0.009	0.923	-0.041	0.791	0.077	0.589
Female mice	10	28.85	7.30	22.73	30.80	Vnn3	10	23.73	22.33	23.50	24.63	Female	-0.061	0.504	-0.036	0.806	-0.096	0.473
Male mice	19	38.17	7.92	31.89	40.25	Cyp2c67	19	40.25	11.73	38.67	40.51	Male	**-0.458**	**4.27E-09**	-0.223	0.082	**-0.364**	**0.003**
Male mice	19	38.17	7.92	31.89	40.25	Pten	19	30.90	8.85	29.37	34.07	Male	**-0.194**	**0.022**	-0.141	0.284	**-0.321**	**0.009**
Male mice	19	38.17	7.92	31.89	40.25	Cyp2c39	19	38.83	42.82	38.67	40.23	Female	0.074	0.408	0.231	0.066	-0.235	0.062
Male mice	19	38.17	7.92	31.89	40.25	Rpl10-ps6	19	38.67	10.84	37.26	40.25	No	**0.170**	**0.046**	0.052	0.729	0.228	0.078
Male mice	19	38.17	7.92	31.89	40.25	Cyp2c38	19	40.22	20.00	37.57	40.25	Female	0.025	0.787	0.153	0.243	-0.194	0.130
Male mice	19	38.17	7.92	31.89	40.25	Asah2	19	32.10	11.98	31.82	32.43	Male	-0.123	0.158	-0.090	0.513	-0.175	0.173
Male mice	19	38.17	7.92	31.89	40.25	Cyp2c68	19	37.59	11.56	37.48	40.25	Female	**0.308**	**1.57E-04**	0.123	0.372	-0.171	0.200
Male mice	19	38.17	7.92	31.89	40.25	Lipo3	19	34.02	19.25	32.91	34.28	Female	0.009	0.924	-0.082	0.553	0.146	0.264
Male mice	19	38.17	7.92	31.89	40.25	Slc16a12	19	34.84	9.45	34.28	36.37	Female	-0.051	0.579	-0.003	0.984	-0.141	0.280
Male mice	19	38.17	7.92	31.89	40.25	Pdlim1	19	40.45	9.51	38.67	40.51	Female	0.047	0.611	0.009	0.953	0.096	0.475
Male mice	19	38.17	7.92	31.89	40.25	Rnls	19	34.02	12.89	32.85	34.28	No	0.151	0.079	-0.067	0.646	0.072	0.619
Male mice	19	38.17	7.92	31.89	40.25	Minpp1	19	32.40	15.52	31.73	32.88	No	**-0.201**	**0.017**	**-0.326**	**0.007**	-0.062	0.654
Male mice	19	38.17	7.92	31.89	40.25	Fra10ac1	19	38.17	10.75	37.57	38.68	Male	0.013	0.894	0.072	0.607	-0.037	0.791
Male mice	19	38.17	7.92	31.89	40.25	Pcgf5	19	36.19	8.89	35.96	37.29	Male	-0.057	0.531	-0.114	0.396	0.036	0.797
Male mice	19	38.17	7.92	31.89	40.25	Fas	19	34.28	10.86	34.04	35.95	Female	0.076	0.396	0.146	0.266	-0.035	0.805
Male mice	19	38.17	7.92	31.89	40.25	Rpl9-ps6	19	32.43	15.73	32.40	32.53	No	0.035	0.705	0.092	0.518	-0.003	0.987

Table 4.13. List of *Pten* SNP that showed significant difference of aortic lesion area and gene expression between heterozygous and homozygous genotypes.

Gene	Maker	Chr	Position	SNP	P-value for aortic lesion area	Aortic lesion area in heterozygous genotype mice	P-value for gene expression (All mice)	P-value for gene expression (Females)	P-value for gene expression (Males)	Gene expression in heterozygous genotype mice	Strains that have minor alleles	Major allele	Minor allele
Pten	UNC30214131	19	32.802285	rs30401869	0.042	Lower	2.00E-07	2.00E-07	2.00E-07	Higher	B6;129;CAST;PWK	T	C

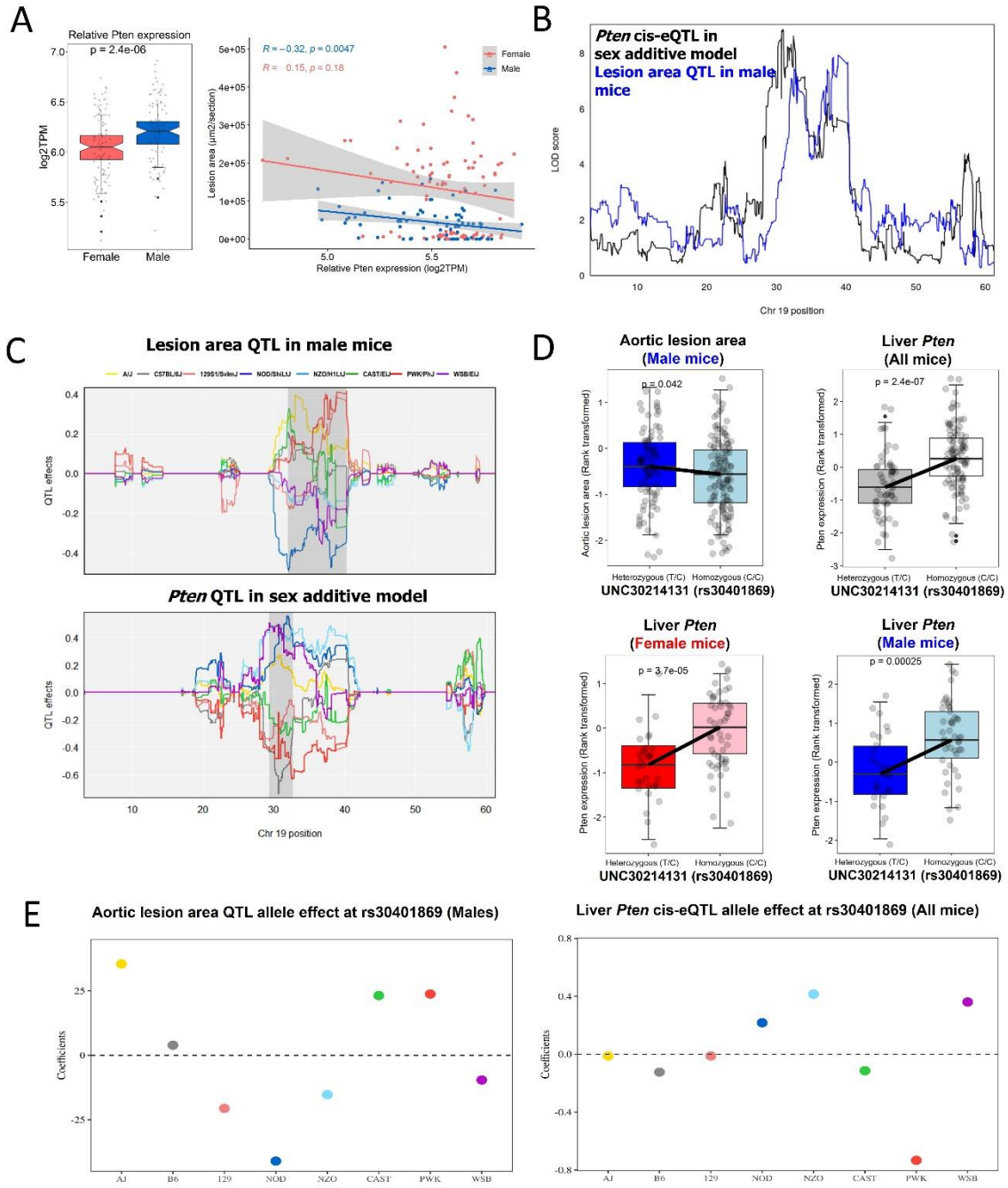

A
Relative Pten expression
p = 2.4e-06
log2TPM
Female
Male
R = −0.32, p = 0.0047
R = 0.15, p = 0.18
Female
Male
Lesion area (um2/section)
Relative Pten expression (log2TPM)
B
Pten cis-eQTL in sex additive model
Lesion area QTL in male mice
LOD score
Chr 19 position
C
Lesion area QTL in male mice
A/J
C57BL/6J
129S1/SvImJ
NOD/ShiLtJ
NZO/H1LtJ
CAST/EiJ
PWK/PhJ
WSB/EiJ
QTL effects
Pten QTL in sex additive model
QTL effects
Chr 19 position
D
Aortic lesion area (Male mice)
p = 0.042
Aortic lesion area (Rank transformed)
Heterozygous (T/C)
Homozygous (C/C)
UNC30214131 (rs30401869)
Liver Pten (All mice)
p = 2.4e-07
Pten expression (Rank transformed)
Heterozygous (T/C)
Homozygous (C/C)
UNC30214131 (rs30401869)
Liver Pten (Female mice)
p = 3.7e-05
Pten expression (Rank transformed)
Heterozygous (T/C)
Homozygous (C/C)
UNC30214131 (rs30401869)
Liver Pten (Male mice)
p = 0.00025
Pten expression (Rank transformed)
Heterozygous (T/C)
Homozygous (C/C)
UNC30214131 (rs30401869)
E
Aortic lesion area QTL allele effect at rs30401869 (Males)
Coefficients
AJ
B6
129
NOD
NZO
CAST
PWK
WSB
Liver Pten cis-eQTL allele effect at rs30401869 (All mice)
Coefficients
AJ
B6
129
NOD
NZO
CAST
PWK
WSB

Figure 4.12. Sex-specific aortic lesion area QTLs co-localized with liver gene expression of *Pten* in males.
(A) *Pten* gene (log2TPM) is highly upregulated in males ($P<1\times10^{-5}$) and negatively correlated with aortic lesion area in males (R = -0.32, P<0.0047), not females (R = -0.14, P<0.21) by Spearman correlation. The p-values were Wilcoxon signed-rank test between sexes for *Pten* gene expressions. (B) LOD profiles on chromosome 19 highlighting a locus significantly associated with the liver *Pten* gene expression and aortic lesion area in male mice. A black line for *cis*-eQTL in sex additive model (N=162) and a blue line for aortic lesion area QTL in male mice (N=226). (C) Estimated founder allele effect plots for aortic lesion area QTL in males and *Pten* *cis*-eQTL in sex additive model. Male aortic lesion area was linked to the high allele in PWK strain and low allele in NOD strain and *Pten* gene expression was linked to the high allele in NOD strain and low allele in PWK strain at the chromosome 19 QTL. (D) Association between a SNP (rs30401869) in *Pten* gene and aortic lesion area in males (blue color) or liver *Pten* gene expression in all mice (grey color), females (red color), and males (blue color). (E) Estimated founder strain levels of aortic lesion area in males and liver *Pten* gene expression in all mice were inferred from the founder strain coefficients observed at the SNP (rs30401869). B6, C57BL/6J; 129, 129S1/SvlmJ; NOD, NOD/ShiLtJ; NZO, NZO/HILtJ; CAST, CAST/EiJ; PWJ, PWK/PhJ; WSB, WSB/EiJ.

4.5.7. Identification of liver transcription factors with sex-specific gene expression, genetic regulation, and gene-trait correlation

We hypothesized that specific differences in the expression of transcription factors be critical determinants of sexually dimorphic gene expression which is supported in the literature (Naqvi et al., 2019). To evaluate whether sex-specific expression of the liver transcriptome and genetic regulations are useful as a means to dissect the molecular basis of liver transcription factors (TFs) for atherosclerosis, we examined the sex-specific expression and genetic regulation of liver TFs and performed correlation analysis with cardiometabolic traits in each sex. First, we prioritized TF binding sites (TFBSs) of 453 mouse TFs identified by large-scale chromatin immunoprecipitation sequencing analysis(Kulakovskiy et al., 2018) in promoter regions (i.e., 2 kb upstream of the transcription start site) and primarily filtered 265 TFs expressed in liver in DO-F1 mice. Of these, 172 TFs (64.9 %) were identified as sex-specific DEGs and biological pathways that were highly enriched in these sex-specific TFs were associated with inflammatory responses such as ‘‘Th17 cell differentiation” (-logP > 13) and “C-type lectin receptor signaling pathway’’ (-logP > 8), among others (**Figure 4.13A**). The 172 TFs include known hormone-related TFs (i.e. *Esr1*), nuclear receptors (i.e. *Nr1h3, Ppara*, and *Rara*), and non-reported sex-related TFs. The greatest difference between female and male-biased expressions of TFs was observed for TFBS of *Cux2*, *Atf3*, *Spi1*, *Irf8*, and *Esr1* (log2fold change (female vs male) > 0.9) and of *Onecut1*, *Bcl6*, *Foxa1*, *Ppara*, and *Xbp1* (log2fold change (female vs male) < -0.5), respectively. Interestingly, among these sex-bised TFs, *Spi1* and *Irf8* localized to a male-specific trans-hotspot on chr10.

Next, we investigated which cardiometabolic traits exhibited sex-specific correlations with transcripts including liver TFs. We selected genes that significantly correlated with each of

the cardiometabolic traits in each sex at the BH-adjusted $p<0.05$ level. In general, most cardiometabolic traits showed sex-specific correlation with liver transcripts except for plasma ALT (**Figure 4.14**), and this trend was the same for liver TFs (**Figure 4.13B**).

We next sought to identify liver TFs that are genetically regulated and contribute to the susceptibility of atherosclerosis in a sex-specific manner. We focused on liver TFs that showed sex-biased *cis*-eQTLs and correlated with atherosclerosis (**Figure 4.13C**). Of the 172 TFs that were differently expressed between sexes, only *Nr1h3* (encoding Liver X receptor alpha) TF, known as a cholesterol sensor (Peet et al., 1998) satisfied the criteria (**Figure 4.13C**). For example, *Nr1h3* expression was positively correlated with aortic lesion area only in females (**Figures 4.13D**). However, *Nr1h3* expression showed a positive correlation with plasma TC regardless of sex, but not with plasma TG (**Figures 4.13D**). In addition, *Nr1h3* had a *cis*-eQTL in both sexes, but with opposite directions of effect (**Figures 4.13E**). Lastly, we examined direct or indirect target genes of *Nr1h3* to determine if the sex-biased expression, genetic regulation, and correlation with aortic lesion area of *Nr1h3* could affect the sex-biased outcome of the genes targeted by *Nr1h3*. Of the 19 *Nr1h3*-dependent genes reported (Edwards et al., 2002; Wang and Tontonoz, 2018), 8 genes upregulated in females showed a strong positive correlation with the *Nr1h3* expression and plasma TC and TG concentrations in males (**Figure 4.13F and Figures 4.15 and 4.16**). These genes are involved in cholesterol efflux (*Abcg1*), inflammation (*Ccl2*, *Ccl7*, *Il1b*, and *Ptgs1*), lipoprotein remodeling (*Lpl* and *Pltp*), and cholesterol synthesis (*Rnf145*). On the other hand, expression of lipogenesis (*Acaca*, *Fasn*, and *Scd1*), phospholipid metabolism (*Lpcat3*), and cholesterol uptake (*Ldlr*)-related genes were significantly increased in males, and there were no significant correlations with *Nr1h3* expression in both sexes. Only the direct target genes of *Nr1h3* TF, *Abca1*, *Abcg1*, and *Lpl*, showed strong positive correlations with *Nr1h3*

expression, aortic lesion area, plasma TC and TG levels in males (**Figure 4.13F and Figures 4.15 and 4.16**). Thus, these results suggest that sex-biased regulation of TFs may contribute to the pathogenesis of sexually dimorphic atherosclerosis.

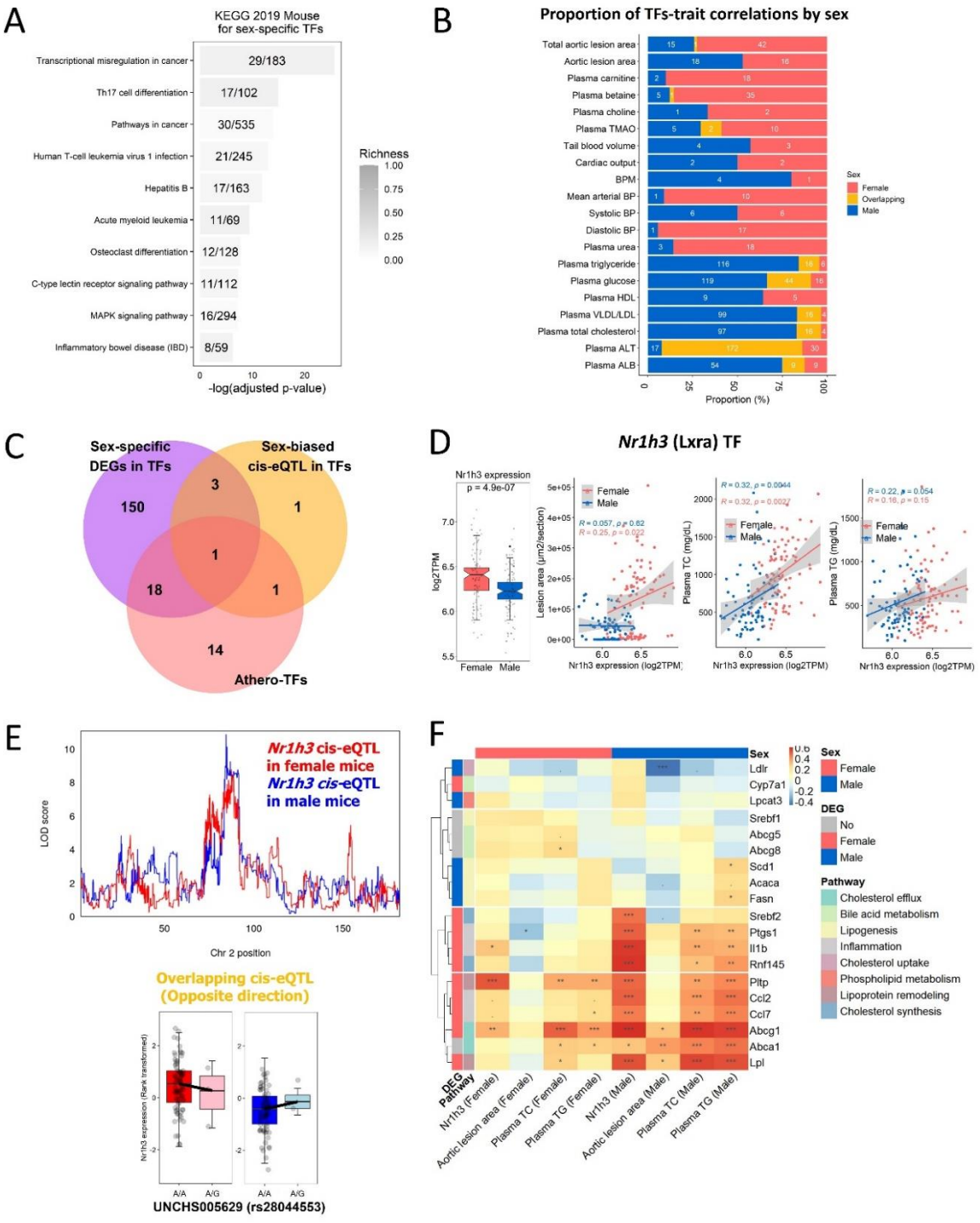


Figure 4.13. Identification of liver transcription factors with sex-specific gene expression, genetic regulation, and gene-trait correlation.
(A) Top10 KEGG pathway of liver transcription factors (TFs) that have sex-specific gene expressions identified in enrichment analysis. Pathways are ordered from top to bottom by significance (highest to lowest) and colored by gene richness. (B) Proportion of 265 TFs associated (adjusted $P < 0.05$) with cardiometabolic traits in each sex. Number in the bar is the number of TFs associated with the trait in each category. Red bar contains genes that correlate with the aortic lesion area in females, blue bar contains genes that correlate with the aortic lesion area in males, and yellow bar contains genes that correlate with the aortic lesion area in both sexes. The p-values were adjusted using the BH FDR procedure ($P<0.05$). (C) Identification of liver TFs with sex-specific gene expression, genetic regulation, and gene-trait correlation. Purple circle contains 172 TFs that have sex-specific gene expressions, orange circle contains 6 TFs that have sex-biased *cis*-eQTLs, and pink circle contains 34 TFs that correlate with the aortic lesion area (adjusted $P < 0.05$) in either sex. (D) *Nr1h3* gene (log2TPM) is significantly upregulated in females ($P<0.001$), positively correlated with aortic lesion area in females ($R = 0.25$, $P<0.05$), not males ($R = 0.057$, $P>0.05$), and positively correlated with plasma total cholesterol in both sexes ($R = 0.32$, $P<0.001$) by Spearman correlation. (E) LOD profiles on chromosome 2 highlighting a locus significantly associated with the liver *Nr1h3* gene expression in both sexes. A red line for *Nr1h3 cis*-eQTL in females (N=85) and a blue line for *Nr1h3 cis*-eQTL in males (N=77). Association between a lead SNP (rs28044553) for *Nr1h3 cis*-eQTL and liver *Nr1h3* gene expression in females (red color), and males (blue color). (F) Spearman correlation between liver *Nr1h3* gene expression, aortic lesion area, and plasma total cholesterol and triglyceride with target genes of *Nr1h3* TF in females (red color) and males (blue color). DEG status (Female: upregulated in females; Male: upregulated in males; NS: non-significant between sexes) for target genes of *Nr1h3* TF were shown along the left axis. The p-values were adjusted using the Benjamini–Hochberg (BH) FDR procedure. "***"$P < 0.001$, "**"$P < 0.01$, "*"$P < 0.05$, "."$P < 0.10$.

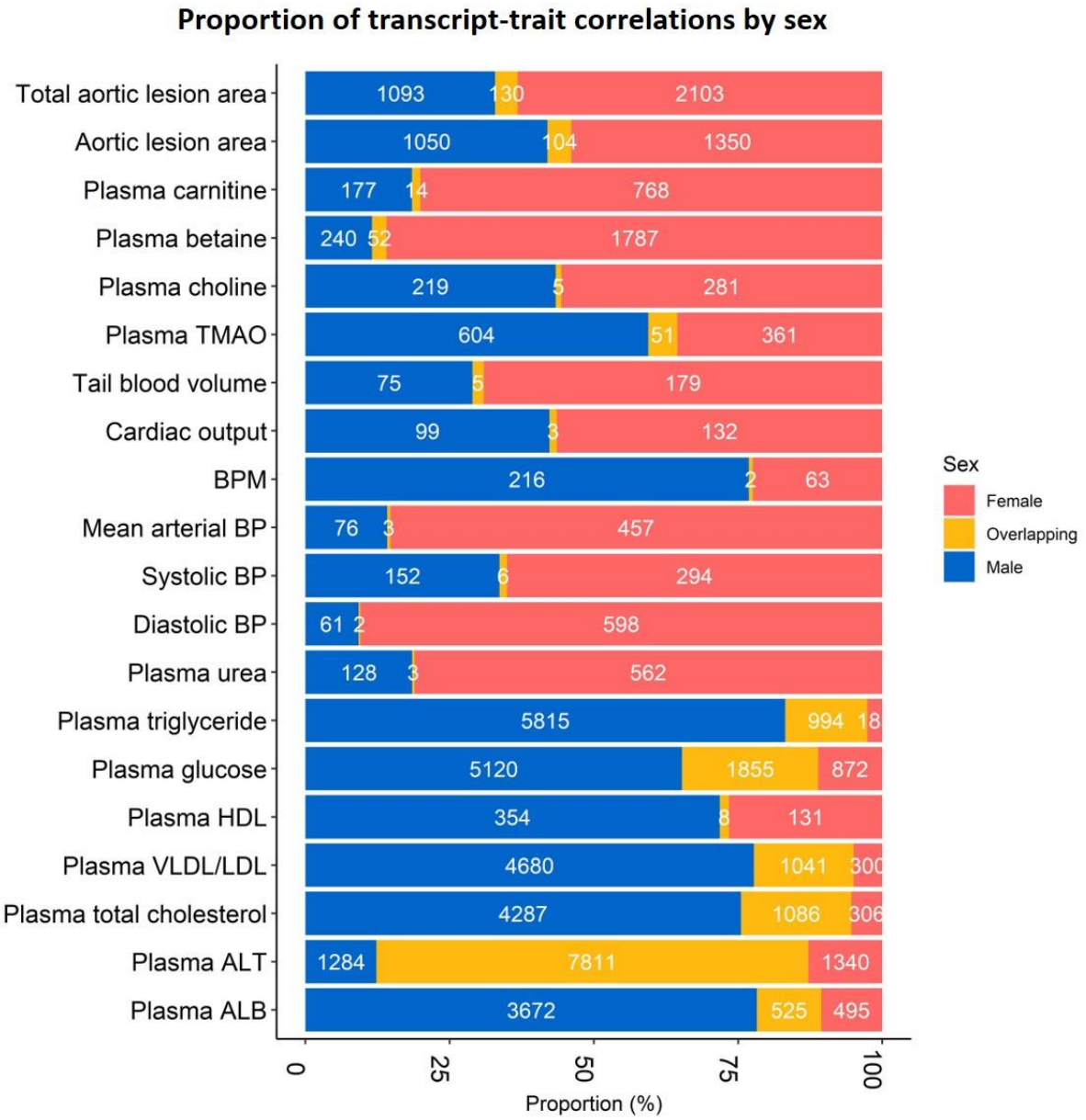


Figure 4.14. Cardiometabolic traits exhibit sex-specific interdependencies in trait-transcript correlations in DO-F1 mice.
Proportion of liver transcripts associated (adjusted $P < 0.05$) with cardiometabolic traits in each sex. Number in the bar is the number of genes associated with the trait in each category. Red bar contains genes that correlate with the aortic lesion area in females, blue bar contains genes that correlate with the aortic lesion area in males, and yellow bar contains genes that correlate with the aortic lesion area in both sexes. The p-values were adjusted using the BH FDR procedure.

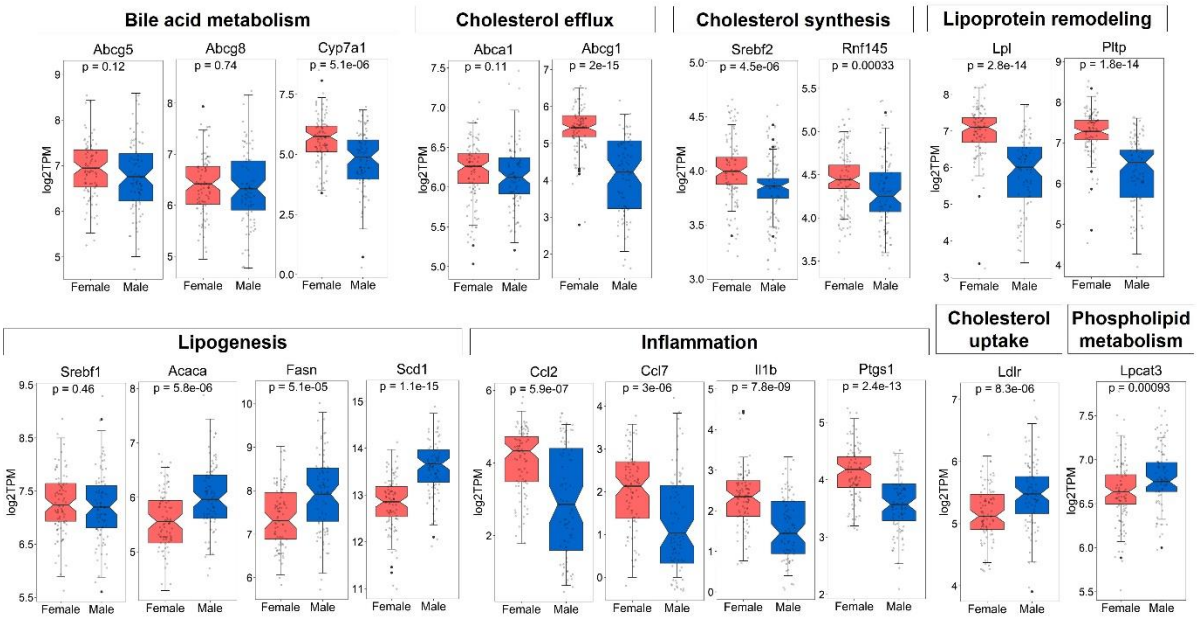
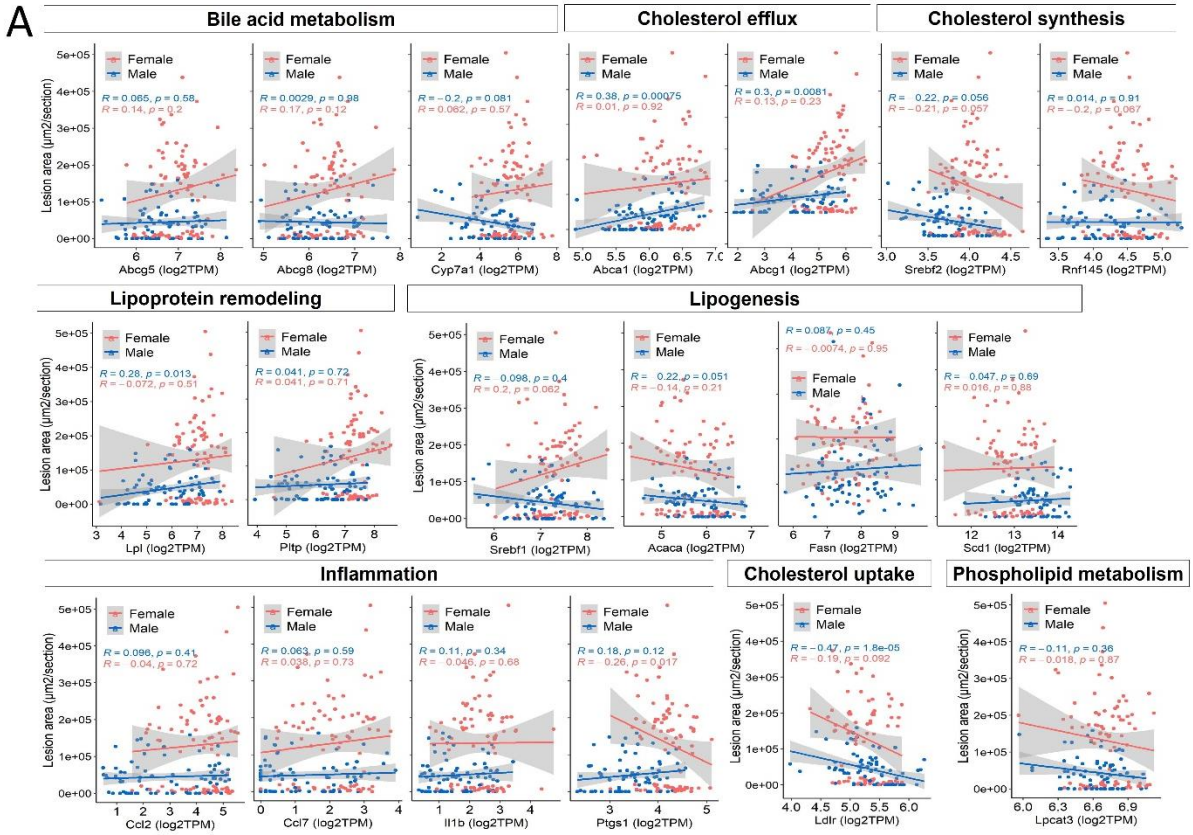


Figure 4.15. Sex-specific gene expressions for target genes of Lxr alpha transcription factor encoded by *Nr1h3*

Expressions of 19 target genes (log2TPM) of Lxr alpha transcription factor encoded by *Nr1h3* between sexes. These genes are either direct targets of Lxr alpha or induced by activation of Lxr alpha and are involved in the following pathways; bile acid metabolism (*Abcg5*, *Abcg8*, and *Cyp7a1*), cholesterol efflux (*Abca1* and *Abcg1*), cholesterol synthesis (*Srebf2* and *Rnf145*), lipoprotein remodeling (*Lpl* and *Pltp*), lipogenesis (*Srebf1*, *Acaca*, *Fasn*, and Scd1), inflammation (*Ccl2*, *Ccl7*, *Il1b*, and *Ptgs1*), cholesterol uptake (*Ldlr*), phospholipid metabolism (*Lpcat3*). The p-values were confirmed by Wilcoxon signed-rank test.


A
Bile acid metabolism
Cholesterol efflux
Cholesterol synthesis
Lipoprotein remodeling
Lipogenesis
Inflammation
Cholesterol uptake
Phospholipid metabolism
Female
Male
Lesion area (µm2/section)
Abcg5 (log2TPM)
Abcg8 (log2TPM)
Cyp7a1 (log2TPM)
Abca1 (log2TPM)
Abcg1 (log2TPM)
Srebf2 (log2TPM)
Rnf145 (log2TPM)
Lpl (log2TPM)
Pltp (log2TPM)
Srebf1 (log2TPM)
Acaca (log2TPM)
Fasn (log2TPM)
Scd1 (log2TPM)
Ccl2 (log2TPM)
Ccl7 (log2TPM)
Il1b (log2TPM)
Ptgs1 (log2TPM)
Ldlr (log2TPM)
Lpcat3 (log2TPM)

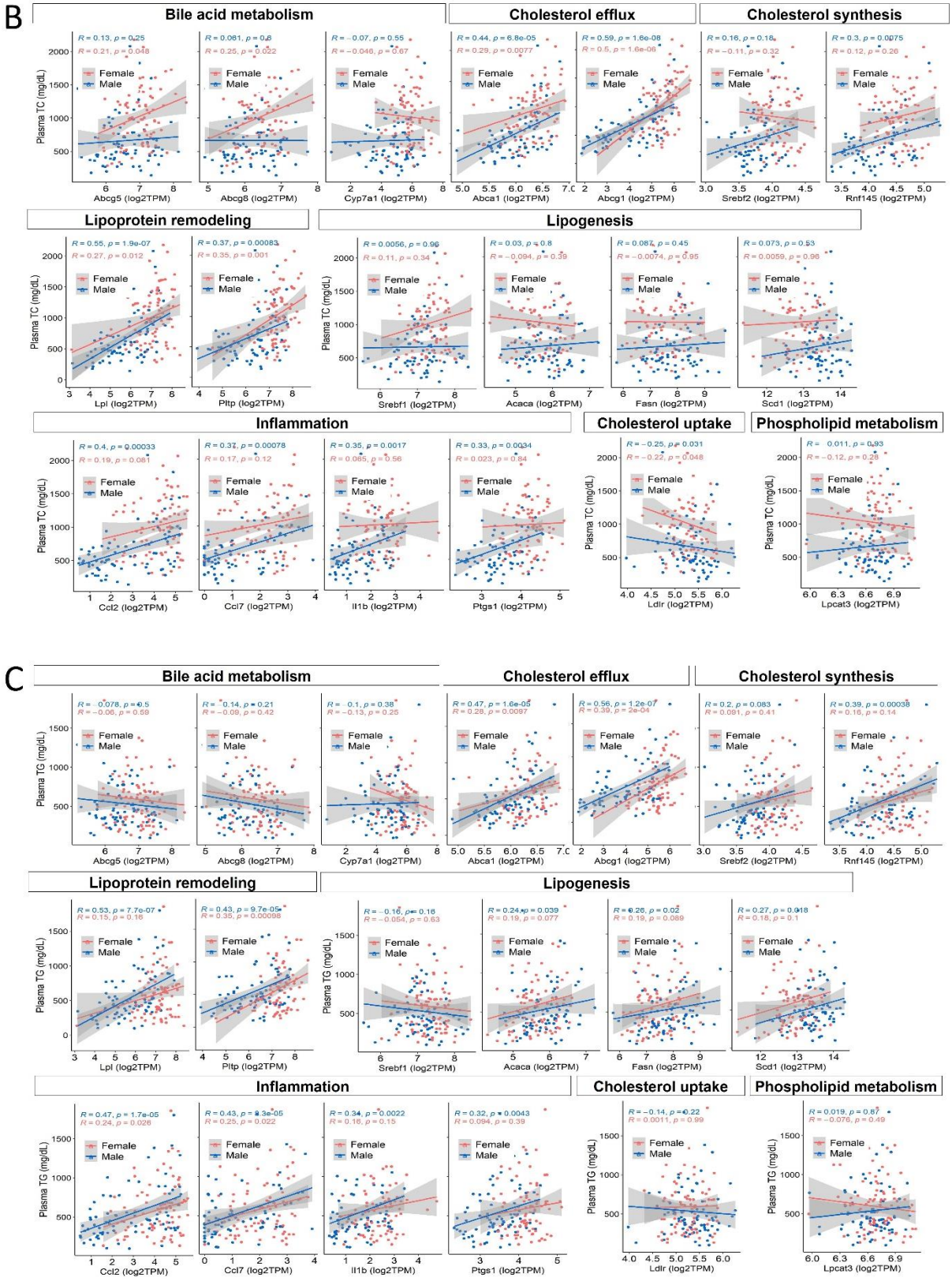

B
Bile acid metabolism
Cholesterol efflux
Cholesterol synthesis
Lipoprotein remodeling
Lipogenesis
Inflammation
Cholesterol uptake
Phospholipid metabolism
Plasma TC (mg/dL)
Female
Male
Abcg5 (log2TPM)
Abcg8 (log2TPM)
Cyp7a1 (log2TPM)
Abca1 (log2TPM)
Abcg1 (log2TPM)
Srebf2 (log2TPM)
Rnf145 (log2TPM)
Lpl (log2TPM)
Pltp (log2TPM)
Srebf1 (log2TPM)
Acaca (log2TPM)
Fasn (log2TPM)
Scd1 (log2TPM)
Ccl2 (log2TPM)
Ccl7 (log2TPM)
Il1b (log2TPM)
Ptgs1 (log2TPM)
Ldlr (log2TPM)
Lpcat3 (log2TPM)
C
Bile acid metabolism
Cholesterol efflux
Cholesterol synthesis
Lipoprotein remodeling
Lipogenesis
Inflammation
Cholesterol uptake
Phospholipid metabolism
Plasma TG (mg/dL)
Female
Male
Abcg5 (log2TPM)
Abcg8 (log2TPM)
Cyp7a1 (log2TPM)
Abca1 (log2TPM)
Abcg1 (log2TPM)
Srebf2 (log2TPM)
Rnf145 (log2TPM)
Lpl (log2TPM)
Pltp (log2TPM)
Srebf1 (log2TPM)
Acaca (log2TPM)
Fasn (log2TPM)
Scd1 (log2TPM)
Ccl2 (log2TPM)
Ccl7 (log2TPM)
Il1b (log2TPM)
Ptgs1 (log2TPM)
Ldlr (log2TPM)
Lpcat3 (log2TPM)

Figure 4.16. Sex-specific gene-trait correlations for target genes of Lxr alpha transcription factor encoded by *Nr1h3*

(A) Spearman correlation between aortic lesion area and 19 target genes (log2TPM) of Lxr alpha transcription factor in each sex. (B) Spearman correlation between plasma total cholesterol and 19 target genes (log2TPM) of Lxr alpha transcription factor in each sex. (C) Spearman correlation between plasma total triglyceride and 19 target genes (log2TPM) of Lxr alpha transcription factor in each sex.

4.6. Discussion

In this study, we observed a significant effect of sex on atherosclerosis and cardiometabolic traits and investigated how sex and gene expression interact to affect these traits. Several conclusions have been drawn from this study. First, we identified predominant sexual dimorphism in the cardiometabolic traits and transcripts and gene-by-sex interaction in the liver. Second, we identified sex-biased genetic regulation of liver transcripts, and a number of these transcripts are associated with cardiometabolic traits. Third, high-resolution mapping in hyperlipidemic DO-F1 mice uncovered sex-specific loci for cardiometabolic traits. Our study provides a higher resolution of aortic lesion area QTLs than the previous study that identified loci on chromosome 10 (Dansky et al., 2002). Fourth, integration of sex-specific gene expression, genetic regulation, and gene-trait correlation discovered multiple sex-biased liver TFBSs associated with atherosclerosis in DO-F1 mice. These results suggest a prominent role of sex in the modulation of atherosclerotic traits and provide new insights into liver sex differences, and may help explain the risk of sex differences in CVD. Each of these points is discussed in turn below.

Sexual dimorphism of atherosclerosis has been observed in various experimental models such as $ApoE^{-/-}$ and $Ldlr^{-/-}$ mice, pigs, and rabbits (Fisher et al., 1967; Freeman et al., 2007; Matthan et al., 2018). Previous studies have also found sexual dimorphism of atherosclerosis in F2 genetic crosses or >100 different inbred mouse strains (Bennett et al., 2015; Su et al., 2006) but have not been studied in the DO-F1 mouse model. Our study reveals profound differences in lesion size, TC, TG, and plasma glucose, TMAO, choline, and betaine concentrations between sexes at 24 weeks, suggesting that females were more susceptible to atherosclerosis than males in the DO-F1 study. Thus, these results imply that therapeutic effects to disease may not work

equally well in all host genotypes or both sexes, and that a personalized approach that takes into account host genotypes, sex, and their interactions may be advantageous.

Sex differences in pathophysiology and disease risk are characterized by many tissues, including the liver (Yang et al., 2006). A large number of sex-biased genes have been identified in mouse liver (Yang et al., 2006). However, previous global gene expression studies that performed RNA-seq in mice mostly have used inbred mice, and lack sufficient statistical power to identify a large number of sex-biased genes. Our study identified more than 8,000 sex-specific DEGs in both autosome and X-chromosome, which suggests the effects of sex on genome-wide regulatory mechanisms. Such sex differences may be biologically relevant, especially when multiple genes within a specific pathway are affected (Subramanian et al., 2005). Using this approach, we have identified a number of liver genes that exhibit sex-biased expression involved in a wide range of critically important biological processes. Specifically, the abundance of genes involved in immune function in the liver of females was higher than in males while genes involved in mitochondrial biogenesis were more abundant in livers from males than females. Surprisingly, despite markedly higher TC concentrations in plasma, gene pathways involved in cholesterol synthesis were not identified in female-specific DEGs. These results are consistent with a previous study that sex differences in plasma lipids were not related to changes in liver gene abundance for the cholesterol synthesis pathway (Link et al., 2015).

Few studies have highlighted the sex specificity of gene expression and their genetic regulation in humans. Previous studies in human blood samples, lymphoblastoid cell lines, or the GTEx found autosome and chromosome X variants (Dimas et al., 2012; Kukurba et al., 2016; Yao et al., 2014). eQTL analyses in mouse models either have identified that eQTLs can be sex-dependent or specific (Norheim et al., 2019; Yang et al., 2006). The current study showed

genotype-by-sex interactions in DO-F1 liver tissue. Consistent with the previous finding, these associations can be classified as sex-specific eQTLs or overlapping eQTLs between sexes with directionalities (Yao et al., 2014). In addition, our study revealed that although thousands of genes were identified as sex-specific DEGs in the liver, the proportion of genes identified as sex-specific eQTL is relatively small.

Identification of sex-specific QTL demonstrates the complexity of atherosclerosis in DO-F1 mice. Sex differences in aortic lesion area and sex-specific QTLs have previously been reported in F2 cross studies (Teupser et al., 2006; Wang et al., 2007). The cause of sexual dimorphism in atherosclerosis susceptibility is likely that sex hormones may play a role, but it has not been fully elucidated. Our results reveal that aortic lesion area QTLs identified in DO-F1 mice show sex bias. The sex specificity of these QTLs reiterates that the genetic regulation of atherosclerosis is complex and that the expression of candidate genes may differ between sexes. This has important implications for developing therapies for atherosclerosis in men and women as the underlying genetic effects and their genes they affect may differ between sexes. In this regard, a recent systematic meta-analysis for observational studies performed on CVD patients revealed sex differences in prescribing cardiovascular medications (Zhao et al., 2020). This analysis included a total of 43 studies involving 2,264,600 participants (28% women) worldwide and reported that women were less likely to be prescribed angiotensin-converting enzyme inhibitors, but more likely with diuretics than men. This study, in turn, shows that sex differences in cardiovascular drug prescription persist in primary care, suggesting that a mechanism study is needed to elucidate the interaction between sex and medications for personalized treatment of atherosclerosis.

Genetic regulation of liver transcriptome may also contribute to complex traits, and colocalization analysis between *cis*-eQTLs and sexually dimorphic atherosclerosis can reveal sex-specific genetic variant-gene-atherosclerosis associations. One of the main goals of our study was to discover novel genes that influence atherosclerosis, and we hypothesized that these genes would reveal new pathways and mechanisms. To identify robust candidate genes associated with atherosclerosis, we considered allele effects, *cis*-eQTL, and correlation with lesion size. Since 8 alleles are segregated in DO mice, high-resolution mapping allowed us to investigate whether each allele was associated with lesion size. One notable gene was in our candidate gene list, identified as having *cis*-eQTL colocalized with atherosclerosis QTL, and also correlated with aortic lesion area. *Pten*, a candidate with overlapping *cis*-eQTL and atherosclerosis QTL, is a tumor suppressor gene that expressed in endothelial cells, sub-endothelial cells, and vascular smooth muscle cells (Lu et al., 2020). *Pten* has been shown to influence the development of atherosclerosis in many studies. In mice, PTEN overexpression reduces plaque area and preserves contractile protein expression in smooth muscle cells (SMC) in atherosclerosis (Lu et al., 2020). PTEN deletion promotes spontaneous vascular remodeling in SMC and PTEN loss correlates with increased atherosclerotic lesion severity in human coronary arteries (Moulton et al., 2018). This negative causal relationship between PTEN and atherosclerosis revealed via the *Pten* gene perturbation is consistent with the correlation analysis in the DO-F1 cross. In our analysis, *Pten* gene and lesion size showed a significant correlation only in males, which is also consistent with previous reports, considering that most functional validations were performed in male mice. Taken together, these findings and reports provide strong evidence that *Pten* may have a pivotal role in atherosclerosis development in mice and humans.

Uncovering the mechanistic basis for sex differences in liver TFs could improve our understanding in precision medicine for atherosclerosis (Lopes-Ramos et al., 2020; Oliva et al., 2020). The integration of liver TFBSs with sex-specific gene expression and genetic regulation yielded sex-biased associations between atherosclerosis and specific TFs. Of the 265 TFs expressed in DO-F1 liver tissue, 172 TFs were upregulated in one sex, and our results not only suggest that hormone-related TFs regulate sex-biased gene expression as expected, but also implicate that additional TFs play a role in sex-biased expression. Furthermore, we identified only *Nr1h3* TF with sex-biased *cis*-eQTLs that significantly correlated with atherosclerosis. The liver X receptor alpha (LXRα), encoded by *Nr1h3* gene, is an important regulator of cholesterol, fatty acid, and glucose homeostasis (Peet et al., 1998). LXRα expression is highest in the liver among all organs, and LXRα target genes, ATP binding cassette (ABC) transporters A1/G1, apolipoprotein E/CI, and members of the Cyp7a family are also highly expressed in liver in both human and mouse (Edwards et al., 2002; Peet et al., 1998). Numerous studies have shown that LXRα is involved in many pathways underlying the development of atherosclerosis and CVD, including lipid metabolism, innate immunity and inflammation (Calkin and Tontonoz, 2010; Joseph et al., 2003). In addition, data from the Phenotype-Genotype Integrator in NCBI (https://www.ncbi.nlm.nih.gov/gap/phegeni) demonstrate that SNPs in the LXRα's target genes are associated with various CVD-related traits across approximately 500,000 individuals (**Table 4.14**). Furthermore, recent studies have shown that LXRα is responsible for the sex differences in lipid metabolism and circadian pattern of plasma corticosterone (Feillet et al., 2016; Jiang et al., 2016). Our study showed that key lipid processing genes, *Abcg1* and *Lpl*, among the target genes of LXRα, were differentially expressed between sexes as well as a strong positive correlation with atherosclerosis in one sex. Notably, sex-biased TF targeting of genes is independent of sex-

specific expression of TFs themselves (Carrel and Willard, 2005). However, if these differences occur at the time point of early development and translate into a more constitutive sex-biased TF binding profile, then the sex-biased expression of TFs may affect the sex-specificity of TF target genes (Leuenberger et al., 2009). Taken together, these sex-based integrated analyzes may help elucidate the origin of the gene-atherosclerosis association, as hypothesized for the effects of candidate sex-specific TFs on sex differences in atherosclerosis.

This is the first comprehensive study of sexual dimorphism of the cardiometabolic traits and liver transcriptome in DO-F1 mice. More than 8,000 genes were found to show significant sex differences in expression and influence sex-specific functional pathways. These results increase our understanding of sex-biased liver biology at the molecular level and provide important insights into our understanding of sex-specific atherosclerosis susceptibility. By understanding how sex differences influence molecular pathways, we can identify novel factors that ultimately affect disease susceptibility and suggest diagnostic and clinical strategies as translational regulation and post-translational modifications. Further studies of sex-biased liver genes at the genetic, regulatory (including translational regulation and post-translational modifications), and functional levels will enhance our understanding of liver physiology and its role in disease states. Overall, our results provide strong evidence for why males and females should be distinguished in biological research, and why females are more susceptible to atherosclerosis than males in the hyperlipidemic outbred mice model.

Table 4.14. Phenome wide association (PheWAS) for Nr1h3 transcription factor and its target genes.

Phenotypes (Human)	SNP_rs (Human)	Context	Gene	Chr	Position (bp, Human)	P_Value ($p<1\times10^{-7}$)	Population
Cholesterol	1883025	intron	ABCA1	9	104902020	6.00E-53	European
Cholesterol	2575876	intron	ABCA1	9	104903458	1.00E-16	East Asian
Cholesterol	2740488	intron	ABCA1	9	104899461	2.00E-22	European
Cholesterol, HDL	1883025	intron	ABCA1	9	104902020	2.00E-65	European
Cholesterol, HDL	2472386	intron	ABCA1	9	104839260	8.00E-11	Hispanic
Cholesterol, HDL	2575876	intron	ABCA1	9	104903458	4.00E-13	East Asian
Cholesterol, HDL	2740488	intron	ABCA1	9	104899461	2.00E-34	European
Cholesterol, HDL	3890182	intron	ABCA1	9	104885374	3.00E-10	NR
Cholesterol, HDL	3905000	intron	ABCA1	9	104894789	9.00E-13	European
Cholesterol, HDL	4149268	intron	ABCA1	9	104884939	1.00E-10	European
Cholesterol, HDL	7024300	intron	ABCA1	9	104827286	1.00E-08	Arab
Cholesterol, HDL	9282541	missense	ABCA1	9	104858554	6.00E-26	Hispanic
Cholesterol, HDL	12686004	intron	ABCA1	9	104891145	2.00E-18	East Asian
Lipids	2575876	intron	ABCA1	9	104903458	2.00E-11	European
Lipoproteins	2575876	intron	ABCA1	9	104903458	3.00E-15	European
Metabolic Syndrome X	1883025	intron	ABCA1	9	104902020	6.00E-10	European
Platelet Function Tests	2066717	intron	ABCA1	9	104829197	7.20E-08	European
Vascular Calcification	4149310	intron	ABCA1	9	104826853	2.00E-10	European
Cholesterol	2081687	intergenic	CYP7A1	8	58476006	9.00E-13	European
Cholesterol	4738684	intergenic	CYP7A1	8	58480714	3.00E-11	European
Cholesterol, LDL	2081687	intergenic	CYP7A1	8	58476006	4.00E-09	European
Cholesterol, LDL	9297994	intergenic	CYP7A1	8	58479765	2.00E-11	European
Atrial Fibrillation	765547	intergenic	LPL	8	20008763	2.00E-44	European
Blood Pressure	15285	UTR-3	LPL	8	19967156	1.00E-10	European
Blood Pressure	765547	intergenic	LPL	8	20008763	3.00E-51	European
Cholesterol	765547	intergenic	LPL	8	20008763	3.00E-51	European
Cholesterol, HDL	325	intron	LPL	8	19961817	8.00E-26	European
Cholesterol, HDL	326	intron	LPL	8	19961928	1.00E-08	African|African American|Hispanic
Cholesterol, HDL	328	STOP-GAIN	LPL	8	19962213	9.00E-23	NR
Cholesterol, HDL	13702	UTR-3	LPL	8	19966981	1.00E-16	European
Cholesterol, HDL	765547	intergenic	LPL	8	20008763	3.00E-51	European
Cholesterol, HDL	2083637	intergenic	LPL	8	20007664	6.00E-18	European
Cholesterol, HDL	10096633	intergenic	LPL	8	19973410	2.00E-09	African|African American
Cholesterol, HDL	10503669	intergenic	LPL	8	19990179	8.00E-43	East Asian
Cholesterol, HDL	12678919	intergenic	LPL	8	19986711	1.00E-149	European
Cholesterol, HDL	17091905	intergenic	LPL	8	19992246	6.00E-12	European
Cholesterol, HDL	17482753	intergenic	LPL	8	19975135	3.00E-11	European
Cholesterol, HDL	28526159	intergenic	LPL	8	20033716	8.00E-09	Hispanic
Cholesterol, HDL	115849089	intergenic	LPL	8	20054859	2.00E-63	European
Coronary Disease	765547	intergenic	LPL	8	20008763	2.00E-44	European
Electrocardiography	765547	intergenic	LPL	8	20008763	3.00E-51	European
Heart Failure	765547	intergenic	LPL	8	20008763	2.00E-44	European
Lipids	331	intron	LPL	8	19962894	1.00E-17	European
Lipids	1059611	UTR-3	LPL	8	19967052	1.00E-20	European
Lipids	115849089	intergenic	LPL	8	20054859	4.00E-15	European
Lipoproteins	75835816	intergenic	LPL	8	20028002	4.00E-14	European
Lipoproteins	115849089	intergenic	LPL	8	20054859	2.00E-25	European
Metabolic Syndrome X	268	missense	LPL	8	19956018	2.00E-12	European
Metabolic Syndrome X	295	intron	LPL	8	19958727	2.00E-09	European
Metabolic Syndrome X	301	intron	LPL	8	19959423	3.00E-11	European
Metabolic Syndrome X	1441756	intergenic	LPL	8	20010875	3.00E-08	European
Metabolic Syndrome X	2083637	intergenic	LPL	8	20007664	2.00E-10	South Asian|Asian
Metabolic Syndrome X	2197089	intergenic	LPL	8	19968862	2.00E-09	European
Metabolic Syndrome X	7841189	intergenic	LPL	8	19987865	1.00E-14	European
Stroke	765547	intergenic	LPL	8	20008763	2.00E-44	European

Triglycerides	326	intron	LPL	8	19961928	5.00E-12	South Asian|Asian|European|Hispanic
Triglycerides	328	STOP-GAIN	LPL	8	19962213	2.00E-28	NR
Triglycerides	13702	UTR-3	LPL	8	19966981	1.00E-16	European
Triglycerides	15285	UTR-3	LPL	8	19967156	1.00E-10	European
Triglycerides	765547	intergenic	LPL	8	20008763	3.00E-51	European
Triglycerides	1441759	intergenic	LPL	8	20008052	2.00E-09	NR
Triglycerides	4244457	intergenic	LPL	8	20041535	1.00E-10	East Asian
Triglycerides	9644568	intergenic	LPL	8	20071071	4.00E-11	Hispanic
Triglycerides	10096633	intergenic	LPL	8	19973410	2.00E-18	European
Triglycerides	10105606	intergenic	LPL	8	19970337	4.00E-26	European
Triglycerides	10503669	intergenic	LPL	8	19990179	7.00E-39	East Asian
Triglycerides	12678919	intergenic	LPL	8	19986711	2.00E-199	European
Triglycerides	17091905	intergenic	LPL	8	19992246	5.00E-15	European
Triglycerides	17410962	intergenic	LPL	8	19990569	7.00E-09	African|African American|Hispanic
Triglycerides	17482753	intergenic	LPL	8	19975135	6.00E-10	European
Triglycerides	79236614	intergenic	LPL	8	20002949	7.00E-09	Hispanic
Triglycerides	115849089	intergenic	LPL	8	20054859	9.00E-84	European
Vascular Calcification	12678919	intergenic	LPL	8	19986711	9.00E-13	European
Cholesterol, HDL	2167079	UTR-5	NR1H3	11	47248704	5.13E-08	Northern Finland Birth Cohort 1966 (NFBC1966)
Cholesterol, HDL	7120118	intron	NR1H3	11	47264739	3.57E-08	Northern Finland Birth Cohort 1966 (NFBC1966)
Cholesterol, HDL	7120118	intron	NR1H3	11	47264739	4.00E-08	European
Metabolic Syndrome X	10838681	intron	NR1H3	11	47253513	1.00E-09	European
Cholesterol, HDL	6065906	intergenic	PLTP	20	45925376	5.00E-40	European
Lipids	6065904	intron	PLTP	20	45906012	4.00E-40	European
Lipids	6065906	intergenic	PLTP	20	45925376	3.00E-49	European
Lipoproteins	6073958	intergenic	PLTP	20	45923216	3.00E-55	European
Vascular Calcification	4810479	intergenic	PLTP	20	45916409	2.00E-42	European
Triglycerides	6065906	intergenic	PLTP	20	45925376	2.00E-34	European
Triglycerides	6073958	intergenic	PLTP	20	45923216	9.00E-21	European
Heart Rate	12941356	intron	SREBF1	17	17813217	5.00E-08	Asian|African|European|Mixed
Cholesterol	4148191	intron	ABCG5	2	43815765	4.00E-09	NR
Cholesterol	6756629	missense	ABCG5	2	43837951	2.00E-11	European
Cholesterol, LDL	6756629	missense	ABCG5	2	43837951	3.00E-10	European
Cholesterol, LDL	11887534	nearGene-5	ABCG5	2	43839108	9.00E-33	NR
Lipoproteins	6756629	missense	ABCG5	2	43837951	1.00E-14	European
Cholesterol	4299376	intron	ABCG8	2	43845437	3.00E-73	European
Cholesterol	6756629	nearGene-5	ABCG8	2	43837951	2.00E-11	European
Cholesterol	76866386	intron	ABCG8	2	43848344	6.00E-33	European
Cholesterol, LDL	4299376	intron	ABCG8	2	43845437	4.00E-72	European
Cholesterol, LDL	6544713	intron	ABCG8	2	43846742	2.00E-20	European
Cholesterol, LDL	6756629	nearGene-5	ABCG8	2	43837951	3.00E-10	European
Cholesterol, LDL	11887534	missense	ABCG8	2	43839108	9.00E-33	NR
Cholesterol, LDL	72875462	intron	ABCG8	2	43852171	2.00E-35	European
Lipids	4245791	intron	ABCG8	2	43847292	1.00E-09	African|African American|Hispanic
Lipids	4299376	intron	ABCG8	2	43845437	2.00E-08	European
Lipoproteins	6756629	nearGene-5	ABCG8	2	43837951	1.00E-14	European
Aortic Aneurysm, Abdominal	6511720	intron	LDLR	19	11091630	8.00E-14	European
Atherosclerosis	6511720	intron	LDLR	19	11091630	1.00E-07	European
Cholesterol	2228671	cds-synon	LDLR	19	11100236	9.00E-24	European
Cholesterol	6511720	intron	LDLR	19	11091630	5.00E-202	European
Cholesterol	112374545	intergenic	LDLR	19	11078223	2.00E-113	European
Cholesterol, LDL	2228671	cds-synon	LDLR	19	11100236	4.00E-14	European
Cholesterol, LDL	2738446	intron	LDLR	19	11116650	2.00E-12	East Asian
Cholesterol, LDL	6511720	intron	LDLR	19	11091630	4.00E-262	European
Cholesterol, LDL	112374545	intergenic	LDLR	19	11078223	7.00E-142	European
Coronary Artery Disease	56289821	intergenic	LDLR	19	11077571	4.00E-15	South Asian|East Asian|Asian|Arab|Afri

Lipids	6511720	intron	LDLR	19	11091630	2.00E-31	can|European|African American|Hispanic European
Lipids	17249141	nearGene-5	LDLR	19	11089332	2.00E-17	African|African American|Hispanic
Lipids	55791371	intergenic	LDLR	19	11077477	8.00E-17	European
Lipoproteins	142130958	intergenic	LDLR	19	11079976	7.00E-47	European
Myocardial Infarction	55791371	intergenic	LDLR	19	11077477	3.00E-08	South Asian|East Asian|Asian|Arab| African|European| American|Hispanic
Plaque, Atherosclerotic	6511720	intron	LDLR	19	11091630	8.25E-08	European
Vascular Calcification	6511720	intron	LDLR	19	11091630	4.00E-09	European
Fatty Acids	12580543	intron	LPCAT3	12	6992881	3.00E-30	NR
Mean Platelet Volume	10076782	intron	RNF145	5	159177955	4.00E-08	European

4.7. Acknowledgments

Mice carrying the human ApoE3 Leiden variant were kindly provided by Dr. Lusis, University of California, Los Angeles, CA, USA. This research was supported in part by NIH grant 5R01HL128572 (B.J.B), USDA project 2032-51530-025-00D (B.J.B).

4.8. Declaration of Interests

The authors declare no conflicts of interest.

4.9. Reference

AlSiraj, Y., Chen, X., Thatcher, S.E., Temel, R.E., Cai, L., Blalock, E., Katz, W., Ali, H.M., Petriello, M., Deng, P., et al. (2019). XX sex chromosome complement promotes atherosclerosis in mice. Nat Commun *10*, 2631.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - A Practical and Powerful Approach to Multiple Testing. J. R. Statst. Soc. B *57*, 289-300.

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

Blum, A., and Blum, N. (2009). Coronary artery disease: Are men and women created equal? Gend Med *6*, 410-418.

Broman, K.W. (2012a). Genotype probabilities at intermediate generations in the construction of recombinant inbred lines. Genetics *190*, 403-412.

Broman, K.W. (2012b). Haplotype probabilities in advanced intercross populations. G3 (Bethesda) *2*, 199-202.

Broman, K.W., Gatti, D.M., Simecek, P., Furlotte, N.A., Prins, P., Sen, Ś., Yandell, B.S., and Churchill, G.A. (2019). R/qtl2 - Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multi-parent Populations. Genetics *211*, 495-502.

Bubb, K.J., Khambata, R.S., and Ahluwalia, A. (2012). Sexual dimorphism in rodent models of hypertension and atherosclerosis. Br J Pharmacol *167*, 298-312.

Calkin, A.C., and Tontonoz, P. (2010). Liver x receptor signaling pathways and atherosclerosis. Arterioscler Thromb Vasc Biol *30*, 1513-1518.

Carrel, L., and Willard, H.F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature *434*, 400-404.

Charlesworth, B. (1996). The evolution of chromosomal sex determination and dosage compensation. Curr Biol *6*, 149-162.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr - interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *15*, 128.

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

Clodfelter, K.H., Holloway, M.G., Hodor, P., Park, S.H., Ray, W.J., and Waxman, D.J. (2006). Sex-dependent liver gene expression is extensive and largely dependent upon signal transducer and activator of transcription 5b (STAT5b): STAT5b-dependent activation of male genes and repression of female genes revealed by microarray analysis. Mol Endocrinol *20*, 1333-1351.

Consortium, G. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318-1330.

Dansky, H.M., Shu, P., Donavan, M., Montagno, J., Nagle, D.L., Smutko, J.S., Roy, N., Whiteing, S., Barrios, J., McBride, T.J., et al. (2002). A phenotype-sensitizing Apoe-deficient genetic background reveals novel atherosclerosis predisposition loci in the mouse. Genetics *160*, 1559-1608.

Dimas, A.S., Nica, A.C., Montgomery, S.B., Stranger, B.E., Raj, T., Buil, A., Giger, T., Lappalainen, T., Gutierrez-Arcelus, M., Mu, T.C., et al. (2012). Sex-biased genetic effects on gene regulation in humans. Genome Res *22*, 2368-2375.

Disteche, C.M. (2012). Dosage Compensation of the Sex Chromosomes. Annual Review of Genetics *46*, 537-560.
Doerge, R.W., and Churchill, G.A. (1996). Permutation tests for multiple loci affecting a quantitative character. Genetics *142*, 285-294.
Edwards, P.A., Kennedy, M.A., and Mak, P.A. (2002). LXRs; oxysterol-activated nuclear receptors that regulate genes controlling lipid homeostasis. Vascul Pharmacol *38*, 249-256.
Fairweather, D. (2014). Sex differences in inflammation during atherosclerosis. Clin Med Insights Cardiol *8*, 49-59.
Fairweather, D., Petri, M.A., Coronado, M.J., and Cooper, L.T. (2012). Autoimmune heart disease: role of sex hormones and autoantibodies in disease pathogenesis. Expert Rev Clin Immunol *8*, 269-284.
Feillet, C., Guerin, S., Lonchampt, M., Dacquet, C., Gustafsson, J.A., Delaunay, F., and Teboul, M. (2016). Sexual Dimorphism in Circadian Physiology Is Altered in LXRalpha Deficient Mice. PLoS One *11*, e0150665.
Feng, M., Whitesall, S., Zhang, Y., Beibel, M., Alecy, L.D., and DiPetrillo, K. (2008). Validation of Volume-Pressure Recording Tail-Cuff Blood Pressure Measurements. American Journal of Hypertension *21*, 1288-1291.
Fisher, H., Griminger, P., and Siller, W.G. (1967). Effect of pectin on atherosclerosis in the cholesterol-fed rabbit. J. Atheroscler. Res. *7*, 381-386.
Freeman, L., Amar, M.J., Shamburek, R., Paigen, B., Brewer, H.B., Jr., Santamarina-Fojo, S., and Gonzalez-Navarro, H. (2007). Lipolytic and ligand-binding functions of hepatic lipase protect against atherosclerosis in LDL receptor-deficient mice. J Lipid Res *48*, 104-113.
Gatti, D.M., Svenson, K.L., Shabalin, A., Wu, L.Y., Valdar, W., Simecek, P., Goodwin, N., Cheng, R., Pomp, D., Palmer, A., et al. (2014). Quantitative trait locus mapping methods for diversity outbred mice. G3 (Bethesda) *4*, 1623-1633.
Hager, R., Cheverud, J.M., Leamy, L.J., and Wolf, J.B. (2008). Sex dependent imprinting effects on complex traits in mice. BMC Evol Biol *8*, 303.
Isensee, J., Witt, H., Pregla, R., Hetzer, R., Regitz-Zagrosek, V., and Noppinger, P.R. (2008). Sexually dimorphic gene expression in the heart of mice and men. J Mol Med (Berl) *86*, 61-74.
Jiang, Z., Huang, X., Huang, S., Guo, H., Wang, L., Li, X., Huang, X., Wang, T., Zhang, L., and Sun, L. (2016). Sex-Related Differences of Lipid Metabolism Induced by Triptolide: The Possible Role of the LXRalpha/SREBP-1 Signaling Pathway. Front Pharmacol *7*, 87.
Joseph, S.B., Castrillo, A., Laffitte, B.A., Mangelsdorf, D.J., and Tontonoz, P. (2003). Reciprocal regulation of inflammation and lipid metabolism by liver X receptors. Nat Med *9*, 213-219.
Kolde, R., Franzosa, E.A., Rahnavard, G., Hall, A.B., Vlamakis, H., Stevens, C., Daly, M.J., Xavier, R.J., and Huttenhower, C. (2018). Host genetic variation and its microbiome interactions within the Human Microbiome Project. Genome Med *10*, 6.
Kukurba, K.R., Parsana, P., Balliu, B., Smith, K.S., Zappala, Z., Knowles, D.A., Fave, M.J., Davis, J.R., Li, X., Zhu, X., et al. (2016). Impact of the X Chromosome and sex on regulatory variation. Genome Res *26*, 768-777.

Kulakovskiy, I.V., Vorontsov, I.E., Yevshin, I.S., Sharipov, R.N., Fedorova, A.D., Rumynskiy, E.I., Medvedeva, Y.A., Magana-Mora, A., Bajic, V.B., Papatsenko, D.A., et al. (2018). HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. Nucleic Acids Res *46*, D252-D259.

Leuenberger, N., Pradervand, S., and Wahli, W. (2009). Sumoylated PPARalpha mediates sex-specific gene repression and protects the liver from estrogen-induced toxicity in mice. J Clin Invest *119*, 3138-3148.

Link, J.C., Chen, X., Prien, C., Borja, M.S., Hammerson, B., Oda, M.N., Arnold, A.P., and Reue, K. (2015). Increased high-density lipoprotein cholesterol levels in mice with XX versus XY sex chromosomes. Arterioscler Thromb Vasc Biol *35*, 1778-1786.

Lopes-Ramos, C.M., Chen, C.Y., Kuijjer, M.L., Paulson, J.N., Sonawane, A.R., Fagny, M., Platig, J., Glass, K., Quackenbush, J., and DeMeo, D.L. (2020). Sex Differences in Gene Expression and Regulatory Networks across 29 Human Tissues. Cell Rep *31*, 107795.

Lu, S., Strand, K.A., Mutryn, M.F., Tucker, R.M., Jolly, A.J., Furgeson, S.B., Moulton, K.S., Nemenoff, R.A., and Weiser-Evans, M.C.M. (2020). PTEN (Phosphatase and Tensin Homolog) Protects Against Ang II (Angiotensin II)-Induced Pathological Vascular Fibrosis and Remodeling-Brief Report. Arterioscler Thromb Vasc Biol *40*, 394-403.

Maas, A.H., and Appelman, Y.E. (2010). Gender differences in coronary heart disease. Neth Heart J. *18*, 598-602.

Man, J.J., Beckman, J.A., and Jaffe, I.Z. (2020). Sex as a Biological Variable in Atherosclerosis. Circ Res *126*, 1297-1319.

Matthan, N.R., Solano-Aguilar, G., Meng, H., Lamon-Fava, S., Goldbaum, A., Walker, M.E., Jang, S., Lakshman, S., Molokin, A., Xie, Y., et al. (2018). The Ossabaw Pig Is a Suitable Translational Model to Evaluate Dietary Patterns and Coronary Artery Disease Risk. J Nutr *148*, 542-551.

Melia, T., and Waxman, D.J. (2020). Genetic Factors Contributing to Extensive Variability of Sex-Specific Hepatic Gene Expression in Diversity Outbred Mice. bioRxiv.

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

Moulton, K.S., Li, M., Strand, K., Burgett, S., McClatchey, P., Tucker, R., Furgeson, S.B., Lu, S., Kirkpatrick, B., Cleveland, J.C., et al. (2018). PTEN deficiency promotes pathological vascular remodeling of human coronary arteries. JCI Insight *3*.

Naqvi, S., Godfrey, A.K., Hughes, J.F., Goodheart, M.L., Mitchell, R.N., and Page, D.C. (2019). Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. Science *365*.

Nedungadi, T.P., and Clegg, D.J. (2009). Sexual dimorphism in body fat distribution and risk for cardiovascular diseases. J Cardiovasc Transl Res *2*, 321-327.

Norheim, F., Hasin-Brumshtein, Y., Vergnes, L., Chella Krishnan, K., Pan, C., Seldin, M.M., Hui, S.T., Mehrabian, M., Zhou, Z., Gupta, S., et al. (2019). Gene-by-Sex Interactions in Mitochondrial Functions and Cardio-Metabolic Traits. Cell Metabolism *29*, 932-949.e934.

Oliva, M., Munoz-Aguirre, M., Kim-Hellmuth, S., Wucher, V., Gewirtz, A.D.H., Cotter, D.J., Parsana, P., Kasela, S., Balliu, B., Vinuela, A., et al. (2020). The impact of sex on gene expression across human tissues. Science *369*.

Parsch, J., and Ellegren, H. (2013). The evolutionary causes and consequences of sex-biased gene expression. Nat Rev Genet *14*, 83-87.
Peet, D.J., Turley, S.D., Ma, W., Janowski, B.A., Lobaccaro, J.M., Hammer, R.E., and Mangelsdorf, D.J. (1998). Cholesterol and Bile Acid Metabolism Are Impaired in Mice Lacking the Nuclear Oxysterol Receptor LXRα. Cell *93*, 693-704.
Pletscher-Frankild, S., Palleja, A., Tsafou, K., Binder, J.X., and Jensen, L.J. (2015). DISEASES: text mining and data integration of disease-gene associations. Methods *74*, 83-89.
Raghupathy, N., Choi, K., Vincent, M.J., Beane, G.L., Sheppard, K.S., Munger, S.C., Korstanje, R., Pardo-Manual de Villena, F., and Churchill, G.A. (2018). Hierarchical analysis of RNA-seq reads improves the accuracy of allele-specific expression. Bioinformatics *34*, 2177-2184.
Rinn, J.L., and Snyder, M. (2005). Sexual dimorphism in mammalian gene expression. Trends in Genetics *21*, 298-305.
Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47.
Rodgers, J.L., Jones, J., Bolleddu, S.I., Vanthenapalli, S., Rodgers, L.E., Shah, K., Karia, K., and Panguluri, S.K. (2019). Cardiovascular Risks Associated with Gender and Aging. J Cardiovasc Dev Dis *6*.
Roy, A.K., and Chatterjee, B. (1983). Sexual dimorphism in the liver. Annu Rev Physiol. *45*, 37-50.
Sen, S., and Churchill, G.A. (2001). A statistical framework for quantitative trait mapping. Genetics *159*, 371-387.
Su, Z., Li, Y., James, J.C., McDuffie, M., Matsumoto, A.H., Helm, G.A., Weber, J.L., Lusis, A.J., and Shi, W. (2006). Quantitative trait locus analysis of atherosclerosis in an intercross between C57BL/6 and C3H mice carrying the mutant apolipoprotein E gene. Genetics *172*, 1799-1807.
Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis - a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. *102*, 15545-15550.
Teupser, D., Tan, M., Persky, A.D., and Breslow, J.L. (2006). Atherosclerosis quantitative trait loci are sex-an ... rcross of C57BL - 6 and FVB - N low-density lipoprotein receptor-mice. Proc Natl Acad Sci U S A *103*, 123-128.
Vorobyev, A., Gupta, Y., Sezin, T., Koga, H., Bartsch, Y.C., Belheouane, M., Kunzel, S., Sina, C., Schilf, P., Korber-Ahrens, H., et al. (2019). Gene-diet interactions associated with complex trait variation in an advanced intercross outbred mouse line. Nat Commun *10*, 4097.
Wang, B., and Tontonoz, P. (2018). Liver X receptors in lipid signalling and membrane homeostasis. Nat Rev Endocrinol *14*, 452-463.
Wang, S.S., Schadt, E.E., Wang, H., Wang, X., Ingram-Drake, L., Shi, W., Drake, T.A., and Lusis, A.J. (2007). Identification of pathways for atherosclerosis in mice: integration of quantitative trait locus analysis and global gene expression data. Circ Res *101*, e11-30.
Wang, Z., Levison, B.S., Hazen, J.E., Donahue, L., Li, X.M., and Hazen, S.L. (2014). Measurement of trimethylamine-N-oxide by stable isotope dilution liquid chromatography tandem mass spectrometry. Anal Biochem *455*, 35-40.

Westerterp, M., van der Hoogt, C.C., de Haan, W., Offerman, E.H., Dallinga-Thie, G.M., Jukema, J.W., Havekes, L.M., and Rensen, P.C. (2006). Cholesteryl ester transfer protein decreases high-density lipoprotein and severely aggravates atherosclerosis in APOE*3-Leiden mice. Arterioscler Thromb Vasc Biol *26*, 2552-2559.

Winham, S.J., de Andrade, M., and Miller, V.M. (2015). Genetics of cardiovascular disease: Importance of sex and ethnicity. Atherosclerosis *241*, 219-228.

Winter, J.M., Gildea, D.E., Andreas, J.P., Gatti, D.M., Williams, K.A., Lee, M., Hu, Y., Zhang, S., Program, N.C.S., Mullikin, J.C., et al. (2017). Mapping Complex Traits in a Diversity Outbred F1 Mouse Population Identifies Germline Modifiers of Metastasis in Human Prostate Cancer. Cell Syst *4*, 31-45 e36.

Yalcin, B., Wong, K., Agam, A., Goodson, M., Keane, T.M., Gan, X., Nellaker, C., Goodstadt, L., Nicod, J., Bhomra, A., et al. (2011). Sequence-based characterization of structural variation in the mouse genome. Nature *477*, 326-329.

Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M., and Price, A.L. (2014). Advantages and pitfalls in the application of mixed-model association methods. Nat Genet *46*, 100-106.

Yang, X., Schadt, E.E., Wang, S., Wang, H., Arnold, A.P., Ingram-Drake, L., Drake, T.A., and Lusis, A.J. (2006). Tissue-specific expression and regulation of sexually dimorphic genes in mice. Genome Res *16*, 995-1004.

Yao, C., Joehanes, R., Johnson, A.D., Huan, T., Esko, T., Ying, S., Freedman, J.E., Murabito, J., Lunetta, K.L., Metspalu, A., et al. (2014). Sex- and age-interacting eQTLs in human complex diseases. Hum Mol Genet *23*, 1947-1956.

Zhang, Y., Klein, K., Sugathan, A., Nassery, N., Dombkowski, A., Zanger, U.M., and Waxman, D.J. (2011). Transcriptional profiling of human liver identifies sex-biased genes associated with polygenic dyslipidemia and coronary artery disease. PLoS One *6*, e23506.

Zhao, M., Woodward, M., Vaartjes, I., Millett, E.R.C., Klipstein-Grobusch, K., Hyun, K., Carcel, C., and Peters, S.A.E. (2020). Sex Differences in Cardiovascular Medication Prescription in Primary Care: A Systematic Review and Meta-Analysis. J Am Heart Assoc *9*, e014742.

CHAPTER 5.

Systems Genetic Analysis of Atherosclerosis and Gut Microbiota in a Hyperlipidemic Diversity Outbred F1 Mouse Population

Myungsuk Kim[1,2], M. Nazmul Huda[2], Excel Que[2], Erik R. Gertz[2], Brian J. Bennett[1,2]

[1] Department of Nutrition, University of California, Davis, CA

[2] USDA-ARS-Western Human Nutrition Research Center, Davis, CA

5.1. Author Contributions

B.J.B designed all the experiments. M.K., M.N.H., and E.R. performed the experiments. M.K., M.N.H., and E.Q. analyzed raw data. M.K. and B.J.B. wrote the manuscript, which was reviewed by all authors.

5.2. Abstract

Atherosclerosis is a precipitating event in the development of cardiovascular disease (CVD). The progression of the disease is prevalent in developed countries and there are currently limited options for prevention and treatment interventions. Recent studies report that gut microbiota contributes to the pathogenesis of cardiometabolic diseases. While host genetic architecture is a known factor that affects atherosclerosis development and gut microbial colonization, the underlying mechanisms are not yet clear. Here, we interrogated atherosclerosis regulatory networks in hyperlipidemic Diversity Outbred mice to reveal key insights into control of atherosclerosis using system genetic approaches of cardiometabolic traits, microbiome, and liver transcriptome. These include identifying the effect of genetic backgrounds on gut microbiota and fecal microbial taxa associated with atherosclerotic traits, defining the functionality of genes

associated with the atherosclerotic traits and gut microbiota, and identifying signatures of functional gene variants predicted to modulate those traits. Trans-omic analysis facilitated identification of *Ptprk* as a novel regulatory gene for atherosclerotic traits and *Lactococcus* abundance. Collectively, this study provides a rich resource for investigating the pathogenesis of atherosclerosis and suggests an opportunity to discover therapeutics and biomarkers in the setting of hyperlipidemia.

5.3. Introduction

Atherosclerosis, a major cause of coronary artery disease (CAD), is a highly complex disease caused by the interaction of a number of genetic and environmental factors. (Hartiala et al., 2017; Khera and Kathiresan, 2017; Sasidhar et al., 2014). Early evidence for the genetic cause of atherosclerosis was based on the demonstration of familial aggregation and heritability estimates (Marenberg et al., 1994; Mayer et al., 2007), which initiated a search to identify risk alleles. The advent of genome-wide association studies (GWAS) has yielded an unbiased genome-wide approach that has identified novel atherosclerosis candidate genes. To date, GWAS has the identification of over 100 different loci for CAD susceptibility (Bis et al., 2011; Franceschini et al., 2018; Nelson et al., 2017) but in aggregate, these loci only explain a fraction of the heritability of atherosclerosis and even less of the overall risk of disease (Marian, 2012). Thus, the causes and determinants of atherosclerosis remain to be elucidated and may include environmental factors including diet, the epigenome, and the microbiota (Jonsson and Backhed, 2017; Khyzha et al., 2017; Wu et al., 2019).

Recent studies have revealed that gut microbiota is an emerging contributor to human physiology and also affects the cardiovascular system (Brown and Hazen, 2018). Many studies

of human cohorts and animal models suggest that alteration of gut microbial diversity and composition influences the risk factors for cardiovascular disease (CVD) such as CAD (Emoto et al., 2016), hyperlipidemia (Fu et al., 2015), hypertension (Li et al., 2017), and heart failure (Luedde et al., 2017). In addition, several mouse and human studies have investigated the role of host genetics in shaping the composition of the gut microbiota (Kurilshikov et al., 2017). GWAS in thousands of individuals has reported genetic variants associated with gut microbiota composition (Bonder et al., 2016; Goodrich et al., 2016; Turpin et al., 2016; Wang et al., 2016). More recent efforts have successfully demonstrated the mechanisms by which the microbiota are involved in disease etiology (Fan and Pedersen, 2021; Lynn et al., 2021; Needham et al., 2020). The current main challenge is to identify the definite effect of gut dysbiosis from genetic effects and to identify how these 2 scales of biological data interact to affect disease susceptibility.

A complementary approach to studying the genetic factors of atherosclerosis and gut microbiota is to use genetic variations that occur in experimental model organisms. These studies benefit from the ability to tightly control the environment applied to the animals and detailed monitoring of clinical and molecular phenotypes allows for a precise assessment of the impact of genetic factors on each phenotype with minimal confounding factors. Quantitative trait locus (QTL) analysis has found hundreds of genetic loci associated with various clinical traits including atherosclerosis and gut microbiota. Specifically, with the recent sequencing of the Hybrid Mouse Diversity Panel (HMDP) or Diversity Outbred (DO) mice that have natural genetic variations, it has become possible to perform high-resolution mapping of complex traits (Bennett et al., 2015; Kemis et al., 2019; Lusis et al., 2016; Saul et al., 2019; Tabrett and Horton, 2020). DO mice, for example, are maintained as an outbred stock by intercrossing eight strains to capture most of the genetic diversity of inbred mouse strains: A/J, C57BL/6J, 129S1/SvImJ,

NOD/ShiLtJ, NZO/HILtJ, CAST/EiJ, PWK/PJ and WSB/EiJ. These mice are superior to other mapping populations in that they have more than 40 million SNPs and have a high average minor allele frequency and a fine recombination block structure (Churchill et al., 2012; Gatti et al., 2014).

We addressed the use of DO-F1 mice to describe the sexual dimorphism of atherosclerosis and liver transcriptome in Chapter 4. We observed strong sex-specific effects and heritability on cardiometabolic traits and liver mRNA transcripts, and identified sex-specific co-mapped locus between traits and transcripts in response to the same genetic variation. In addition, we investigated gut microbiota and liver transcriptomes from eight DO founder strains used to generate DO mouse panels, and observed differences in cardiometabolic traits, gut microbial diversity, and hepatic gene networks between diets and among the strains (Kim et al., 2021; O'Connor et al., 2014).

In this study, we characterized the microbial taxa associated with atherosclerosis in the DO-F1 mice fed a high-fat and high-cholesterol (HFHC) diet and performed QTL to identify host genetic loci associated with atherosclerotic traits, gut microbiota composition (microbiome), and molecular traits (transcriptome). We focused on searching for pleiotropic loci, which was defined as a genetic locus identified in co-localization of QTLs in trans-omics datasets. For instance, our analysis revealed a locus associated with aortic lesion area and *Lactococcus* abundance attributed to the same eight DO founder haplotypes in female mice. Genes identified in the locus were prioritized with liver gene expression QTL (eQTL) results, and we found an unknown relationship with the *Ptprk* gene that was associated with aortic lesion area and *Lactococcus* abundance in DO-F1 mice and confirm a significant association human *Ptprk* SNPs with CVD traits and dysbiosis-related diseases in human GWAS. This study demonstrates how

genetic variation affects complex traits such as atherosclerosis and dysbiosis-related diseases, and how to define new targets for improving disease progression in high-risk CAD patients with utility of QTL and system genetics.

5.4. Method

5.4.1. Ethics Statement

We followed all NIH animal welfare guidelines. The animal care and study protocols were approved by IACUC at the University of California Davis.

5.4.2. Animals: Hyperlipidemic DO-F1 Mice

We utilized hyperlipemic CETP/ApoE3 Leiden males, hemizygous to the CETP and ApoE3 transgenes (Tg), which were kindly provided by Dr. Lusis (Bennett et al., 2015) to induce atherosclerosis susceptibility in two mouse populations. A total of 200 DO females (JAX stock number 009376, outbreeding generation # 26, 28) were obtained from the Jackson Laboratory (Bar Harbor, ME) and crossed with CETP/ApoE3 Leiden males. This cross resulted in 238 (CETP/ApoE3 Leiden × J:DO) F1 females and 234 (CETP/ApoE3 Leiden × J:DO) F1 males which were used in this study. All progeny were genotyped to confirm the presence of CETP and ApoE3-Leiden transgenes. DO-F1 mice were maintained in the Mouse Biology Protocol vivarium at the University of California Davis. Mice were fed a synthetic diet, AIN-76A (D10001, Research Diets, New Brunswick, NJ) up to 8 weeks of age. At 8 weeks of age, mice were fed with a synthetic HFHC (33 kcal % fat from cocoa butter,1.25% cholesterol) diet (Research Diets D121083) (see **Table S1**) *ad libitum*. Mice were euthanized for tissue collection after being fed this diet for 16 weeks. Animals were maintained on a 12 h light and dark cycle

under temperature- and humidity-controlled conditions. Euthanasia of all mice was performed by cervical dislocation after anesthesia with isoflurane, and aorta and liver were collected.

5.4.3. Measurement of triglycerides and total cholesterol in the liver

TG and TC in the liver were quantified via Folch extraction. Mouse liver was collected, frozen, and stored at −80°C prior to analysis. Cut 30 ± 8 mg of frozen mouse liver tissue was thoroughly homogenized for 5 min in 500 µl of a 2:1 v/v chloroform/methanol mix and then equilibrated for 15 min at room temperature. After adding 100 µL of 0.9% w/v NaCl to each sample, the samples were vortexed for 1 min and centrifuged at 2000 × g for 15 min at 4°C. The lower organic phase was separated and evaporated in Eppendorf tubes overnight in a fume hood. After evaporation, each tube was resuspended with 500 µL of a 0.5% Triton X-100/PBS solution, sonicated with 20% amplitude for 15 seconds using BRANSON Digital Sonifier, and placed in a drying bath at 55°C for 5 min. Hepatic TG/TC were measured using a colorimetric assay (Infinity™, Thermo Scientific), according to the manufacturer's instructions as follows: 2 µL of the standards, samples, and blanks were pipetted into a 96-well plate in duplicate and 200 µL of the Infinity reagent was added to the 96-well plate. Absorbance (500/660 nm) was measured on a 96-well plate reader.

5.4.4. Fecal DNA Extraction

The DO-F1 mice at 24 weeks of age were singly housed for 24 h prior to fecal collection. During the last 4 hours of this period, feces were collected from each cage with food removed and used for gut microbiota analysis. DNA was isolated from feces using a ZymoBIOMICS 96 MagBead DNA Kit (Zymo Research, Irvine, CA, USA) with an epMOTION (Eppendorf, Hamburg, Germany) automatic robotic system. In brief, mouse feces (about 100 mg) were resuspended in 750 µl of lysis solution in silica beads containing tubes. Samples were mechanically

homogenized at 6 m/s for 2 min using a FastPrep Automated Homogenizer (MP Biomedicals, Solon, OH; maximum setting for 1 min at room temperature). The suspension was centrifuged for 1 min at full speed. 200 µl of supernatant was transferred to the 96 deep-well plates (Fisher Scientific, Waltham, MA). The isolated DNA was eluted in nuclease-free water. DNA concentration and quality were verified using a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE) and stored at -20°C until further use.

5.4.5. 16S rRNA amplicon library preparation sequencing

To amplify variable region 4 of 16S rRNA gene, PCR was performed using universal primer set 515F and barcoded 806R in triplicate as described earlier (Nazmul Huda et al., 2020). Each reaction consisted of 25 µl GoTaq Green Master Mix® (Madison, WI), 10 ng genomic DNA, 10 µM each primer, and water were included in the final reaction volume of 25 µl. PCR was performed under the following conditions: initial denaturation at 95°C for 3 min, followed by 25 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s and elongation at 72°C for 30 s, and a final elongation step at 72°C for 5 min. PCR products were confirmed using denaturation gradient gel electrophoresis. PCR products were gel-purified using a Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI) according to the manufacturer's protocol. Samples were sequenced with the MiSeq (2×250 bp paired-end v2 kit) (Illumina, San Diego, CA, USA) by the UC Davis DNA Technologies Core at the Genome Center using custom sequencing primers.

5.4.6. 16S Analysis

The raw reads were processed, filtered, and analyzed with QIIME2-DADA2 pipeline (Bolyen et al., 2019; Callahan et al., 2016) to determine the ASV. This produced a total of 7,355,669 sequences with 15,852 ± 6,221 sequences for DO-F1 fecal samples. The ASVs were aligned with

the mafft aligner (Katoh and Standley, 2013) using the q2-alignment plug-in. The q2-phylogeny plugin was used to reconstruct phylogeny via FastTree (Price et al., 2010). Taxonomic classification was assigned using classify-sklearn (Bokulich et al., 2018) against the SILVA database release 138 (Yilmaz et al., 2014). Alpha- (Shannon diversity index, Faith's phylogenetic diversity, Observed ASVs, and Plelou's evenness) and Beta-diversity (phylogeny-based [Weighted UniFrac and Unweighted UniFrac] and abundance-based [Bray-Curtis]) (Lozupone and Knight, 2005) analyses were performed at a rarefaction depth of 6,000 sequences per sample. Six samples in DO-F1 mice were removed from subsequent analysis because of the lower sequencing depth.

Subsequent processing and analysis were performed using Phyloseq to import data generated from QIIME2 into R (v.3.5.3) (McMurdie and Holmes, 2013). Hierarchical clustering (Euclidean distance) was performed using relative abundance data. Box plots were generated for each classification level indicating the abundance of taxa within the classification level for the samples. Differential abundance of genera and ASV analyses were performed by using ANCOM 2.1 (Kaul et al., 2017) upon adjusting for the covariates such as sex in DO-F1 mice and generation number in DO females (F0 outbreeding generation # 26, 28). Significance was defined as ANCOM W>0.7.

Taxonomic distributions were produced by collapsing ASV counts into higher taxonomic levels (genus to phylum) by phylogenetic rank. We excluded unclassified, uncultured, and unidentified genus levels from the analysis. A total of 242 ASVs and 69 collapsed microbial taxonomies were identified in fecal samples in DO-F1 mice. We defined a core measurable microbiota (CMM) as a taxon found in over 20% of the mice for subsequent analysis (Benson et

al., 2010). This was done to eliminate the effects of excessive variation in the data due to the low abundance and/or sparsely distributed bacterial taxa.

For functional inferences on the identified ASVs in DO-F1 mice, we used the the PICRUSt2 algorithm (v2.2.0-b) (Douglas et al., 2020). Phylogenetic placement in PICRUSt2 is based on the results of three analysis tools; 1) HMMER (http://www.hmmer.org) to place ASVs, 2) EPA-ng (Barbera et al., 2019) to determine the optimal position of these placed ASVs in a reference phylogeny, and 3) GAPPA (Czech and Stamatakis, 2019) to create a new tree incorporating the ASV placements. This results in a phylogenetic tree containing both the reference genome and environmentally sampled organisms used to predict the number of individual gene family copy numbers for each ASV. Metabolic Pathway Database (MetaCyc) pathway abundance, the main high-level predictions output, was calculated in PICRUSt2 through structured mapping of gene families for Enzyme Commission (EC) gene families to pathways (Caspi et al., 2016), and ANCOM was used to identify differentially abundant MetaCyc pathways or EC gene families with many zero abundances after adjusting sex and generation number in DO-F1 mice. For microbial quantitative trait loci (mbQTL) mapping, the CMM, MetaCyc pathways, and EC gene families were transformed to rank normal scores using the 'rankZ' function in the DOQTL R package (Gatti et al., 2014).

5.4.7. Statistical testing of gut microbiota diversity and cardiometabolic traits

We assessed whether variations of the three atherosclerotic traits (aortic lesion area, plasma total cholesterol, and hepatic total cholesterol) could be explained by gut microbiota alpha- and beta-diversity using a linear mixed-effect regression model. First, we confirmed normal distribution of traits by using the Shapiro-Wilk test in R. Non-normal traits were log-transformed and aortic lesion area was transformed to rank normal scores using the 'rankZ' function in the DOQTL R

package. Second, we assessed whether variations in traits could be explained by gut microbial diversity in the linear mixed-effects regression model with adjusted sex, generation number, and genetic relatedness (kinship) using the relmatLmer function from the R package lme4qtl (Ziyatdinov et al., 2018).

5.4.8. Quantitative trait loci mapping for microbial abundance

QTL mapping was performed using the R package R/QTL2 (v0.20). Marker genotypes from a GeneSeek final report were filtered for GigaMUGA-containing single nucleotide polymorphism (SNPs) and encoded into a hetero/homozygous set (e.g. AA, AB) (https://kbroman.org/qtl2/). These genotype codes were processed with GigaMUGA genotype codes of the DO founder strains using a Hidden Markov model and haplotypes were defined as previously reported (Broman, 2012a, b). We estimated the posterior probability to have one of the eight possible genotype states at each SNP. We performed association mapping by fitting a linear mixed model at each SNP, where we regressed the traits on the diplotype probabilities. A kinship matrix was incorporated as a random effects term to adjust for the genetic relatedness between mice. Genotype probabilities were reduced to eight founder allele probabilities and were used to generate a kinship matrix using the "leave-one-chromosome-out" method to reduce bias from same chromosome SNPs (Yang et al., 2014).

Genome scans were performed using three different models using the scan1 function in R/qtl2 as previously reported: sex additive, female mice, and male mice. Reported mapping statistics were logarithms of odds ratios (LODs) which describe the log-scaled likelihood difference in full and null genome scan models. Confidence intervals for QTL were calculated by a 95% Bayesian credible interval (Sen and Churchill, 2001). Using a similar mixed-effect model with allele probabilities as random effects, the contribution of each founder strain genotype at

each QTL was determined using Best Linear Unbiased Predictor (BLUP). Candidate genes found in QTL were identified by position based on the Wellcome Trust Sanger mouse genomes database (www.sanger.ac.uk), release 1303, based on genome assembly GRCm38 (Yalcin et al., 2011).

Permutation analysis was used for subsequent filtering to reduce the initial output to a highly reliable QTLs set. The significance threshold at $P < 0.05$ and the suggestive threshold at $P < 0.1$ of all reported QTL was empirically determined by permutation analysis, where rows of the genotype data were randomized for each trait and a maximum LOD score was generated (Doerge and Churchill, 1996). We described QTL analysis for cardiometabolic traits and RNA-seq methods from DO-F1 mice in Chapter 4. In this study, aortic lesion area QTL, mbQTL, and eQTL using generation as an additive covariate were filtered out at the 95th quantile of the 1,000 times null distributions. For QTL mapping, aortic lesion area was transformed to rank normal scores using the 'rankZ' function in the DOQTL R package. eQTL was defined as *cis*-eQTL when the SNP with the maximum LOD score was within ±4 Mb at the transcription start site.

5.4.9. Heritability

To determine the extent to which phenotypic variation is affected by genotypic variation, a linear mixed-effect model was used to estimate the narrow-sense heritability scores of the gut microbial taxa. This was performed using the function est_herit in R/qtl2 by submitting a kinship matrix and each trait value.

5.4.10. Weighted Gene Coexpression Network Analysis (WGCNA)

To best assess the correlation between hepatic gene network, cardiometabolic traits, and gut microbiome, the liver transcriptome was analyzed individually for sex. Coexpression gene

modules were calculated using WGCNA version 1.13 (Zhang and Horvath, 2005), which performs network construction by module detection. For the WGCNA analysis, log2 transformed 13,094 transcripts greater than 1 TPM measured in 164 DO-F1 liver samples were included and seven Y chromosome genes (*Ddx3y*, *Eif2s3y*, *Gm29650*, *Gm37158*, *Gm47283*, *Kdm5d*, and *Uty*) were removed in female-specific analysis due to a higher number of missing values detected using the "goodSampleGenes" test implemented in WGCNA. We used a soft thresholding power of 12 by the scale-free topology criterion in the WGCNA package using the "pickSoftThreshold" function. For network parameters for both female and male networks, we chose the "unsigned" network type to maintain the relationship of the negatively correlated gene and the "signed" topological overlap matrix (TOM) to exclude the connections influenced by the noise (Langfelder and Horvath, 2008; Zhang and Horvath, 2005). We obtained 9 modules in females and 7 modules in males through dynamic branch cutting using 0.25 as the merge threshold. The network connectivity of each gene was calculated as the sum of the intensity of connectivity with all genes in the other network. Spearman's correlations were performed between module eigengenes (MEs) and cardiometabolic traits or gut microbiome (microbial diversity, microbial genus levels, MetaCyc pathways, and EC gene families). The average number of transcripts per module in females was 1,120, ranging from 45 (darkgrey module in females) to 6,457 (darkred module in females) and the average number of transcripts per module in males was 1,447, ranging from 130 (black module in males) to 4,351 (turquoise module in males).

5.4.11. Enrichment analysis

Enrichment analyses for sex-specific DEGs, modules and genes that were significantly correlated with aortic lesion area in each sex were performed using enrichR (Chen et al., 2013) to generate enrichment categories from the GO Biological Process 2018, KEGG 2019 Mouse, and Jenson

Diseases (Pletscher-Frankild et al., 2015). This analysis identifies differential enrichment in the functional categories of transcripts. The GO Biological process 2018 contains 5,103 terms and 14,433 genes. While it is clear that individual GO terms can be found in related classes of ontology, GO terms do not occupy strictly fixed levels in a hierarchy. Each GO term identified is associated with a unique GO annotation number that relates to a specific function. Both the Gene Ontology website (http://geneontology.org/docs/faq/) and enrichR tool do not utilize a specific hierarchy thus all available terms are used in the analysis.

5.4.12. Other statistical analysis

All statistical analyses were performed in R (v.3.5.3) (R Core Team). Sex differences of alpha-microbial diversity were assessed using Wilcoxon signed-rank test in DO-F1 mice. Spearman's correlation was used to correlate the cardiometabolic traits, liver transcripts, and abundance of microbial taxa. The p-values were adjusted for multiple comparisons using the Benjamini-Hochberg (BH) procedure (Benjamini and Hochberg, 1995), and correlation coefficients and adjusted p-value were visualized using the 'pheatmap' package (Kolde et al., 2018). Significance was determined with a $P<0.05$. Summary statistics were calculated to evaluate the magnitude of variability of the cardiometabolic traits and microbial abundance by sex. PERMDISP2 function of Vegan R Package (Oksanen et al., 2018)was used to determine the homogeneity difference in the dispersion of microbial compositions between groups with 999 permutations. A nonparametric-based PERMANOVA statistical test (McArdle and Anderson, 2001) was used to compare the composition of microbiota between sexes using the Vegan R package (Oksanen et al., 2018).

5.5. Results

5.5.1. Effects of genetic variation and sex on microbial diversity, taxonomic ranks, and functional profiling

In Chapter 4, we identified a wide range of cardiometabolic responses and substantial sex differences of the traits in DO-F1 mice. In this study, we characterized the gut microbiota by sequencing the 16S rRNA gene from DNA extracted from fecal samples to investigate the effects of genetic variation and sex on gut microbial diversity and compositions in DO-F1 mice (n=464). We defined a core measurable microbiota (CMM) as a taxon found in over 20% of the mice for subsequent analysis (Benson et al., 2010). This was done to eliminate the effects of excessive variation in the data due to the low abundance and/or sparsely distributed bacterial taxa. For instance, females showed higher alpha-diversity ($P \leq 0.001$ and range: 1.6-5.7 for Shannon index; $P \leq 0.05$ and range: 16-142 for Observed ASV; $P \leq 0.001$ and range: 0.3-0.9 for Pielou's evenness) than males (**Figures 5.1A**). Principal Coordinates Analysis (PCoA) of unweighted UniFrac distances (phylogeny-based) and Bray-Curtis (abundance-based) showed also significant differences between sexes in the gut microbial community ($P \leq 0.001$), but the extent to which the variance of beta-diversity is explained by sex was less than 3% (**Figures 5.1B and Table 5.1**). Weighted UniFrac distances matrix (phylogeny and abundance-based) showed no significant differences between sexes ($P \leq 0.831$; **Figure 5.1B and Table 5.1**).

We also identified highly inter-individual variability in microbial taxa and observed a significant differential abundance of microbial taxa at phylum, class, order, family levels between sexes in DO-F1 mice. The abundance of phyla such as Actinobacteria, Bacteroidetes, and Firmicutes ranged from 0-19.27%, 0.06–40.57%, and 37.36-98.13% (**Table 5.2**). In addition, analysis of composition of microbiotas (ANCOM) was used to identify differential taxa

abundance between sexes at each taxonomic level. Taxa with ANCOM W statistics >0.7 were considered differentially abundant. For instance, significant sex differences were found only in Bacilli, Clostridia, and Erysipelotrichia at the class level (**Table 5.2**). A total of 11 genera and 8 ASVs (**Figures 5.2A and 5.2B**) were found to have significant differential abundance by sex.

Lastly, we evaluated differences in functional profiling of microbial communities using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States 2 (PICRUSt2). The number of predicted Metabolic Pathway Database (MetaCyc) pathways and Enzyme Commission (EC) gene families inferred by PICRUSt2 for ASVs was 297 and 1367, respectively (data not shown). Among the identified pathways, we found the 6 MetaCyc pathways (**Figure 5.2C and Table 5.3**) and 241 EC gene families (**Table 5.4**) that showed statistically significant differences between sexes by ANCOM with W cut off 0.9. Consistent with the studies that the gut microbiota is affected by sex, sex differences were found in the gut microbial diversity, specific taxa, and functional profiling in hyperlipidemic DO-F1 mice.

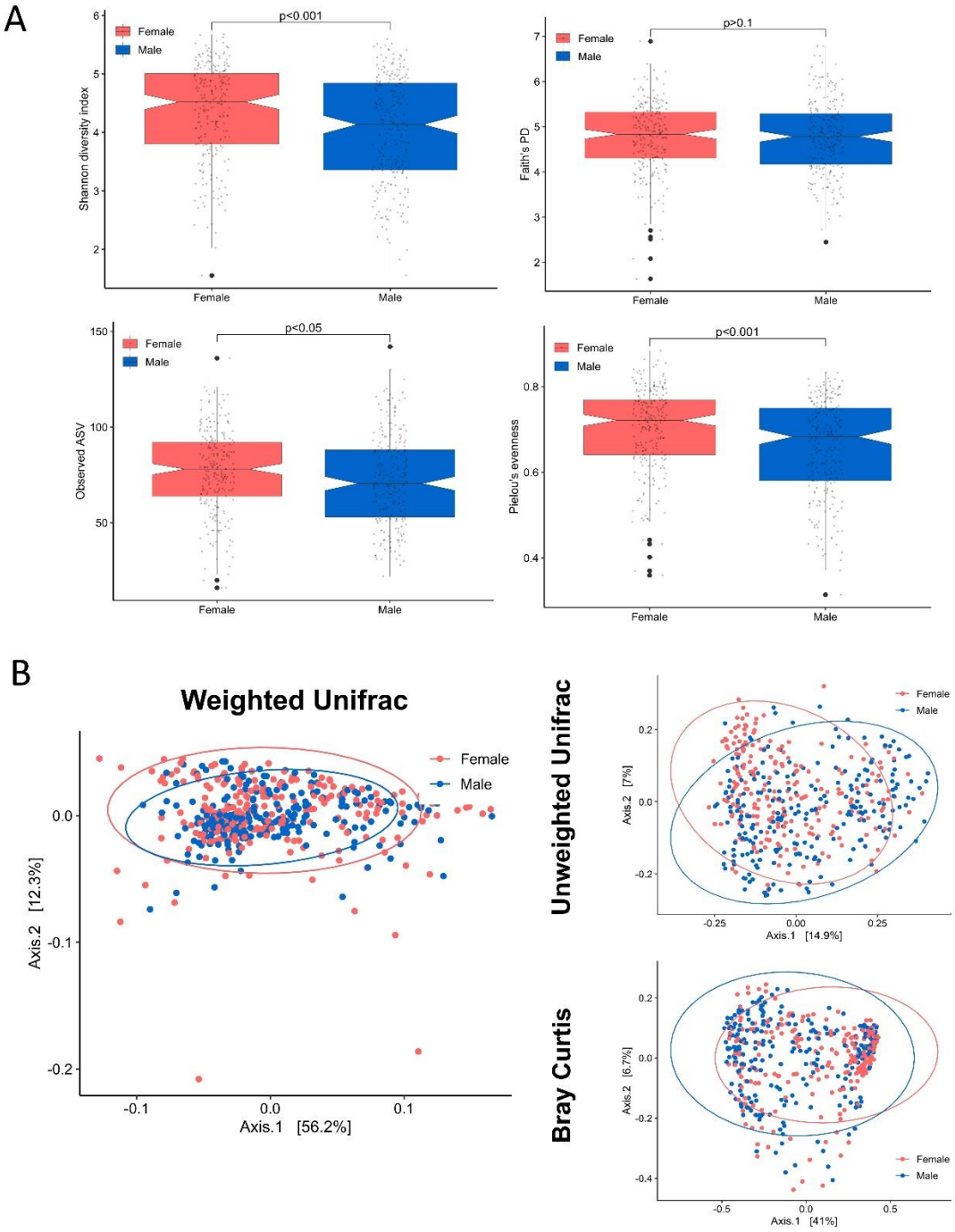


Figure 5.1. Sex differences in microbial diversity in DO-F1 mice.
(A) Shannon diversity, observed amplicon sequence variants (ASV), Faith's phylogenetic diversity and Plelou's evenness indices by sex at 24 weeks of age. The p-values were unpaired Wilcoxon signed-rank test between groups for alpha-diversity. (B) Weighted UniFrac (PERANOVA = 0.831), Unweighted UniFrac (PERANOVA: <0.001), and Bray-Curtis (PERANOVA <0.001) beta diversity principal coordinate plot by sex.

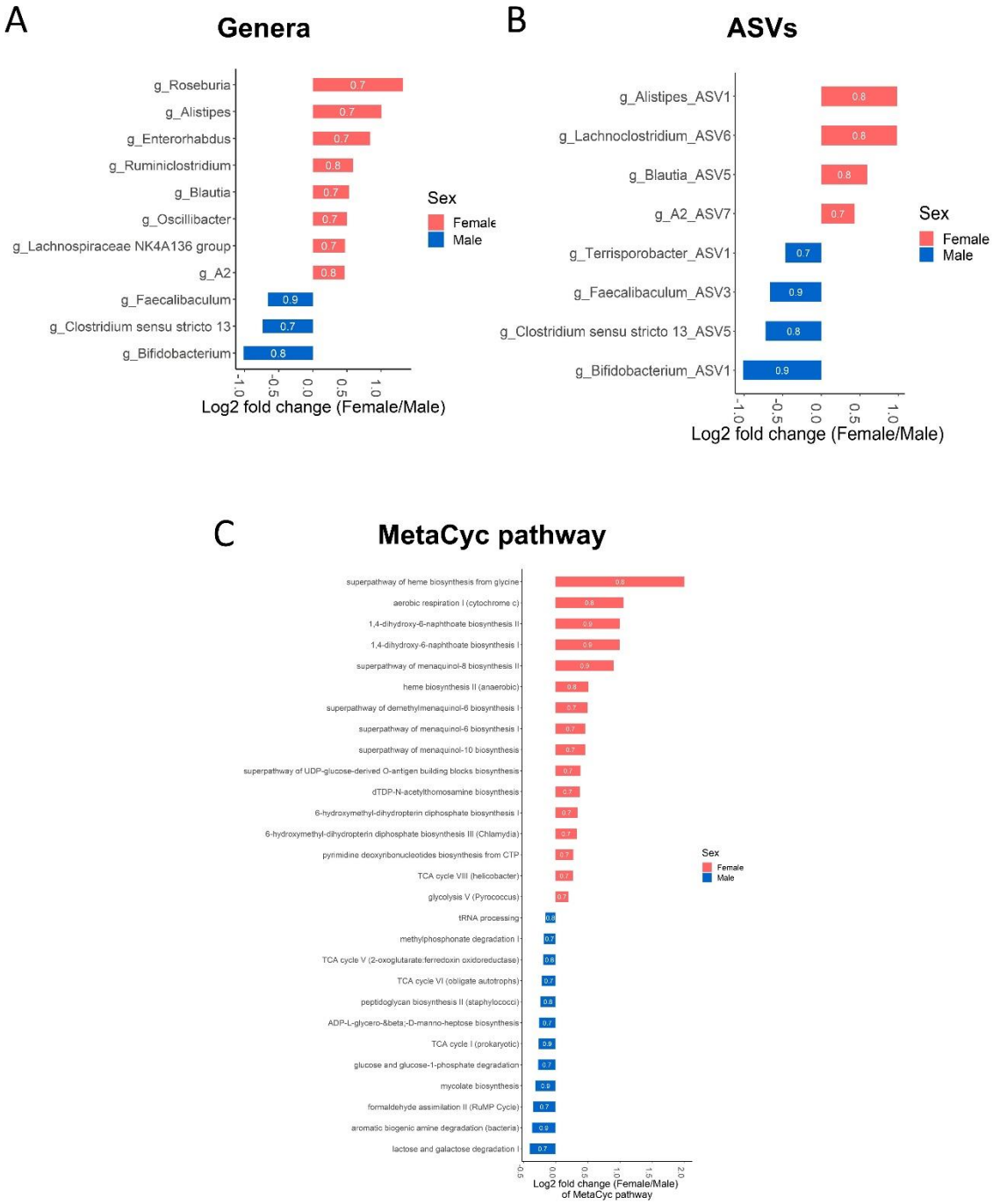


Figure 5.2. Sex differences in specific taxa and MetaCyc pathways in DO-F1 mice.
(A-C) Analysis of composition of microbiomes (ANCOM) analysis of microbial genera, ASVs, MetaCyc pathways by sex at 24 weeks after adjusting generation number. All identified taxa and pathways by ANCOM met W>0.7 significant level. Differentially abundant genera, ASVs, and pathways were ordered from top to bottom by log2 fold change of mean between female (red color) and male (blue color). All W-statistic cutoffs from ANCOM output (0.7, 0.8, and 0.9) were denoted for each taxon.

Table 5.1. Multivariate homogeneity of groups dispersions (betadisper) and Permutational Multivariate Analysis of Variance (ADONIS) analyses of the microbial beta-diversity between sexes in DO-F1 mice.

	β-dispersion		ADONIS		
	F	**P**	**F.Model**	**R^2**	**P**
24 weeks					
Weighted UniFrac	9.3493	0.002	0.39079	0.00085	0.831
Unweighted UniFrac	0.8566	0.358	6.8719	0.01466	<0.001
Bray-Curis	11.646	<0.001	12.623	0.0266	<0.001

Table 5.2. Sex differences and narrow sense heritability in taxon (phylum, class, order, family, and genus) and microbial amplicon sequence variants (ASVs) except for uncultured, and unidentified taxa detected in at least 20% of DO-F1 mice at 24 weeks. n = 464 (235 females and 229 males).

Taxonomy	Taxa	% of presence in mice	W cutoff (Females vs males) (ANCOM)[a]	Male mice	Female mice	Narrow sense heritability		
				Mean ± SD	Mean ± SD	Entire mice	Female mice	Male mice
class	c_Actinobacteria	47.63	0.7	49.63 ± 79.41	24.09 ± 53.03	0.52	0.76	0.63
class	c_Bacilli	100.00	0.7	563.3 ± 458.78	781.81 ± 628.66	0.43	0.80	0.49
class	c_Clostridia	100.00	0.7	1991.03 ± 1187.62	2519.23 ± 1169.09	0.42	0.51	0.66
class	c_Erysipelotrichia	99.35	0.7	2213.21 ± 1481.37	1482.09 ± 1411.17	0.41	0.68	0.67
order	o_Bifidobacteriales	47.63	0.7	49.63 ± 79.41	24.09 ± 53.03	0.52	0.76	0.63
order	o_Lactobacillales	100.00	0.7	561.52 ± 457.7	779.15 ± 623.47	0.43	0.79	0.50
order	o_Clostridiales	100.00	0.7	1991.03 ± 1187.62	2519.23 ± 1169.09	0.42	0.51	0.66
order	o_Erysipelotrichales	99.35	0.7	2213.21 ± 1481.37	1482.09 ± 1411.17	0.41	0.68	0.67
familiy	f_Bifidobacteriaceae	49.57	0.8	40.44 ± 75.76	44.18 ± 76.84	0.16	0.10	0.09
familiy	f_Eggerthellaceae	91.81	0.7	33.95 ± 30.91	30.9 ± 31.15	0.03	0.20	0.00
familiy	f_Rikenellaceae	63.15	0.8	46.93 ± 95.23	38.16 ± 86.55	0.28	0.50	0.22
familiy	f_Lachnospiraceae	100.00	0.7	1677.2 ± 1115.33	1420.66 ± 1092.18	0.15	0.16	0.17
familiy	f_Erysipelotrichaceae	99.14	0.7	1781.41 ± 1518.52	2016.67 ± 1504.68	0.15	0.10	0.07
genus	g_Bifidobacterium	47.63	0.8	49.63 ± 79.41	24.09 ± 53.03	0.52	0.76	0.63
genus	g_Enterorhabdus	92.03	0.7	23.95 ± 22.8	43.69 ± 36.52	0.39	0.80	0.35
genus	g_Alistipes	64.44	0.7	28.06 ± 62.1	57.36 ± 104.42	0.42	0.68	0.73
genus	g_Clostridium sensu stricto 13	36.64	0.7	68.09 ± 121.02	40.52 ± 92.08	0.41	0.77	0.68
genus	g_A2	83.41	0.8	306.37 ± 329.02	422.76 ± 331.99	0.34	0.62	0.71
genus	g_Blautia	85.78	0.7	254.92 ± 319.19	368.16 ± 348.77	0.47	0.67	0.59
genus	g_Lachnospiraceae NK4A136 group	88.15	0.7	88.13 ± 112.15	122.59 ± 129.42	0.39	0.40	0.66
genus	g_Roseburia	45.91	0.7	28.77 ± 70.89	73.43 ± 140.98	0.53	0.79	0.55
genus	g_Oscillibacter	89.01	0.7	61.14 ± 64.57	86.75 ± 76.75	0.48	0.75	0.60
genus	g_Ruminiclostridium	79.96	0.8	53.24 ± 70.15	80.76 ± 80.2	0.52	0.66	0.73
genus	**g_Faecalibaculum**	93.10	**0.9**	1764.79 ± 1368.18	1116.09 ± 1229.38	0.44	0.64	0.72
ASV	g_A2_ASV7	78.45	0.7	265.79 ± 289.26	358.72 ± 292.59	0.33	0.68	0.72
ASV	g_Alistipes_ASV1	62.28	0.8	25.19 ± 59.29	50.69 ± 97.42	0.43	0.65	0.70
ASV	**g_Bifidobacterium_ASV1**	47.63	**0.9**	49.63 ± 79.41	24.09 ± 53.03	0.52	0.76	0.63
ASV	g_Blautia_ASV5	70.91	0.8	123.9 ± 195.71	187.57 ± 241.64	0.39	0.42	0.55
ASV	g_Clostridium sensu stricto 13_ASV5	31.90	0.8	53.19 ± 99.92	31.88 ± 78.07	0.37	0.81	0.58
ASV	**g_Faecalibaculum_ASV3**	**93.10**	**0.9**	1710.66 ± 1327.52	1079.97 ± 1190.21	0.43	0.64	0.69
ASV	g_Lachnoclostridium_ASV6	53.88	0.8	13.65 ± 28.38	20.03 ± 29.9	0.46	0.57	1.00
ASV	g_Terrisporobacter_ASV1	33.19	0.7	59.28 ± 110.33	42.68 ± 108.69	0.45	1.00	0.70

[a]ANCOM2.1 was performed after adjusting the generation number in the model. All W-statistic cutoffs from ANCOM output (0.7, 0.8, and 0.9) were denoted for each taxon.

Table 5.3. Sex differences in MetaCyc pathways detected in at least 20% of DO-F1 mice at 24 weeks. n = 464 (232 females and 232 males).

Pathway	Description	% of presence in mice	W cutoff (ANCOM)[a]	Male mice	Female mice	Narrow sense heritability		
				Mean ± SD	Mean ± SD	Entire mice	Female mice	Male mice
PWY_6263	**superpathway of menaquinol-8 biosynthesis II**	67.24	**0.9**	66.5 ± 124.08	125.22 ± 185.48	0.54	0.61	0.79
PWY_7371	**1,4-dihydroxy-6-naphthoate biosynthesis II**	65.73	**0.9**	28.99 ± 62.51	58.87 ± 104.62	0.45	0.60	0.59
PWY_7374	**1,4-dihydroxy-6-naphthoate biosynthesis I**	64.87	**0.9**	28.23 ± 61.97	57.25 ± 103.62	0.46	0.76	0.50
PWY_7431	**aromatic biogenic amine degradation (bacteria)**	76.29	**0.9**	97.13 ± 203.45	75.17 ± 162.74	0.40	0.52	0.73
PWYG_321	**mycolate biosynthesis**	50.22	**0.9**	37.49 ± 53.61	29.95 ± 58.48	0.40	0.52	0.73
TCA	**TCA cycle I (prokaryotic)**	96.55	**0.9**	468.62 ± 388.92	391.15 ± 452.8	0.45	0.58	0.51
HEMESYN2_PWY	heme biosynthesis II (anaerobic)	100.00	0.8	237.36 ± 189.05	338.23 ± 218.65	0.39	0.61	0.70
PWY_3781	aerobic respiration I (cytochrome c)	31.47	0.8	39.61 ± 245.99	83.31 ± 293.52	0.27	0.61	0.48
PWY_5265	peptidoglycan biosynthesis II (staphylococci)	94.83	0.8	693.25 ± 496.75	588.85 ± 509.67	0.54	0.61	0.80
PWY_5920	superpathway of heme biosynthesis from glycine	33.19	0.8	1.6 ± 5.52	9.42 ± 44.59	0.49	0.96	0.75
PWY_6969	TCA cycle V (2-oxoglutarate:ferredoxin oxidoreductase)	96.55	0.8	638.86 ± 424.56	560.18 ± 496.97	0.27	0.10	0.98
PWY0_1479	tRNA processing	72.63	0.8	67.89 ± 81.2	60.83 ± 101.7	0.36	0.52	0.66
GLUCOSE1PMETAB _PWY	glucose and glucose-1-phosphate degradation	77.80	0.7	99.72 ± 168.51	82.51 ± 199.3	0.35	0.48	0.47
LACTOSECAT_PWY	lactose and galactose degradation I	96.55	0.7	241.68 ± 277.81	182.66 ± 249.84	0.20	0.34	0.45
P341_PWY	glycolysis V (Pyrococcus)	84.27	0.7	137.52 ± 168.16	158.18 ± 138.36	0.50	0.66	0.85
PWY_1861	formaldehyde assimilation II (RuMP Cycle)	100.00	0.7	1364.8 ± 707.77	1073.2 ± 676.62	0.41	0.48	0.73
PWY_5850	superpathway of menaquinol-6 biosynthesis I	99.57	0.7	152.42 ± 76.68	210.44 ± 120.59	0.40	0.48	0.73
PWY_5860	superpathway of demethylmenaquinol-6 biosynthesis I	99.57	0.7	110.31 ± 57.33	156.15 ± 102.93	0.36	0.55	0.37
PWY_5896	superpathway of menaquinol-10 biosynthesis	99.57	0.7	152.42 ± 76.68	210.44 ± 120.59	0.41	0.61	0.89
PWY_5913	TCA cycle VI (obligate autotrophs)	98.92	0.7	538.36 ± 463.09	464.45 ± 493.37	0.33	0.77	0.37
PWY_6147	6-hydroxymethyl-dihydropterin diphosphate biosynthesis I	100.00	0.7	1638.58 ± 909.17	2081.43 ± 858.48	0.30	0.57	0.48
PWY_7210	pyrimidine deoxyribonucleotides biosynthesis from CTP	77.59	0.7	59.05 ± 106.45	71.77 ± 198.79	0.60	0.99	0.64
PWY_7315	dTDP-N-acetylthomosamine biosynthesis	100.00	0.7	858.98 ± 651.15	1114.81 ± 678.27	0.44	0.64	0.89
PWY_7328	superpathway of UDP-glucose-derived O-antigen building blocks biosynthesis	100.00	0.7	603.24 ± 439.63	788.34 ± 439.59	0.43	0.61	0.86
PWY_7539	6-hydroxymethyl-dihydropterin diphosphate biosynthesis III (Chlamydia)	100.00	0.7	1371.27 ± 691.36	1723.39 ± 638.66	0.46	0.57	0.77
PWY0_1241	ADP-L-glycero-β-D-manno-heptose biosynthesis	77.80	0.7	69.69 ± 116.75	58.15 ± 136.03	0.51	0.56	0.63
PWY0_1533	methylphosphonate degradation I	84.70	0.7	37.38 ± 80.02	32.76 ± 69.02	0.42	0.69	0.73
REDCITCYC	TCA cycle VIII (helicobacter)	100.00	0.7	524.7 ± 473.8	632.61 ± 447.64	0.39	0.59	0.84

[a]ANCOM2.1 was performed after adjusting the generation number in the model. All W-statistic cutoffs from ANCOM output (0.7, 0.8, and 0.9) were denoted for each taxon.

Table 5.4. Sex differences and narrow sense heritability in EC gene families detected in at least 20% of DO-F1 mice.

EC gene families	Description	% of presence in mice	W cutoff 0.9 (ANCOM)[a]	Male mice	Female mice	Narrow sense heritability		
				Mean ± SD	Mean ± SD	Entire mice	Female mice	Male mice
EC.1.2.99.7	Aldehyde dehydrogenase (FAD-independent)	93.62	0.9	30.58 ± 56.45	22.27 ± 58.64	0.34	0.70	0.47
EC.1.21.98.1	Cyclic dehypoxanthinyl futalosine synthase	96.72	0.9	29.22 ± 62.66	59.2 ± 104.65	0.41	0.58	0.75
EC.1.3.3.4	Protoporphyrinogen oxidase	95.00	0.9	13.79 ± 123.57	26.2 ± 104.34	0.43	0.37	0.61
EC.1.4.3.4	Monoamine oxidase	97.63	0.9	76.16 ± 164.71	58.56 ± 129.38	0.46	0.54	0.71
EC.1.6.1.2	NAD(P)(+) transhydrogenase (Re/Si-specific)	97.91	0.9	172.42 ± 245.46	100.09 ± 179.19	0.34	0.65	0.80
EC.2.1.1.219	tRNA (adenine(57)-N(1)/adenine(58)-N(1))-methyltransferase	94.85	0.9	50.12 ± 79.35	25.21 ± 53.85	0.61	1.00	0.61
EC.2.1.1.220	tRNA (adenine(58)-N(1))-methyltransferase	94.85	0.9	50.12 ± 79.35	25.21 ± 53.85	0.18	0.36	0.44
EC.2.1.4.1	Glycine amidinotransferase	99.25	0.9	24.02 ± 22.57	43.87 ± 36.55	0.47	0.63	0.73
EC.2.4.1.11	Glycogen(starch) synthase	96.44	0.9	28.06 ± 62.1	57.36 ± 104.42	0.18	0.34	0.33
EC.2.4.1.57	Phosphatidylinositol alpha-mannosyltransferase	94.76	0.9	49.63 ± 79.41	24.09 ± 53.03	0.20	0.34	0.30
EC.2.4.99.16	Starch synthase (maltosyl-transferring)	94.76	0.9	49.63 ± 79.41	24.09 ± 53.03	0.10	0.49	0.02
EC.2.5.1.120	Aminodeoxyfutalosine synthase	96.72	0.9	29.22 ± 62.66	59.2 ± 104.65	0.42	1.00	0.60
EC.2.5.1.26	Alkylglycerone-phosphate synthase	94.68	0.9	101.22 ± 169.8	63.15 ± 143.57	0.40	0.79	0.52
EC.2.5.1.56	N-acetylneuraminate synthase	97.20	0.9	15.45 ± 26.73	30.55 ± 40.44	0.54	0.61	0.77
EC.2.5.1.76	Cysteate synthase	96.44	0.9	28.06 ± 62.1	57.36 ± 104.42	0.27	0.89	0.56
EC.2.6.1.57	Aromatic-amino-acid transaminase	99.83	0.9	546.44 ± 399.51	378.69 ± 371.36	0.51	0.57	0.64
EC.2.7.1.63	Polyphosphate--glucose phosphotransferase	94.76	0.9	49.63 ± 79.41	24.09 ± 53.03	0.51	0.57	0.64
EC.2.7.7.42	[Glutamate--ammonia-ligase] adenylyltransferase	97.91	0.9	61.4 ± 83.51	38 ± 66.65	0.11	0.54	0.02
EC.2.7.7.43	N-acylneuraminate cytidylyltransferase	99.68	0.9	445.06 ± 483.34	632.04 ± 522.39	0.42	0.51	0.76
EC.2.7.7.49	RNA-directed DNA polymerase	98.32	0.9	89.6 ± 159.11	144.77 ± 219.85	0.36	0.57	0.82
EC.2.7.7.59	[Protein-PII] uridylyltransferase	97.95	0.9	61.89 ± 83.39	39.13 ± 67.11	0.47	0.96	0.78
EC.2.7.8.6	Undecaprenyl-phosphate galactose phosphotransferase	99.81	0.9	737.22 ± 436.88	532.68 ± 451.51	0.37	0.63	0.72
EC.3.1.26.12	Ribonuclease E	97.97	0.9	71.67 ± 94.85	48.29 ± 80.5	0.40	0.67	0.79
EC.3.1.3.71	2-phosphosulfolactate phosphatase	98.02	0.9	116.23 ± 187.57	84.07 ± 189.24	0.51	0.57	0.64
EC.3.2.1.85	6-phospho-beta-galactosidase	99.78	0.9	519.31 ± 397.96	361 ± 355.28	0.54	0.60	0.77
EC.3.2.2.26	Futalosine hydrolase	96.49	0.9	27.78 ± 61.51	56.77 ± 103.62	0.08	0.30	0.26
EC.3.4.11.19	D-stereospecific aminopeptidase	96.59	0.9	100.03 ± 151.87	65.72 ± 136.05	0.47	0.65	0.86
EC.3.4.19.1	Acylaminoacyl-peptidase	99.46	0.9	488.79 ± 373.78	313.69 ± 339.01	0.03	0.00	0.00
EC.3.4.21.26	Prolyl oligopeptidase	96.49	0.9	29.41 ± 65.14	64.22 ± 131.64	0.51	0.57	0.64
EC.3.4.21.83	Oligopeptidase B	97.91	0.9	61.4 ± 83.51	38 ± 66.65	0.27	0.95	0.72
EC.3.4.24.28	Bacillolysin	93.28	0.9	61.2 ± 116.17	35.41 ± 86.1	0.26	0.53	0.31
EC.3.5.1.26	N(4)-(beta-N-acetylglucosaminyl)-L-asparaginase	99.89	0.9	796.47 ± 490.27	586.32 ± 522.09	0.27	0.61	0.47
EC.4.1.2.43	3-hexulose-6-phosphate synthase	99.63	0.9	506.15 ± 398.87	324.66 ± 353.13	0.45	0.62	0.92
EC.4.2.1.151	Chorismate dehydratase	96.57	0.9	28.68 ± 62.23	58.49 ± 104.49	0.35	0.65	0.71
EC.4.2.1.22	Cystathionine beta-synthase	94.83	0.9	49.66 ± 79.4	24.66 ± 53.17	0.45	0.52	0.70
EC.5.1.3.22	L-ribulose-5-phosphate 3-epimerase	98.94	0.9	48.43 ± 105.46	42.4 ± 106.8	0.42	0.79	0.40
EC.5.1.3.7	UDP-N-acetylglucosamine 4-epimerase	99.44	0.9	439.44 ± 478.54	613.75 ± 508.28	0.20	0.80	0.41

EC.5.4.99.16	Maltose alpha-D-glucosyltransferase	98.94	0.9	59.09 ± 80.69	37.37 ± 94.58	0.55	0.61	0.79
EC.5.4.99.28	tRNA pseudouridine(32) synthase	98.02	0.9	62.74 ± 83.22	44.86 ± 106.19	0.44	0.51	0.75
EC.5.4.99.29	23S rRNA pseudouridine(746) synthase	98.02	0.9	62.74 ± 83.22	44.86 ± 106.19	0.43	0.65	0.89
EC.6.2.1.5	Succinate--CoA ligase (ADP-forming)	99.66	0.9	260.63 ± 355.16	197.54 ± 383.41	0.45	0.56	0.85
EC.6.4.1.3	Propionyl-CoA carboxylase	96.34	0.9	92.73 ± 115.63	62.37 ± 159.22	0.46	0.67	0.82

[a]ANCOM2.1 was performed after adjusting the generation number in the model. All W-statistic cutoffs from ANCOM output (0.9) were denoted for each taxon.

5.5.2. Gut microbiota diversity associated with cardiometabolic traits

Since we identified significant sex differences on cardiometabolic traits in Chapter 4, we next performed Spearman correlation analysis in all mice, and both females and males separately to elucidate the relationship between 20 cardiometabolic traits and gut microbial diversity. In the DO-F1 mice, traits such as aortic lesion area, plasma TC, plasma VLDL/LDL-C, hepatic TC, and BPM among 24 traits positively correlated with the Shannon diversity index, Observed ASV, Plelou's evenness, Weighted UniFrac PC2, and Bray Curtis PC1, and showed a negative correlation with Unweighted UniFrac PC1 (**Figure 5.3A**). These patterns were similar in correlation analysis with cardiometabolic traits in females and males, respectively (**Figures 5.3B and 5.3C**). Among the correlations we found, aortic lesion area, plasma TC, and hepatic TC, showed the strongest effect on the gut microbial diversities. These traits we define as "atherosclerotic traits".

Next, we performed a multivariate regression analysis to evaluate how gut microbiota diversity contributes to the atherosclerotic traits (aortic lesion area, plasma TC, and hepatic TC). In multivariate regression analysis, a significant association with gut microbiota diversity was observed for three atherosclerotic traits in a model with adjusted sex, generation number, and kinship matrix (**Table 5.5**). Similar to Spearman correlation, aortic lesion area, plasma TC, and hepatic TC were associated with increased alpha-diversity, Weighted UniFrac PC1, and Bray Curtis PC1, and associated with decreased Unweighted UniFrac PC1. Only aortic lesion area showed significant association with beta-diversity, not alpha-diversity. In addition, the microbial diversity contributing to the proportion of variance in traits was high in the index considering abundance or abundance and phylogenetic diversity together. For example, the Shannon index in alpha-diversity, which considers phylogenetic diversity and abundance, and abundance-based

Bray Curtis index in beta-diversity explained the highest proportion of variance in traits (**Table 5.5**). Therefore, these results suggest that proportion of variance in key atherosclerotic traits can be explained by gut microbial diversity affected by the HFHC diet.

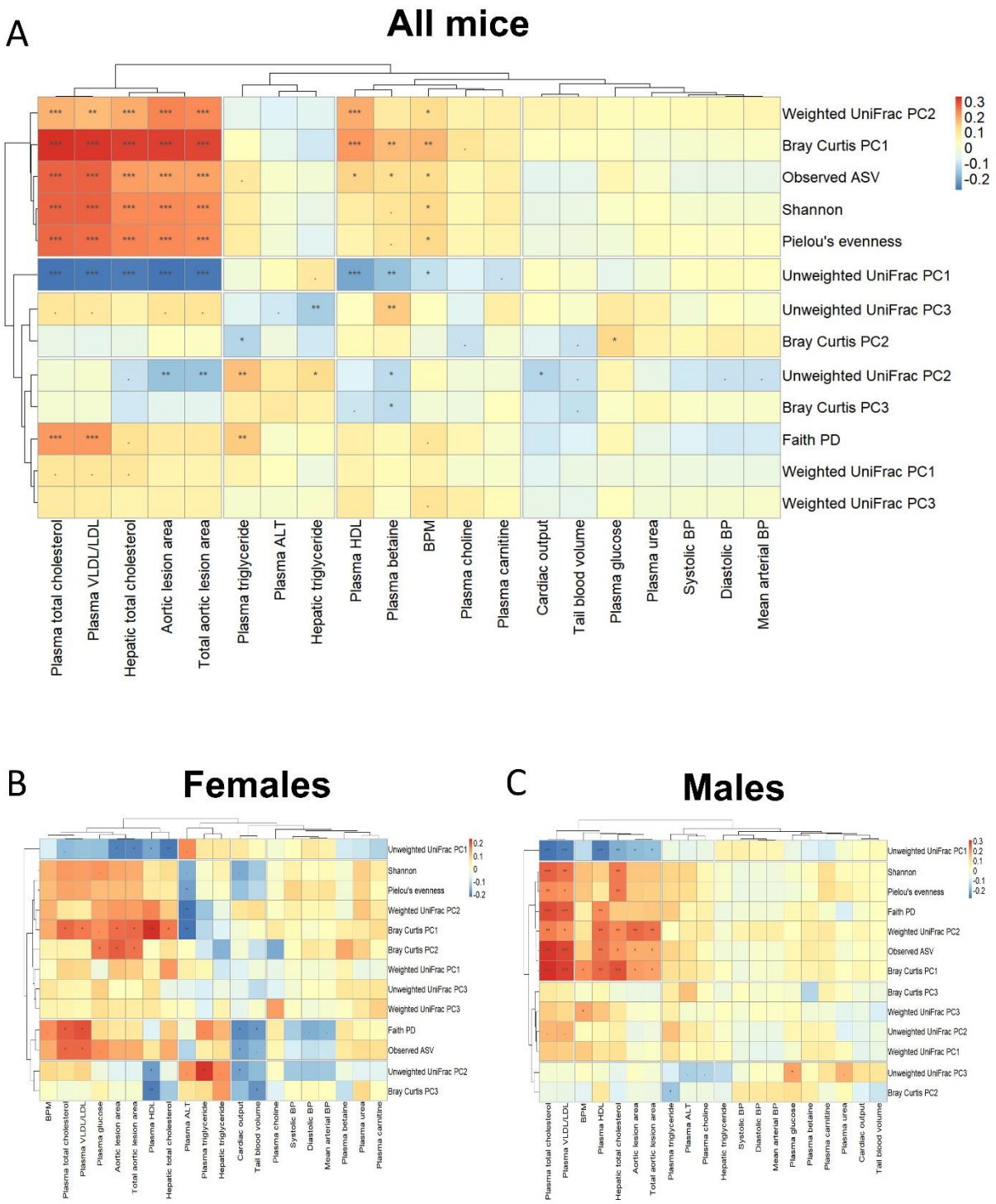


Figure 5.3. Cardiometabolic traits associated with gut microbiota diversity in DO-F1 mice. (A-C) Spearman correlation between microbial diversity and cardio-metabolic traits in all mice (A), females (B), and males (C). The p-values were adjusted using the Benjamini-Hochberg (BH) FDR procedure. "***"$P < 0.001$, "**"$P < 0.01$, "*"$P < 0.05$, "."$P < 0.10$.

Table 5.5. A multivariate linear mixed-effect regression model assessing three atherosclerotic traits associated with gut microbiota diversity in DO-F1 mice.

		Adjusted for sex, generation number, and kinship[a]				
Aortic lesion area		Estimate	CI_Lo[b]	CI_Hi[b]	R^2 explained (%)	p-value
	Shannon	0.691	-0.305	1.687	5.0	0.174
α-diversity	Observed ASVs	0.395	-0.263	1.053	3.8	0.240
	Faith PD	0.005	-1.250	1.261	0.8	0.993
	Weighted Unifrac PC1	-0.110	-4.409	4.188	7.6	0.960
β-diversity	Unweighted Unifrac PC1	-2.003	-3.441	-0.564	0.3	**0.006**
	Bray curtis PC1	1.106	0.296	1.917	9.1	**0.007**
Plasma total cholesterol		Estimate	CI_Lo	CI_Hi	R^2 explained (%)	p-value
	Shannon	0.347	0.174	0.520	7.2	**8.48E-05**
α-diversity	Observed ASVs	0.301	0.188	0.414	7.2	**1.93E-07**
	Faith PD	0.460	0.244	0.676	4.3	**3.00E-05**
	Weighted Unifrac PC1	0.923	0.166	1.680	7.7	**0.017**
β-diversity	Unweighted Unifrac PC1	-0.654	-0.907	-0.402	1.0	**3.66E-07**
	Bray curtis PC1	0.323	0.180	0.466	10.5	**9.51E-06**
Hepatic total cholesterol		Estimate	CI_Lo	CI_Hi	R^2 explained (%)	p-value
	Shannon	0.147	0.035	0.259	5.1	**0.010**
α-diversity	Observed ASVs	0.112	0.039	0.186	4.0	**0.003**
	Faith PD	0.117	-0.021	0.256	1.2	0.097
	Weighted Unifrac PC1	0.688	0.206	1.169	6.9	**0.005**
β-diversity	Unweighted Unifrac PC1	-0.348	-0.512	-0.183	1.0	**3.56E-05**
	Bray curtis PC1	0.199	0.107	0.292	9.3	**2.56E-05**

[a]Multivariate associations were assessed by using linear mixed-effect regression models after adjusting sex, generation number, and genetic relatedness (kinship).

[b]CI_Lo, Lower confidence interval; CI_Hi: Higher confidence interval

5.5.3. Differentially abundant microbial genera and functional profiling by atherosclerotic traits

Next, we investigated which microbial genera or functional pathways were commonly associated with the three atherosclerotic traits that showed the most significant association with gut microbial diversity. We performed ANCOM based on the median value of each three atherosclerotic traits to detect potential genera that most likely explain the difference between the two groups. Based on W statistics by ANCOM, we found differentially abundant 23 genera for aortic lesion area (**Figure 5.4A**), 17 genera for plasma TC (**Figure 5.4B**), and 10 genera for hepatic TC (**Figure 5.4C**), between high and low groups of each trait after adjusting sex and generation (**Table 5.6**). Among them, genera that showed association with at least three atherosclerotic traits including aortic lesion area were Lachnospiraceae bacterium A2, *Alistipes*, *Blautia*, Lachnospiraceae NK4A136 group, *Ruminiclostridium*, *Bifidobacterium*, *Clostridium sensu stricto* 13, *Escherichia-Shigella*, and *Faecalibaculum*, and *Terrisporobacter*. In general, Lachnospiraceae bacterium A2, *Alistipes*, *Blautia*, Lachnospiraceae NK4A136 group, and *Ruminiclostridium* were significantly abundant in the high group for aortic lesion area, plasma TC, and hepatic TC while *Bifidobacterium*, *Clostridium sensu stricto* 13, *Escherichia-Shigella*, *Faecalibaculum*, *Terrisporobacter* were significantly enriched in the low group for these traits (**Figures 5.4A-C and Table 5.6**).

We similarly investigated the MetaCyc pathways and EC gene families that are commonly associated with three atherosclerotic traits. We identified differentially abundant MetaCyc pathways (W cutoff >0.7, **Table 5.7**) and EC gene families (W cutoff >0.9, **Table 5.8**) between high and low groups for aortic lesion area, plasma TC, and hepatic TC, and 64 pathways and 448 EC gene families were differentially abundant in all three atherosclerotic traits. The

pathways that have the strongest difference with W cutoff 0.9 level and |log2 fold change|> were "1,4-dihydroxy-6-naphthoate biosynthesis I" and "1,4-dihydroxy-6-naphthoate biosynthesis II", and "superpathway of menaquinol-8 biosynthesis II" (**Figure 5.5A-5.5C and Table 5.7**). These three pathways are associated with the synthesis of menaquinones, which are lipophilic components of the cytoplasmic membrane (Hiratsuka et al., 2008). Similarly, EC gene families with the greatest difference with the W cutoff level of 0.9 and the |log2 fold change|>1 were "Glycogen (starch) synthase", "Cysteate synthase", "Chorismate dehydratase", "Cyclic dehypoxanthinyl futalosine synthase", "Aminodeoxyfutalosine synthase", "Prolyl oligopeptidase", "Futalosine hydrolase" (**Figure 5.5D-5.5F and Table 5.8**). Among them, “Chorismate dehydratase”, “Cyclic dehypoxanthinyl futalosine synthase”, “Aminodeoxyfutalosine synthase” and “Futalosine hydrolase” ECs are involved in three pathways associated with the synthesis of menaquinones (Caspi et al., 2016). These results suggest that specific microbial genera and their associated functional microbial pathways may influence the pathogenesis of atherosclerosis.

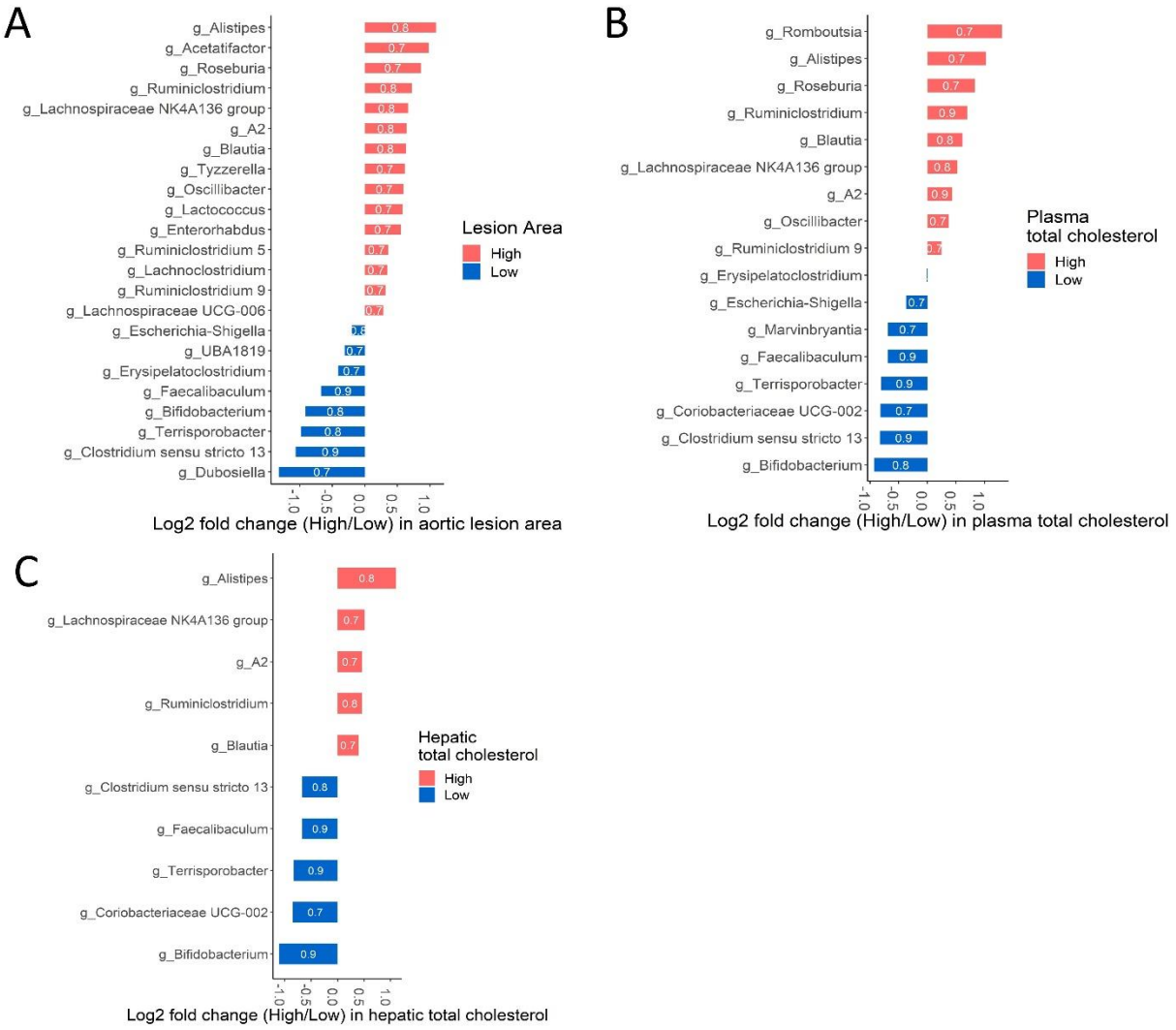


Figure 5.4. Differentially abundant taxa by three atherosclerotic traits in DO-F1 mice. (A-C) ANCOM analysis of microbial genera by median value of aortic lesion area (A), plasma total cholesterol (B), and hepatic total cholesterol (C) at 24 weeks after adjusting sex and generation number. Differentially abundant taxa were ordered from top to bottom by log2 fold change of mean between high (red color) and low group (blue color) in three key traits. All W-statistic cutoffs from ANCOM output (0.7, 0.8, and 0.9) were denoted for each taxon; for interpretation, significance was defined as W>0.7.

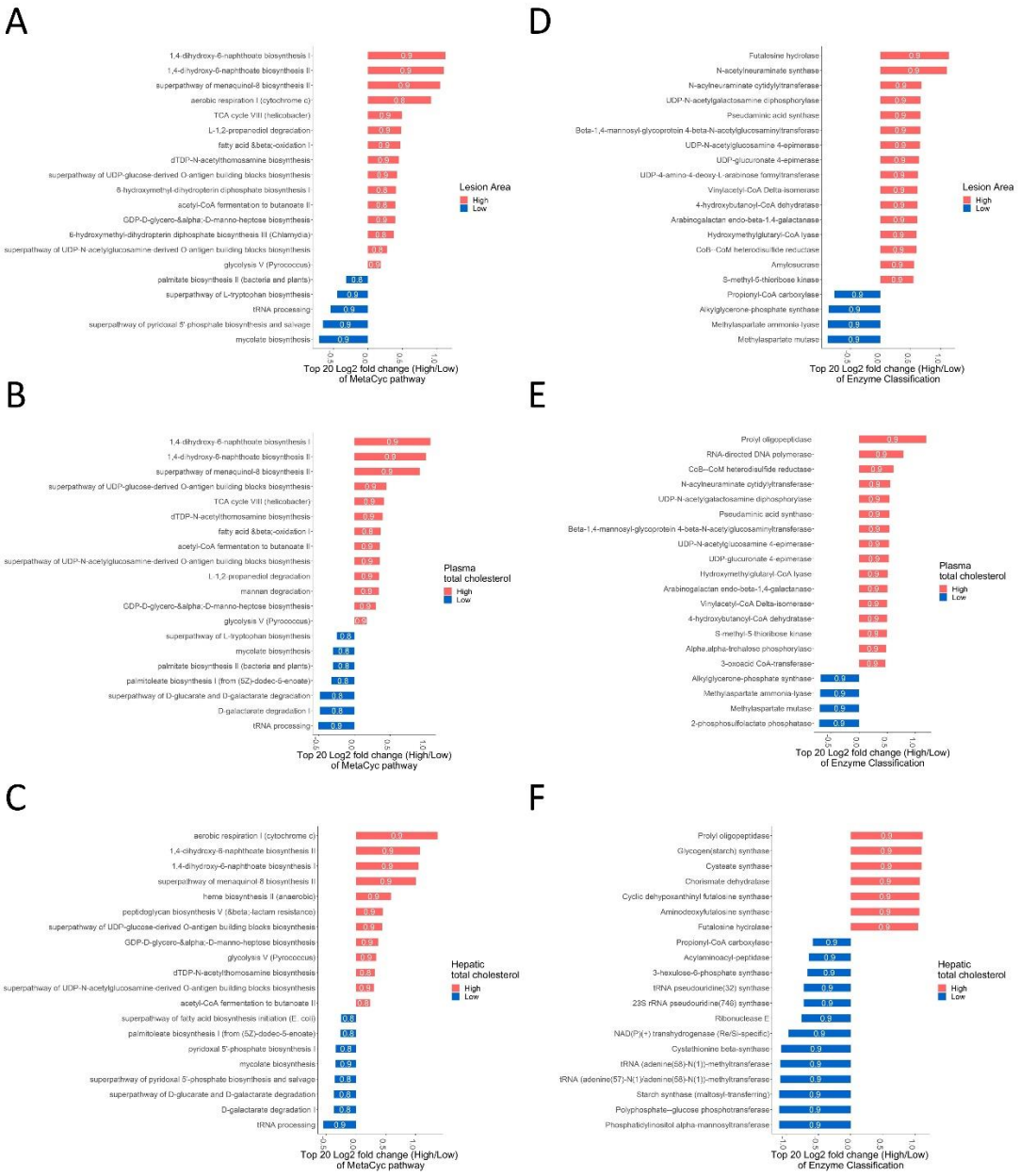


Figure 5.5. Differentially abundant MetaCyc pathways and EC gene families by three atherosclerotic traits in DO-F1 mice. (A-C) ANCOM analysis of MetaCyc pathways by a median value of aortic lesion area (A), plasma total cholesterol (B), and hepatic total cholesterol (C) after adjusting sex and generation number. (D-F) ANCOM analysis of EC gene families by a median value of aortic lesion area (D), plasma total cholesterol (E), and hepatic total cholesterol (F) after adjusting sex and generation number. The top 20 differentially abundant MetaCyc pathways and EC gene families that have the highest W values were ordered from top to bottom by log2 fold change of mean between high (red color) and low group (blue color) in three key traits. All W-statistic cutoffs from ANCOM output (0.7, 0.8, and 0.9) were denoted for each MetaCyc pathway; for interpretation, significance was defined as W>0.7.

Table 5.6. Differentially abundant genera by three atherosclerotic traits in DO-F1 mice at 24 weeks. n = 464 (232 females and 232 males).

Genera	Aortic lesion area		Plasma total cholesterol		Hepatic total cholesterol	
	W cutoff (ANCOM)[a]	Log2FC (High/Low)	W cutoff (ANCOM)[a]	Log2FC (High/Low)	W cutoff (ANCOM)[a]	Log2FC (High/Low)
g_Faecalibaculum	**0.9**	**-0.7**	**0.9**	**-0.7**	**0.9**	**-0.7**
g_Clostridium sensu stricto 13	**0.9**	**-1.1**	**0.9**	**-0.8**	0.8	**-0.7**
g_Terrisporobacter	0.8	**-1.0**	**0.9**	**-0.8**	**0.9**	**-0.8**
g_Ruminiclostridium	0.8	**0.7**	**0.9**	**0.7**	0.8	**0.5**
g_A2	0.8	**0.6**	**0.9**	**0.4**	0.7	**0.5**
g_Bifidobacterium	0.8	**-0.9**	0.8	**-0.9**	**0.9**	**-1.1**
g_Lachnospiraceae NK4A136 group	0.8	**0.7**	0.8	**0.5**	0.7	**0.5**
g_Blautia	0.8	**0.6**	0.8	**0.6**	0.7	**0.4**
g_Alistipes	0.8	**1.1**	0.7	**1.0**	0.8	**1.1**

[a]ANCOM2.1 was performed after adjusting sex and generation number in the model. All W-statistic cutoffs from ANCOM output (0.7, 0.8, and 0.9) were denoted for each genus.

Table 5.7. Differentially abundant MetaCyc pathways by three atherosclerotic traits in DO-F1 mice at 24 weeks.

Pathway	Aortic lesion area		Plasma total cholesterol		Hepatic total cholesterol	
	W cutoff (ANCOM)[a]	Log2FC (High/Low)	W cutoff (ANCOM)[a]	Log2FC (High/Low)	W cutoff (ANCOM)[a]	Log2FC (High/Low)
superpathway of menaquinol-8 biosynthesis II	**0.9**	**1.0**	**0.9**	**0.9**	**0.9**	**1.0**
1,4-dihydroxy-6-naphthoate biosynthesis II	**0.9**	**1.1**	**0.9**	**1.0**	**0.9**	**1.1**
1,4-dihydroxy-6-naphthoate biosynthesis I	**0.9**	**1.1**	**0.9**	**1.1**	**0.9**	**1.0**
tRNA processing	**0.9**	**-0.5**	**0.9**	**-0.5**	**0.9**	**-0.5**
glycolysis V (Pyrococcus)	**0.9**	**0.2**	**0.9**	**0.2**	**0.9**	**0.3**
superpathway of UDP-glucose-derived O-antigen building blocks biosynthesis	**0.9**	**0.4**	**0.9**	**0.4**	**0.9**	**0.4**
GDP-D-glycero-α-D-manno-heptose biosynthesis	**0.9**	**0.4**	**0.9**	**0.3**	**0.9**	**0.4**
dTDP-N-acetylthomosamine biosynthesis	**0.9**	**0.4**	**0.9**	**0.4**	0.8	**0.3**
L-1,2-propanediol degradation	**0.9**	**0.5**	**0.9**	**0.3**	0.7	**0.2**
TCA cycle VIII (helicobacter)	**0.9**	**0.5**	**0.9**	**0.4**	0.7	**0.3**
mycolate biosynthesis	**0.9**	**-0.7**	0.8	**-0.3**	**0.9**	**-0.3**
superpathway of L-tryptophan biosynthesis	**0.9**	**-0.4**	0.8	**-0.2**	0.7	**-0.2**
superpathway of pyridoxal 5'-phosphate biosynthesis and salvage	**0.9**	**-0.7**	0.7	**-0.4**	0.8	**-0.4**
superpathway of UDP-N-acetylglucosamine-derived O-antigen building blocks biosynthesis	0.8	**0.3**	**0.9**	**0.4**	**0.9**	**0.3**
acetyl-CoA fermentation to butanoate II	0.8	**0.4**	**0.9**	**0.4**	0.8	**0.2**
heme biosynthesis II (anaerobic)	0.8	**0.4**	0.8	**0.5**	**0.9**	**0.6**
superpathway of fatty acid biosynthesis initiation (E. coli)	0.8	**-0.3**	0.8	**-0.3**	0.8	**-0.3**
palmitate biosynthesis II (bacteria and plants)	0.8	**-0.3**	0.8	**-0.3**	0.8	**-0.2**
stearate biosynthesis II (bacteria and plants)	0.8	**-0.3**	0.8	**-0.3**	0.8	**-0.3**
palmitoleate biosynthesis I (from (5Z)-dodec-5-enoate)	0.8	**-0.3**	0.8	**-0.3**	0.8	**-0.3**
8-amino-7-oxononanoate biosynthesis I	0.8	**-0.4**	0.8	**-0.3**	0.8	**-0.2**
oleate biosynthesis IV (anaerobic)	0.8	**-0.3**	0.8	**-0.3**	0.8	**-0.2**
D-galactarate degradation I	0.8	**-0.4**	0.8	**-0.5**	0.8	**-0.4**
superpathway of D-glucarate and D-galactarate degradation	0.8	**-0.4**	0.8	**-0.5**	0.8	**-0.4**
biotin biosynthesis I	0.8	**-0.4**	0.8	**-0.3**	0.7	**-0.2**
superpathway of lipopolysaccharide biosynthesis	0.8	**-0.4**	0.8	**-0.3**	0.7	**-0.2**
6-hydroxymethyl-dihydropterin diphosphate biosynthesis I	0.8	**0.4**	0.8	**0.3**	0.7	**0.3**
superpathway of L-arginine, putrescine, and 4-aminobutanoate degradation	0.8	**-0.3**	0.8	**-0.3**	0.7	**-0.3**
enterobacterial common antigen biosynthesis	0.8	**-0.2**	0.8	**-0.3**	0.7	**-0.3**
superpathway of glycol metabolism and degradation	0.8	**-0.3**	0.8	**-0.4**	0.7	**-0.3**
superpathway of methylglyoxal degradation	0.8	**-0.2**	0.8	**-0.4**	0.7	**-0.3**
superpathway of L-arginine and L-ornithine degradation	0.8	**-0.3**	0.8	**-0.3**	0.7	**-0.3**

peptidoglycan biosynthesis V (β-lactam resistance)	0.8	**0.5**	0.7	**0.4**	**0.9**	**0.4**
pyridoxal 5'-phosphate biosynthesis I	0.8	**-0.6**	0.7	**-0.3**	0.8	**-0.3**
glucose and glucose-1-phosphate degradation	0.8	**-0.7**	0.7	**-0.5**	0.7	**-0.3**
6-hydroxymethyl-dihydropterin diphosphate biosynthesis III (Chlamydia)	0.8	**0.4**	0.7	**0.3**	0.7	**0.3**
L-arginine degradation II (AST pathway)	0.8	**-0.2**	0.7	**-0.4**	0.7	**-0.3**
3-phenylpropanoate and 3-(3-hydroxyphenyl)propanoate degradation to 2-oxopent-4-enoate	0.8	**-0.2**	0.7	**-0.4**	0.7	**-0.3**
3-phenylpropanoate degradation	0.8	**-0.2**	0.7	**-0.3**	0.7	**-0.3**
2-methylcitrate cycle II	0.8	**-0.2**	0.7	**-0.3**	0.7	**-0.3**
superpathway of phenylethylamine degradation	0.8	**-0.2**	0.7	**-0.3**	0.7	**-0.3**
cinnamate and 3-hydroxycinnamate degradation to 2-oxopent-4-enoate	0.8	**-0.2**	0.7	**-0.4**	0.7	**-0.3**
sulfoglycolysis	0.8	**-0.2**	0.7	**-0.4**	0.7	**-0.3**
superpathway of heme biosynthesis from uroporphyrinogen-III	0.8	**-0.2**	0.7	**-0.3**	0.7	**-0.2**
phenylacetate degradation I (aerobic)	0.8	**-0.2**	0.7	**-0.3**	0.7	**-0.3**
2-methylcitrate cycle I	0.8	**-0.2**	0.7	**-0.3**	0.7	**-0.3**
polymyxin resistance	0.8	**-0.3**	0.7	**-0.2**	0.7	**-0.2**
L-lysine biosynthesis II	0.8	**-0.2**	0.7	**-0.1**	0.7	**-0.1**
superpathway of ubiquinol-8 biosynthesis (prokaryotic)	0.7	**-0.3**	0.8	**-0.3**	0.7	**-0.2**
fatty acid elongation -- saturated	0.7	**0.3**	0.7	**0.3**	0.7	**0.3**
ADP-L-glycero-β-D-manno-heptose biosynthesis	0.7	**-0.6**	0.7	**-0.5**	0.7	**-0.3**
superpathway of ornithine degradation	0.7	**-0.2**	0.7	**-0.3**	0.7	**-0.3**
ubiquinol-7 biosynthesis (prokaryotic)	0.7	**-0.3**	0.7	**-0.3**	0.7	**-0.2**
ubiquinol-9 biosynthesis (prokaryotic)	0.7	**-0.3**	0.7	**-0.3**	0.7	**-0.2**
ubiquinol-10 biosynthesis (prokaryotic)	0.7	**-0.3**	0.7	**-0.3**	0.7	**-0.2**
ubiquinol-8 biosynthesis (prokaryotic)	0.7	**-0.3**	0.7	**-0.3**	0.7	**-0.2**
3-phenylpropanoate and 3-(3-hydroxyphenyl)propanoate degradation	0.7	**-0.3**	0.7	**-0.3**	0.7	**-0.3**
superpathway of chorismate metabolism	0.7	**-0.3**	0.7	**-0.2**	0.7	**-0.1**
D-glucarate degradation I	0.7	**-0.3**	0.7	**-0.2**	0.7	**-0.2**
superpathway of (Kdo)2-lipid A biosynthesis	0.7	**-0.4**	0.7	**-0.2**	0.7	**-0.2**
catechol degradation I (meta-cleavage pathway)	0.7	**-0.2**	0.7	**-0.3**	0.7	**-0.2**
methylphosphonate degradation I	0.7	**-0.6**	0.7	**-0.6**	0.7	**-0.4**
TCA cycle I (prokaryotic)	0.7	**-0.4**	0.7	**-0.2**	0.7	**-0.3**
ppGpp biosynthesis	0.7	**-0.2**	0.7	**-0.3**	0.7	**-0.2**

[a]ANCOM2.1 was performed after adjusting sex and generation number in the model. All W-statistic cutoffs from ANCOM output (0. 7, 0.8, and 0.9) were denoted for each MetaCyc pathways.

Table 5.8. Differentially abundant EC gene families by three atherosclerotic traits in DO-F1 mice.

EC gene families	Description	Aortic lesion area		Plasma total cholesterol		Hepatic total cholesterol	
		W cutoff 0.9 (ANCOM)[a]	Log2FC (High/Low)	W cutoff 0.9 (ANCOM)[a]	Log2FC (High/Low)	W cutoff 0.9 (ANCOM)[a]	Log2FC (High/Low)
EC.1.1.1.61	4-hydroxybutyrate dehydrogenase	0.9	-0.68	0.9	-0.61	0.9	-0.65
EC.1.2.99.7	Aldehyde dehydrogenase (FAD-independent)	0.9	-0.96	0.9	-0.78	0.9	-0.81
EC.1.21.98.1	Cyclic dehypoxanthinyl futalosine synthase	0.9	1.10	0.9	1.00	0.9	1.07
EC.1.6.1.2	NAD(P)(+) transhydrogenase (Re/Si-specific)	0.9	-0.79	0.9	-0.84	0.9	-0.97
EC.1.8.98.1	CoB--CoM heterodisulfide reductase	0.9	0.60	0.9	0.61	0.9	0.52
EC.2.4.1.11	Glycogen(starch) synthase	0.9	1.09	0.9	1.03	0.9	1.10
EC.2.4.1.144	Beta-1,4-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase	0.9	0.66	0.9	0.54	0.9	0.39
EC.2.5.1.120	Aminodeoxyfutalosine synthase	0.9	1.10	0.9	1.00	0.9	1.07
EC.2.5.1.26	Alkylglycerone-phosphate synthase	0.9	-0.85	0.9	-0.68	0.9	-0.51
EC.2.5.1.76	Cysteate synthase	0.9	1.09	0.9	1.03	0.9	1.10
EC.2.5.1.97	Pseudaminic acid synthase	0.9	0.66	0.9	0.54	0.9	0.39
EC.2.7.7.42	[Glutamate--ammonia-ligase] adenylyltransferase	0.9	-0.72	0.9	-0.79	0.9	-0.89
EC.2.7.7.43	N-acylneuraminate cytidylyltransferase	0.9	0.68	0.9	0.55	0.9	0.40
EC.2.7.7.83	UDP-N-acetylgalactosamine diphosphorylase	0.9	0.66	0.9	0.54	0.9	0.38
EC.3.1.3.71	2-phosphosulfolactate phosphatase	0.9	-0.74	0.9	-0.71	0.9	-0.65
EC.3.2.2.26	Futalosine hydrolase	0.9	1.13	0.9	1.08	0.9	1.05
EC.3.4.21.26	Prolyl oligopeptidase	0.9	1.10	0.9	1.19	0.9	1.12
EC.3.4.21.83	Oligopeptidase B	0.9	-0.72	0.9	-0.79	0.9	-0.89
EC.3.5.1.59	N-carbamoylsarcosine amidase	0.9	-0.96	0.9	-0.79	0.9	-0.81
EC.3.6.3.20	Glycerol-3-phosphate-transporting ATPase	0.9	-0.72	0.9	-0.41	0.9	-0.39
EC.4.2.1.151	Chorismate dehydratase	0.9	1.09	0.9	1.01	0.9	1.08
EC.4.3.1.2	Methylaspartate ammonia-lyase	0.9	-0.87	0.9	-0.69	0.9	-0.68
EC.5.1.3.22	L-ribulose-5-phosphate 3-epimerase	0.9	-0.61	0.9	-0.55	0.9	-0.48
EC.5.1.3.6	UDP-glucuronate 4-epimerase	0.9	0.64	0.9	0.53	0.9	0.38
EC.5.1.3.7	UDP-N-acetylglucosamine 4-epimerase	0.9	0.66	0.9	0.53	0.9	0.39
EC.5.4.99.1	Methylaspartate mutase	0.9	-0.87	0.9	-0.70	0.9	-0.68

[a]ANCOM2.1 was performed after adjusting sex and generation number in the model. All W-statistic cutoffs from ANCOM output (0.9) were denoted for each EC gene family

5.5.4. Microbial taxa and functional profiling are associated with host genetics

We next performed quantitative trait loci (QTL) mapping using the R/qtl2 package (Broman et al., 2019) to investigate the effect of host genetics on microbial taxa and functional profiling, and to identify loci that are genetically regulated concurrently with cardiometabolic traits and gut microbiota. Specifically in the atherosclerosis mouse model, since females are more susceptible than males, and the effects on host genetics also perturb sexual dimorphism (AlSiraj et al., 2019; Bennett et al., 2015), we calculated heritability and conducted QTL analysis in entire mice, female mice, or male mice, respectively.

First, we estimated narrow-sense heritability (h^2) using a linear mixed model implemented in the R/qtl2 package to predict the effects of the genotype on the microbial taxa, which is proportional to the genetic relatedness between the mice. Therefore, h^2 was calculated by including the traits, kinship matrix, and generation number. In entire mice, we used sex and generation as covariates in h^2 calculation. Sex-specific estimates of h^2 suggest that there is considerable sex specificity in host genetics to the observed variance in microbial taxa (**Figures 5.6, and Table 5.2**), MetaCyc pathways (**Table 5.3**), and EC gene families (**Table 5.4**). In the 40 microbial genera levels identified, h^2 was generally higher in females (60% of genera levels) than in males, but the proportion of genera whose h^2 difference was less than 0.1 between the two sexes was almost half of the total (47.5% of genera levels). In both sexes, genera with h^2 less than 0.5 include 8 genera (*Staphylococcus*, *Lactobacillus*, Lachnospiraceae FCS020 group, Lachnospiraceae UCG-004, Lachnospiraceae UCG-006, Anaerotruncus, GCA-900066225, and Ruminococcaceae UCG-010) (**Table 5.2**). On the other hand, h^2 was higher in males in 64% of MetaCyc pathways and EC gene families (**Table 5.3 and 5.4**).

Second, we performed the QTL analysis for microbial taxa, MetaCyc pathways, and EC gene families, respectively. In entire mice, we used sex and generation number as covariates and identified significant QTLs ($p<0.05$) for 31 microbial taxa (16 ASVs, 9 genera, 3 family, 1 class, 1 order, and 1 phylum) (**Table 5.9**), 93 MetaCyc pathways (**Table 5.10**), and 499 EC gene families (data not shown). In addition, we found a substantial effect of host genetics on the sex-specific mbQTL results (**Table 5.9-5.10**). For example, most of the QTLs we identified in each sex showed significant association in only one sex and different chromosome in even same-sex. We applied a highly suggestive threshold ($P<0.1$) to visualize sex-specific microbial genera QTL in the form of circos plots in order to capture a more complete picture of the genetic architecture of the microbiota. **Figure 5.6** showed sex-biased 18 out of the 40 microbial genus levels.

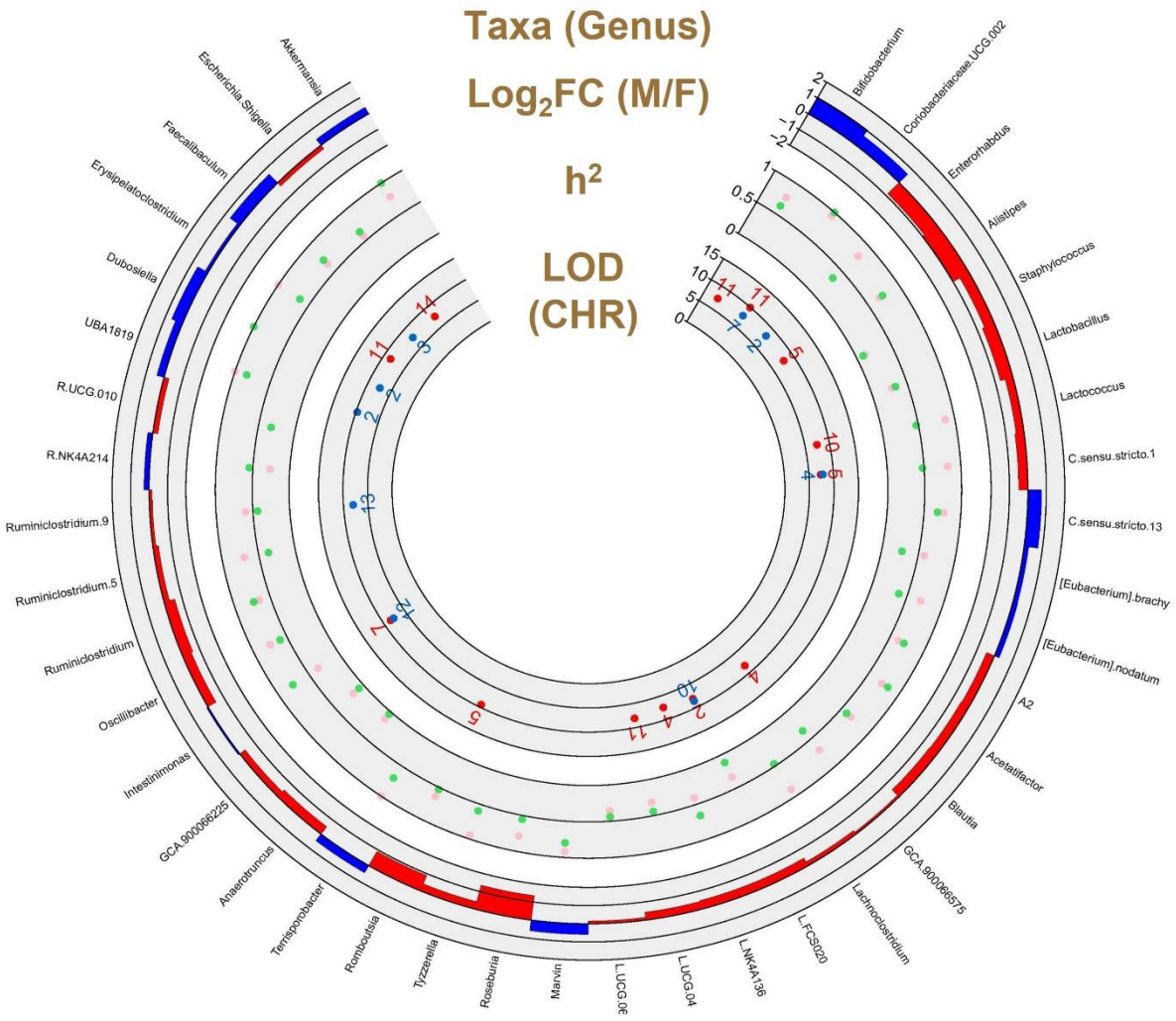


Figure 5.6. Genetic architecture of quantitative trait loci (QTL) for microbial taxa abundance in DO-F1 mice.

Circos plot showing an overview of genetic regulation of microbial genus levels at 24 weeks. The outermost track (blue bars for males and red bars for females) show the relative log2fold change (FC) (male versus female). The middle track represents narrow-sense heritability (h^2) estimates for the microbial genus levels in females (pink) and males (green). The innermost track represents logarithm of the odds (LOD) scores of highly suggestive (P<0.1) quantitative trait loci (QTL) in females (red) and males (blue), with their respective chromosomes indicated below (male) or above (female) dots. Abbreviation:.sensu.stricto.1, *Clostridium sensu stricto 1*; C.sensu.stricto.13 , *Clostridium sensu stricto 13*; L.FCS020, Lachnospiraceae FCS020; L.NK4A136, Lachnospiraceae NK4A136; L.UCG.04, Lachnospiraceae UCG004; L.UCG.06, Lachnospiraceae UCG006; R.NK4A214, Ruminococcaceae NK4A214; R.UCG.010, Ruminococcaceae UCG-010.

Table 5.9. Significant QTL results for gut microbial taxa at 24 weeks in three models. n = 454 (229 females and 225 males)

Model	Taxonomy	Taxa	% of presence in entire mice	Chr	Position (Mbp)	CI (low)[a]	CI(hi)[a]	LOD	LOD threshold (p < 0.05)	#Genes	Marker
Female mice	ASV	a_.Eubacterium..brachy.group_ASV1	20.26	3	73.60	72.77	75.56	7.92	7.82	4	UNCHS009043
Female mice	ASV	a_Acetatifactor_ASV3	41.59	4	69.54	68.48	71.28	9.86	7.71	4	JAX00553788
Female mice	ASV	a_Coriobacteriaceae.UCG.002_ASV1	32.11	11	102.36	101.63	103.25	9.93	7.96	50	JAX00320303
Female mice	ASV	a_Erysipelatoclostridium_ASV1	87.72	1	82.92	78.27	87.80	7.39	7.39	59	UNCHS001435
Female mice	ASV	a_Erysipelatoclostridium_ASV1	87.72	7	44.07	40.71	44.63	7.54	7.39	50	JAX00637206
Female mice	ASV	a_Erysipelatoclostridium_ASV1	87.72	11	53.80	53.24	56.97	8.77	7.39	44	JAX00311270
Female mice	ASV	a_GCA.900066575_ASV14	69.18	17	64.96	63.60	66.15	7.95	7.71	16	JAX00443727
Female mice	ASV	a_GCA.900066575_ASV15	44.83	7	29.55	29.40	29.80	8.93	7.62	4	UNC12559214
Female mice	ASV	a_Intestinimonas_ASV1	58.62	7	124.48	123.52	124.60	8.46	7.74	0	UNC13780242
Female mice	ASV	a_Lachnospiraceae.NK4A136.group_ASV16	20.69	8	111.92	110.70	112.08	8.55	8.10	29	UNC15503960
Female mice	ASV	a_Lachnospiraceae.NK4A136.group_ASV5	79.53	8	118.71	118.35	119.17	7.71	7.65	0	UNCHS024341
Female mice	ASV	a_Lachnospiraceae.UCG.006_ASV3	54.96	3	50.84	48.29	51.74	7.82	7.82	9	UNC5241793
Female mice	ASV	a_Lactococcus_ASV1	100	10	26.71	25.13	27.06	7.45	7.73	8	UNC17630038
Female mice	ASV	a_Romboutsia_ASV1	23.92	5	64.73	64.44	66.48	8.68	7.83	21	UNCHS014527
Female mice	ASV	a_Romboutsia_ASV2	27.59	2	38.96	32.48	39.33	8.31	7.74	67	UNCHS004706
Female mice	ASV	a_Romboutsia_ASV2	27.59	5	65.91	64.49	66.48	9.14	7.74	21	UNCHS014562
Female mice	ASV	a_Ruminiclostridium.5_ASV10	35.13	13	95.15	95.08	95.58	8.29	7.87	4	UNC23132968
Female mice	Genus	g_Alistipes	64.44	5	137.01	136.75	137.27	7.83	7.67	10	UNC10262562
Female mice	Genus	g_Coriobacteriaceae.UCG.002	32.11	11	102.36	101.63	103.25	9.93	7.97	50	JAX00320303
Female mice	Genus	g_Erysipelatoclostridium	88.58	11	53.80	53.24	56.97	8.48	7.52	44	JAX00311270
Female mice	Genus	g_GCA.900066575	90.73	4	58.16	57.88	59.35	8.27	7.60	13	UNC7328598
Female mice	Genus	g_Intestinimonas	58.62	7	124.48	123.52	124.60	8.46	7.74	0	UNC13780242
Female mice	Genus	g_Lachnospiraceae.FCS020.group	37.28	2	110.79	108.92	114.09	8.05	7.82	19	UNC3750106
Female mice	Genus	g_Lactococcus	100	10	26.71	25.13	27.06	7.45	7.79	8	UNC17630038
Female mice	Genus	g_Romboutsia	31.03	2	35.00	32.48	39.33	8.24	8.00	67	UNCHS004572
Female mice	Genus	g_Romboutsia	31.03	5	65.91	64.49	66.39	9.43	8.00	21	UNCHS014561
Male mice	ASV	a_Coriobacteriaceae.UCG.002_ASV1	32.11	7	134.85	134.81	138.04	7.73	7.62	7	UNC13917049
Male mice	ASV	a_Enterorhabdus_ASV2	69.18	5	117.75	117.74	118.46	7.98	7.80	8	UNC10001966
Male mice	ASV	a_Enterorhabdus_ASV4	32.54	2	165.96	165.86	166.19	8.52	8.16	2	UNC4425598
Male mice	ASV	a_Enterorhabdus_ASV4	32.54	7	122.98	122.61	123.97	8.18	8.16	7	UNC13763557
Male mice	ASV	a_GCA.900066575_ASV9	64.87	16	66.88	65.54	68.44	8.55	7.82	0	UNC27018343
Male mice	ASV	a_Ruminiclostridium.9_ASV4	74.57	13	45.86	45.61	47.27	8.72	7.87	12	UNCHS035791
Male mice	ASV	a_Ruminiclostridium.9_ASV5	53.23	5	133.25	132.36	134.79	9.06	7.90	12	UNCHS016012
Male mice	ASV	a_Ruminiclostridium.9_ASV5	53.23	11	61.72	60.70	63.08	8.67	7.90	46	UNCHS030977
Male mice	ASV	a_Ruminiclostridium.9_ASV6	32.33	4	107.23	106.60	108.19	8.07	7.89	28	UNCHS012323
Male mice	ASV	a_Ruminiclostridium.9_ASV6	32.33	11	84.84	83.58	85.95	8.01	7.89	30	UNC20039526
Male mice	Class	c_Coriobacteriia	26.72	2	54.43	52.65	56.56	7.75	7.66	6	UNC3086503
Male mice	Class	c_Mollicutes	30.60	6	139.10	138.67	139.21	8.48	8.01	0	UNC12215650
Male mice	Genus	g_Clostridium.sensu.stricto.1	20.47	4	62.98	57.67	64.81	7.84	7.79	45	UNCHS011495
Male mice	Genus	g_Enterorhabdus	90.73	2	181.82	180.82	181.99	8.08	7.96	34	UNCHS008005
Male mice	Genus	g_Lachnospiraceae.FCS020.group	85.56	10	67.51	67.26	67.99	8.63	8.17	4	JAX00019079

Male mice	Genus	g_Lachnospiraceae.FCS020.group	85.56	14	122.69	121.42	122.98	8.30	8.17	12	UNC24928446
Male mice	Genus	g_Ruminiclostridium.9	100.00	13	45.86	45.61	47.22	8.05	7.79	12	UNCHS035791
Male mice	Genus	g_UBA1819	98.49	2	77.37	74.43	78.27	9.74	7.78	22	UNC3375234
Male mice	Order	o_Coriobacteriales	99.14	2	54.43	52.65	56.56	7.75	7.70	6	UNC3086503
Male mice	Order	o_Mollicutes.RF39	63.15	6	139.10	138.67	139.21	8.48	7.98	0	UNC12215650
Male mice	Phylum	p_Tenericutes	98.06	6	139.10	138.67	139.21	8.48	8.17	0	UNC12215650
Entire mice	ASV	a_.Eubacterium..nodatum.group_ASV1	30.39	5	143.64	142.65	144.40	7.97	7.79	26	UNCHS016241
Entire mice	ASV	a_A2_ASV3	36.21	15	50.99	48.95	51.60	9.56	7.86	1	UNC25574114
Entire mice	ASV	a_Acetatifactor_ASV3	41.59	8	9.11	8.68	9.92	8.22	7.78	0	UNC14126015
Entire mice	ASV	a_Acetatifactor_ASV3	41.59	4	71.23	68.63	75.96	8.93	7.78	10	JAX00120272
Entire mice	ASV	a_Akkermansia_ASV1	25.43	4	57.54	57.19	57.88	9.34	8.04	1	UNCHS011398
Entire mice	ASV	a_Blautia_ASV1	23.49	1	83.74	83.69	86.47	8.76	8.17	19	UNC1063044
Entire mice	ASV	a_Blautia_ASV5	70.91	2	116.95	116.21	117.99	8.11	7.84	4	UNC3812685
Entire mice	ASV	a_Enterorhabdus_ASV2	69.18	5	117.75	117.39	117.79	8.25	7.73	1	UNC10001966
Entire mice	ASV	a_Faecalibaculum_ASV2	25.86	9	14.73	14.69	15.69	8.54	8.22	17	JAX00686181
Entire mice	ASV	a_Lachnospiraceae.NK4A136.group_ASV5	79.53	14	121.37	120.32	121.51	9.36	7.80	4	UNCHS039440
Entire mice	ASV	a_Lactobacillus_ASV6	55.39	12	56.35	56.08	56.50	8.08	7.73	1	UNCHS033569
Entire mice	ASV	a_Romboutsia_ASV1	23.92	5	64.73	64.44	66.48	8.09	7.92	21	UNCHS014527
Entire mice	ASV	a_Romboutsia_ASV2	27.59	10	117.97	117.66	119.03	8.83	7.93	7	UNC18812646
Entire mice	ASV	a_Ruminiclostridium.9_ASV6	32.33	2	78.17	78.06	80.32	9.64	7.77	8	UNC3385056
Entire mice	ASV	a_Ruminococcaceae.NK4A214.group_ASV1	31.25	6	72.73	71.71	73.82	8.38	7.80	23	UNC11350970
Entire mice	ASV	a_Terrisporobacter_ASV1	33.19	16	14.44	13.73	17.44	9.01	7.85	38	UNCHS041843
Entire mice	Class	c_Verrucomicrobiae	25.43	4	57.54	57.19	57.88	9.34	7.94	1	UNCHS011398
Entire mice	Family	f_Akkermansiaceae	25.43	4	57.54	57.19	57.88	9.34	7.98	1	UNCHS011398
Entire mice	Family	f_Bifidobacteriaceae	49.57	8	34.64	33.41	34.84	8.43	7.95	14	UNCHS022638
Entire mice	Family	f_Streptococcaceae	100.00	4	35.89	35.75	36.48	8.23	7.73	0	UNC7055184
Entire mice	Genus	g_.Eubacterium..nodatum.group	30.39	5	143.64	142.65	144.40	7.97	7.74	26	UNCHS016241
Entire mice	Genus	g_Akkermansia	25.43	4	57.54	57.19	57.88	9.34	7.87	1	UNCHS011398
Entire mice	Genus	g_Clostridium.sensu.stricto.1	43.97	16	14.25	13.02	17.16	8.06	7.85	38	UNCHS041835
Entire mice	Genus	g_Lachnospiraceae.FCS020.group	37.28	14	122.96	120.87	123.66	8.46	7.81	16	UNC24930227
Entire mice	Genus	g_Lachnospiraceae.NK4A136.group	88.15	14	121.38	121.04	121.51	8.49	7.78	1	UNC24910435
Entire mice	Genus	g_Romboutsia	31.03	10	117.97	117.55	119.03	8.19	7.87	8	UNC18812646
Entire mice	Genus	g_Ruminococcaceae.NK4A214.group	31.25	6	72.73	71.71	73.82	8.38	7.93	23	UNC11350970
Entire mice	Genus	g_Ruminococcaceae.UCG.010	24.35	1	172.48	172.35	173.39	8.64	7.87	17	UNCHS003307
Entire mice	Genus	g_Terrisporobacter	33.62	16	14.44	13.73	17.60	9.02	7.77	45	UNCHS041843
Entire mice	Order	o_Verrucomicrobiales	25.43	4	57.54	57.19	57.88	9.34	8.01	1	UNCHS011398
Entire mice	Phylum	p_Verrucomicrobia	25.43	4	57.54	57.19	57.88	9.34	7.80	1	UNCHS011398

Significance was considered only when QTL results were P<0.05.

[a]95% Bayesian credible interval was calculated as implemented by the bayesint function in R/QTL2.

Table 5.10. Significant QTL results for MetaCyc pathways at 24 weeks in three models. n = 454 (229 females and 225 males)

Model	Name	Chr	Position (Mbp)	CI (low)[a]	CI(hi)[a]	LOD	LOD threshold (p < 0.05)	#Genes	Marker
Female mice	superpathway of arginine and polyamine biosynthesis	1	119.29	118.85	121.06	8.05	7.71	13	UNCHS002076
Female mice	superpathway of polyamine biosynthesis I	1	119.29	118.98	120.80	8.40	7.68	13	UNCHS002076
Female mice	peptidoglycan biosynthesis II (staphylococci)	2	6.97	5.70	9.36	7.90	7.60	11	JAX00482647
Female mice	guanosine nucleotides degradation III	8	118.48	117.61	118.74	8.40	7.73	4	UNC15604827
Female mice	adenosine nucleotides degradation II	8	118.48	118.45	118.81	9.20	7.67	0	UNC15604827
Female mice	purine nucleotides degradation II (aerobic)	8	118.51	117.64	118.81	8.09	7.61	4	UNC15605042
Female mice	preQ0 biosynthesis	12	29.10	28.58	29.34	7.88	7.83	6	UNC20800862
Female mice	formaldehyde assimilation II (RuMP Cycle)	12	81.28	73.62	82.14	8.01	7.58	58	UNCHS034114
Female mice	fucose degradation	12	84.38	83.46	86.41	8.40	7.63	50	UNC21522803
Female mice	thiazole biosynthesis II (Bacillus)	12	100.32	99.98	100.51	7.65	7.54	4	UNC21745067
Female mice	superpathway of thiamin diphosphate biosynthesis II	12	100.32	99.98	100.48	8.21	7.67	4	UNC21745067
Female mice	reductive TCA cycle I	14	58.40	58.27	59.74	7.62	7.57	8	UNC24064293
Female mice	superpathway of chorismate metabolism	14	58.58	58.27	59.74	8.54	7.77	8	B6_rs30650651
Female mice	superpathway of L-arginine, putrescine, and 4-aminobutanoate degradation	14	58.58	58.35	59.74	9.13	7.62	8	B6_rs30650651
Female mice	L-arginine degradation II (AST pathway)	14	58.58	58.35	59.74	8.72	7.70	8	B6_rs30650651
Female mice	biotin biosynthesis I	14	58.58	58.35	59.74	8.70	7.79	8	B6_rs30650651
Female mice	enterobacterial common antigen biosynthesis	14	58.58	58.35	59.74	8.59	7.74	8	B6_rs30650651
Female mice	superpathway of fatty acid biosynthesis initiation (E. coli)	14	58.58	58.35	59.74	8.78	7.76	8	B6_rs30650651
Female mice	D-galactarate degradation I	14	58.58	58.27	59.74	9.40	7.65	8	B6_rs30650651
Female mice	D-glucarate degradation I	14	58.58	58.27	59.74	9.03	7.76	8	B6_rs30650651
Female mice	superpathway of D-glucarate and D-galactarate degradation	14	58.58	58.27	59.74	9.40	7.74	8	B6_rs30650651
Female mice	superpathway of glycol metabolism and degradation	14	58.58	58.35	59.74	9.51	7.73	8	B6_rs30650651
Female mice	3-phenylpropanoate and 3-(3-hydroxyphenyl)propanoate degradation to 2-oxopent-4-enoate	14	58.58	58.35	59.74	8.72	7.70	8	B6_rs30650651
Female mice	superpathway of (Kdo)2-lipid A biosynthesis	14	58.58	58.35	59.74	8.84	7.65	8	B6_rs30650651
Female mice	superpathway of lipopolysaccharide biosynthesis	14	58.58	58.35	59.74	8.69	7.68	8	B6_rs30650651
Female mice	superpathway of methylglyoxal degradation	14	58.58	58.35	59.74	8.99	7.59	8	B6_rs30650651
Female mice	superpathway of L-arginine and L-ornithine degradation	14	58.58	58.35	59.74	9.13	7.78	8	B6_rs30650651
Female mice	superpathway of ornithine degradation	14	58.58	58.35	59.74	9.09	7.79	8	B6_rs30650651
Female mice	3-phenylpropanoate degradation	14	58.58	58.27	59.74	8.69	7.56	8	B6_rs30650651
Female mice	ppGpp biosynthesis	14	58.58	58.35	59.57	8.42	7.75	6	B6_rs30650651
Female mice	2-methylcitrate cycle II	14	58.58	58.35	59.74	8.73	7.70	8	B6_rs30650651
Female mice	ubiquinol-7 biosynthesis (prokaryotic)	14	58.58	58.35	59.74	8.88	7.69	8	B6_rs30650651
Female mice	ubiquinol-9 biosynthesis (prokaryotic)	14	58.58	58.35	59.74	8.88	7.75	8	B6_rs30650651
Female mice	ubiquinol-10 biosynthesis (prokaryotic)	14	58.58	58.35	59.74	8.88	7.86	8	B6_rs30650651
Female mice	palmitate biosynthesis II (bacteria and plants)	14	58.58	58.35	59.74	8.70	7.64	8	B6_rs30650651
Female mice	stearate biosynthesis II (bacteria and plants)	14	58.58	58.35	59.74	8.74	7.84	8	B6_rs30650651
Female mice	superpathway of phenylethylamine degradation	14	58.58	58.27	59.74	8.59	7.75	8	B6_rs30650651
Female mice	palmitoleate biosynthesis I (from (5Z)-dodec-5-enoate)	14	58.58	58.35	59.74	8.77	7.77	8	B6_rs30650651
Female mice	8-amino-7-oxononanoate biosynthesis I	14	58.58	58.35	59.74	8.69	7.71	8	B6_rs30650651
Female mice	superpathway of L-tryptophan biosynthesis	14	58.58	58.35	59.74	8.88	7.69	8	B6_rs30650651
Female mice	cinnamate and 3-hydroxycinnamate degradation to 2-oxopent-4-enoate	14	58.58	58.35	59.74	8.72	7.67	8	B6_rs30650651
Female mice	ubiquinol-8 biosynthesis (prokaryotic)	14	58.58	58.35	59.74	8.88	7.61	8	B6_rs30650651

Female mice	sulfoglycolysis	14	58.58	58.35	59.74	8.72	7.79	8	B6_rs30650651
Female mice	oleate biosynthesis IV (anaerobic)	14	58.58	58.35	59.74	8.72	7.65	8	B6_rs30650651
Female mice	3-phenylpropanoate and 3-(3-hydroxyphenyl)propanoate degradation	14	58.58	58.27	59.74	8.50	7.58	8	B6_rs30650651
Female mice	polymyxin resistance	14	58.58	58.27	60.70	7.81	7.71	11	B6_rs30650651
Female mice	superpathway of heme biosynthesis from uroporphyrinogen-III	14	58.58	58.27	59.74	8.35	7.60	8	B6_rs30650651
Female mice	phenylacetate degradation I (aerobic)	14	58.58	58.35	59.74	8.58	7.69	8	B6_rs30650651
Female mice	2-methylcitrate cycle I	14	58.58	58.35	59.74	8.67	7.81	8	B6_rs30650651
Female mice	(5Z)-dodec-5-enoate biosynthesis	14	58.58	58.27	59.74	7.99	7.67	8	B6_rs30650651
Female mice	superpathway of ubiquinol-8 biosynthesis (prokaryotic)	14	58.58	58.35	59.74	8.81	7.60	8	B6_rs30650651
Female mice	methylphosphonate degradation I	14	58.62	58.27	59.74	8.05	7.79	8	UNCHS038389
Female mice	superpathway of UDP-N-acetylglucosamine-derived O-antigen building blocks biosynthesis	14	71.11	68.33	74.96	8.32	7.70	52	UNC24208098
Female mice	superpathway of 2,3-butanediol biosynthesis	14	83.07	81.91	86.17	7.79	7.74	1	UNCHS038947
Female mice	TCA cycle V (2-oxoglutarate:ferredoxin oxidoreductase)	14	113.13	112.56	117.71	8.47	7.58	3	UNCHS039348
Female mice	TCA cycle I (prokaryotic)	14	113.13	106.23	118.61	8.10	7.59	10	UNCHS039348
Female mice	peptidoglycan biosynthesis IV (Enterococcus faecium)	14	119.22	58.08	121.34	7.87	7.61	219	UNC24876844
Female mice	superpathway of glycolysis, pyruvate dehydrogenase, TCA, and glyoxylate bypass	14	121.32	107.61	121.51	7.91	7.64	20	UNC24909376
Female mice	L-arginine biosynthesis I (via L-ornithine)	17	48.58	48.10	49.49	8.09	7.63	14	UNC27962616
Female mice	L-arginine biosynthesis II (acetyl cycle)	17	48.58	48.10	49.49	8.09	7.62	14	UNC27962616
Female mice	L-ornithine biosynthesis	17	48.58	48.10	49.49	7.83	7.55	14	UNC27962616
Female mice	L-arginine biosynthesis IV (archaebacteria)	17	48.58	48.10	49.49	8.07	7.68	14	UNC27962616
Male mice	superpathway of sulfur oxidation (Acidianus ambivalens)	1	24.62	24.36	25.57	7.81	7.70	1	UNCHS000273
Male mice	glycogen degradation I (bacterial)	1	24.85	24.24	25.57	7.92	7.80	2	UNC300940
Male mice	N10-formyl-tetrahydrofolate biosynthesis	1	24.88	24.24	25.57	8.19	7.60	2	UNC301459
Male mice	L-arginine biosynthesis I (via L-ornithine)	1	24.88	24.24	25.57	7.83	7.80	2	UNC301459
Male mice	L-arginine biosynthesis II (acetyl cycle)	1	24.88	24.24	25.57	7.84	7.52	2	UNC301459
Male mice	superpathway of L-aspartate and L-asparagine biosynthesis	1	24.88	23.84	25.57	8.03	7.75	6	UNC301459
Male mice	superpathway of N-acetylglucosamine, N-acetylmannosamine and N-acetylneuraminate degradation	1	24.88	23.52	25.53	7.72	7.65	7	UNC301459
Male mice	L-ornithine biosynthesis	1	24.88	24.24	25.57	7.82	7.69	2	UNC301459
Male mice	glycogen biosynthesis I (from ADP-D-Glucose)	1	24.88	24.17	25.55	8.72	7.53	3	UNC301459
Male mice	methylerythritol phosphate pathway I	1	24.88	24.24	25.57	7.72	7.61	2	UNC301459
Male mice	pentose phosphate pathway (non-oxidative branch)	1	24.88	24.24	25.57	8.00	7.61	2	UNC301459
Male mice	peptidoglycan biosynthesis I (meso-diaminopimelate containing)	1	24.88	24.24	25.57	7.76	7.74	2	UNC301459
Male mice	pyruvate fermentation to acetate and lactate II	1	24.88	23.84	25.54	8.15	7.75	6	UNC301459
Male mice	L-isoleucine biosynthesis III	1	24.88	24.17	25.57	7.68	7.57	3	UNC301459
Male mice	L-glutamate and L-glutamine biosynthesis	1	24.88	23.75	25.57	8.40	7.67	7	UNC301459
Male mice	chorismate biosynthesis from 3-dehydroquinate	1	24.88	24.17	25.57	7.63	7.59	3	UNC301459
Male mice	UDP-N-acetylmuramoyl-pentapeptide biosynthesis I (meso-diaminopimelate containing)	1	24.88	24.24	25.57	7.81	7.65	2	UNC301459
Male mice	L-arginine biosynthesis IV (archaebacteria)	1	24.88	24.24	25.57	7.87	7.54	2	UNC301459
Male mice	superpathway of L-serine and glycine biosynthesis I	1	24.88	23.75	25.54	8.31	7.65	7	UNC301459
Male mice	UDP-N-acetyl-D-glucosamine biosynthesis I	1	25.24	24.14	32.87	7.95	7.80	8	UNCHS000280
Male mice	dTDP-L-rhamnose biosynthesis I	1	25.24	24.24	25.64	7.60	7.58	2	UNCHS000280
Male mice	O-antigen building blocks biosynthesis (E. coli)	1	25.24	24.36	25.57	8.69	7.66	1	UNCHS000280
Male mice	acetylene degradation	2	77.36	76.29	77.37	7.79	7.67	9	UNCHS005499
Male mice	adenosine deoxyribonucleotides de novo biosynthesis II	2	77.36	76.29	78.75	8.03	7.76	12	UNCHS005499
Male mice	guanosine deoxyribonucleotides de novo biosynthesis II	2	77.36	76.29	78.75	8.03	7.69	12	UNCHS005499
Male mice	D-galacturonate degradation I	3	107.43	105.19	107.63	8.20	7.61	28	UNCHS009752

Male mice	superpathway of hexuronide and hexuronate degradation	3	107.43	105.19	107.63	8.20	7.65	28	UNC5918145
Male mice	superpathway of β-D-glucuronide and D-glucuronate degradation	3	107.43	103.69	107.63	8.44	7.72	48	UNC5918145
Male mice	peptidoglycan maturation (meso-diaminopimelate containing)	3	107.51	107.20	107.63	7.79	7.55	7	UNC5919053
Male mice	superpathway of 2,3-butanediol biosynthesis	3	122.49	109.68	123.43	7.72	7.69	52	UNC6116208
Male mice	pyruvate fermentation to acetate and lactate II	6	94.69	50.87	94.88	7.76	7.75	247	UNCHS018216
Male mice	superpathway of L-phenylalanine biosynthesis	7	29.55	28.72	29.56	7.99	7.66	26	UNC12559214
Male mice	superpathway of L-tyrosine biosynthesis	7	29.55	28.72	29.56	8.22	7.61	26	UNC12559214
Male mice	L-methionine biosynthesis III	7	136.29	135.68	137.25	8.83	7.69	3	UNCHS021798
Male mice	superpathway of L-methionine biosynthesis (transsulfuration)	7	136.58	135.68	137.26	8.53	7.81	3	UNC13940878
Male mice	L-methionine biosynthesis I	7	136.58	136.09	137.26	8.64	7.72	2	UNC13940878
Male mice	superpathway of S-adenosyl-L-methionine biosynthesis	7	136.58	136.04	137.26	8.53	7.76	2	UNC13940878
Male mice	NAD biosynthesis I (from aspartate)	11	48.03	46.98	48.14	8.28	7.69	0	UNCHS030503
Male mice	succinate fermentation to butanoate	11	54.81	52.71	79.55	7.87	7.72	434	UNCHS030727
Male mice	superpathway of L-alanine biosynthesis	11	115.20	114.64	116.66	8.07	7.73	72	UNC20458147
Male mice	allantoin degradation to glyoxylate III	12	12.00	10.55	12.60	8.76	7.72	6	UNC20635327
Male mice	superpathway of thiamin diphosphate biosynthesis II	14	104.31	55.60	104.94	7.99	7.76	248	UNC24656741
Male mice	superpathway of 2,3-butanediol biosynthesis	16	15.92	14.25	23.24	7.82	7.69	123	UNCHS041877
Male mice	glycolysis III (from glucose)	17	58.70	57.89	60.59	7.75	7.70	3	JAX00442802
Male mice	superpathway of branched amino acid biosynthesis	17	58.70	57.89	60.59	8.08	7.57	3	JAX00442802
Male mice	Calvin-Benson-Bassham cycle	17	58.70	57.89	60.59	8.02	7.70	3	JAX00442802
Male mice	L-isoleucine biosynthesis I (from threonine)	17	58.70	57.89	64.60	8.29	7.76	8	JAX00442802
Male mice	methylerythritol phosphate pathway I	17	58.70	57.87	60.59	7.84	7.61	3	JAX00442802
Male mice	pentose phosphate pathway (non-oxidative branch)	17	58.70	57.89	60.59	8.47	7.61	3	JAX00442802
Male mice	L-lysine biosynthesis III	17	58.70	57.87	64.60	7.88	7.60	8	JAX00442802
Male mice	superpathway of L-isoleucine biosynthesis I	17	58.70	57.87	60.59	7.86	7.62	3	JAX00442802
Male mice	L-isoleucine biosynthesis II	17	58.70	57.89	64.60	8.13	7.71	8	JAX00442802
Male mice	L-isoleucine biosynthesis III	17	58.70	57.89	60.59	7.90	7.57	3	JAX00442802
Male mice	L-isoleucine biosynthesis IV	17	58.70	57.89	64.66	8.23	7.72	9	JAX00442802
Male mice	5-aminoimidazole ribonucleotide biosynthesis I	17	58.70	57.89	60.59	7.87	7.68	3	JAX00442802
Male mice	5-aminoimidazole ribonucleotide biosynthesis II	17	58.70	57.89	60.59	8.11	7.70	3	JAX00442802
Male mice	inosine-5'-phosphate biosynthesis I	17	58.70	57.87	60.59	7.96	7.73	3	JAX00442802
Male mice	superpathway of 5-aminoimidazole ribonucleotide biosynthesis	17	58.70	57.89	60.59	8.11	7.55	3	JAX00442802
Male mice	peptidoglycan biosynthesis III (mycobacteria)	17	58.70	57.87	60.59	7.71	7.68	3	JAX00442802
Male mice	UDP-N-acetylmuramoyl-pentapeptide biosynthesis II (lysine-containing)	17	58.70	57.89	60.59	7.96	7.64	3	JAX00442802
Male mice	pyruvate fermentation to isobutanol (engineered)	17	58.70	57.89	64.35	8.01	7.65	8	JAX00442802
Male mice	purine ribonucleosides degradation	17	58.70	57.89	64.60	8.37	7.65	8	JAX00442802
Male mice	phosphatidylglycerol biosynthesis I (plastidic)	17	58.70	57.87	60.59	7.90	7.62	3	JAX00442802
Male mice	phosphatidylglycerol biosynthesis II (non-plastidic)	17	58.70	57.87	60.59	7.90	7.78	3	JAX00442802
Male mice	superpathway of L-threonine biosynthesis	17	58.70	57.87	64.60	7.81	7.67	8	JAX00442802
Male mice	L-valine biosynthesis	17	58.70	57.89	64.60	8.29	7.75	8	JAX00442802
Male mice	adenine and adenosine salvage III	17	58.90	57.89	60.59	8.13	7.65	3	JAX00442831
Male mice	L-rhamnose degradation I	17	60.59	58.06	64.85	8.96	7.73	10	UNC28131597
Male mice	superpathway of fucose and rhamnose degradation	17	60.59	57.89	64.87	8.63	7.63	10	UNCHS044633
Entire mice	mannan degradation	1	57.68	56.29	59.25	8.77	7.63	23	UNC733960
Entire mice	queuosine biosynthesis	1	112.36	108.21	115.53	7.78	7.53	3	UNC1422808
Entire mice	superpathway of L-phenylalanine biosynthesis	2	117.39	114.36	117.99	7.87	7.56	7	JAX00500783
Entire mice	superpathway of L-tyrosine biosynthesis	2	117.39	114.36	117.99	7.71	7.57	7	JAX00500783
Entire mice	superpathway of purine nucleotides de novo biosynthesis II	4	58.32	57.89	61.86	7.77	7.54	22	UNCHS011408
Entire mice	superpathway of guanosine nucleotides de novo biosynthesis II	4	58.32	57.89	59.85	7.96	7.54	19	UNCHS011408
Entire mice	pyrimidine deoxyribonucleotides de novo biosynthesis I	4	58.32	57.89	61.86	7.94	7.65	22	UNCHS011408

Entire mice	superpathway of pyrimidine ribonucleosides salvage	4	58.32	57.89	59.85	8.09	7.58	19	UNCHS011408
Entire mice	pyrimidine deoxyribonucleotide phosphorylation	4	58.32	57.89	59.85	7.92	7.47	19	UNCHS011408
Entire mice	superpathway of pyrimidine deoxyribonucleoside salvage	4	58.32	57.89	61.86	8.09	7.61	22	UNCHS011408
Entire mice	superpathway of guanosine nucleotides de novo biosynthesis I	4	58.32	57.89	59.85	7.99	7.50	19	UNCHS011408
Entire mice	superpathway of purine nucleotides de novo biosynthesis I	4	58.32	57.89	61.86	7.94	7.72	22	UNCHS011408
Entire mice	superpathway of pyrimidine deoxyribonucleotides de novo biosynthesis (E. coli)	4	58.32	57.89	61.86	7.95	7.53	22	UNCHS011408
Entire mice	superpathway of L-methionine biosynthesis (by sulfhydrylation)	4	58.32	57.06	58.80	7.67	7.65	9	UNC7330906
Entire mice	superpathway of pyrimidine deoxyribonucleotides de novo biosynthesis	4	58.32	57.89	58.80	8.47	7.68	7	UNC7330906
Entire mice	sulfate reduction I (assimilatory)	4	58.32	57.19	58.51	8.29	7.54	6	UNCHS011409
Entire mice	superpathway of sulfate assimilation and cysteine biosynthesis	4	58.32	57.19	58.52	8.11	7.63	6	UNCHS011409
Entire mice	L-methionine biosynthesis I	6	88.03	87.79	88.78	7.72	7.46	15	UNC11551488
Entire mice	L-methionine biosynthesis III	6	88.15	87.79	88.78	7.67	7.59	15	UNC11552932
Entire mice	purine ribonucleosides degradation	9	35.35	35.01	36.88	8.13	7.65	15	UNC16152533
Entire mice	adenine and adenosine salvage III	9	35.38	34.97	36.88	7.88	7.64	15	UNC16152834
Entire mice	adenosine deoxyribonucleotides de novo biosynthesis II	9	122.01	121.68	122.02	7.72	7.51	9	UNCHS027368
Entire mice	guanosine deoxyribonucleotides de novo biosynthesis II	9	122.01	121.68	122.02	7.72	7.67	9	UNCHS027368
Entire mice	superpathway of glucose and xylose degradation	10	81.60	79.60	84.05	7.76	7.59	128	UNCHS028871
Entire mice	L-ornithine biosynthesis	11	48.03	46.98	48.14	7.73	7.53	0	UNCHS030503
Entire mice	allantoin degradation to glyoxylate III	12	12.18	10.55	12.60	7.71	7.68	6	UNC20637838
Entire mice	TCA cycle I (prokaryotic)	14	107.66	106.04	110.05	10.76	7.66	10	UNC24909376
Entire mice	TCA cycle V (2-oxoglutarate:ferredoxin oxidoreductase)	14	107.66	105.96	112.77	10.52	7.65	3	UNCHS039291
Entire mice	pyridoxal 5'-phosphate biosynthesis I	14	108.20	105.75	119.30	7.69	7.61	17	UNCHS039304
Entire mice	L-ornithine biosynthesis	14	118.04	117.69	119.36	7.65	7.53	10	UNC24855486
Entire mice	L-glutamate and L-glutamine biosynthesis	14	118.06	117.75	119.36	7.55	7.50	10	UNCHS039387
Entire mice	L-1,2-propanediol degradation	14	118.62	117.85	119.36	7.68	7.61	10	UNC24866496
Entire mice	superpathway of N-acetylglucosamine, N-acetylmannosamine and N-acetylneuraminate degradation	14	118.63	117.69	119.36	7.69	7.53	10	UNC24866697
Entire mice	heme biosynthesis I (aerobic)	14	119.98	112.76	121.56	8.37	7.56	19	UNCHS039412
Entire mice	superpathay of heme biosynthesis from glutamate	14	119.98	112.76	121.57	8.26	7.73	19	UNCHS039412
Entire mice	enterobactin biosynthesis	14	121.32	112.76	121.51	8.87	7.52	18	UNC24909376
Entire mice	superpathway of glycolysis, pyruvate dehydrogenase, TCA, and glyoxylate bypass	14	121.32	117.77	121.51	9.44	7.57	15	UNC24909376
Entire mice	glyoxylate cycle	14	121.32	117.85	121.51	9.38	7.57	15	UNC24909376
Entire mice	superpathway of hexitol degradation (bacteria)	14	121.32	106.05	121.58	8.04	7.53	22	UNC24909376
Entire mice	superpathway of methylglyoxal degradation	14	121.32	37.04	121.58	7.63	7.56	405	UNC24909376
Entire mice	TCA cycle IV (2-oxoglutarate decarboxylase)	14	121.32	112.76	121.51	8.71	7.61	18	UNC24909376
Entire mice	superpathway of glyoxylate bypass and TCA	14	121.32	118.74	121.51	9.37	7.74	1	UNCHS039291
Entire mice	superpathway of L-arginine, putrescine, and 4-aminobutanoate degradation	14	121.32	37.00	121.58	7.55	7.54	2	UNC28185909
Entire mice	L-arginine degradation II (AST pathway)	14	121.46	37.04	121.58	7.64	7.64	405	UNC24911982
Entire mice	enterobacterial common antigen biosynthesis	14	121.46	37.04	121.58	7.68	7.63	405	UNC24911982
Entire mice	3-phenylpropanoate degradation	14	121.46	37.04	121.58	7.61	7.59	405	UNC24911982
Entire mice	catechol degradation I (meta-cleavage pathway)	14	121.46	36.85	121.58	7.95	7.50	407	UNC24911982
Entire mice	ubiquinol-7 biosynthesis (prokaryotic)	14	121.46	37.04	121.58	7.64	7.49	405	UNC24911982
Entire mice	ubiquinol-9 biosynthesis (prokaryotic)	14	121.46	37.04	121.58	7.64	7.51	405	UNC24911982
Entire mice	palmitate biosynthesis II (bacteria and plants)	14	121.46	37.04	121.58	7.65	7.56	405	UNC24911982
Entire mice	palmitoleate biosynthesis I (from (5Z)-dodec-5-enoate)	14	121.46	37.04	121.58	7.58	7.56	405	UNC24911982
Entire mice	superpathway of L-tryptophan biosynthesis	14	121.46	37.04	121.58	7.69	7.61	405	UNC24911982
Entire mice	superpathway of heme biosynthesis from uroporphyrinogen-III	14	121.46	37.00	121.58	7.72	7.67	406	UNC24911982
Entire mice	tRNA processing	14	121.46	120.60	121.58	8.51	7.57	4	UNC24911982

Entire mice	2-methylcitrate cycle I	14	121.46	37.04	121.58	7.69	7.59	405	UNC24911982
Entire mice	superpathway of ubiquinol-8 biosynthesis (prokaryotic)	14	121.46	37.04	121.58	7.68	7.67	405	UNC24911982
Entire mice	formaldehyde oxidation I	15	13.56	12.20	18.96	8.01	7.65	8	JAX00394051
Entire mice	superpathway of purine nucleotides de novo biosynthesis II	15	28.36	27.41	28.51	7.71	7.54	5	UNCHS039854
Entire mice	superpathway of guanosine nucleotides de novo biosynthesis II	15	28.36	27.41	28.51	7.65	7.54	5	UNCHS039854
Entire mice	pyrimidine deoxyribonucleotide phosphorylation	15	28.36	27.41	28.51	7.72	7.47	5	UNCHS039854
Entire mice	superpathway of pyrimidine deoxyribonucleoside salvage	15	28.36	27.41	28.51	7.70	7.61	5	UNCHS039854
Entire mice	superpathway of guanosine nucleotides de novo biosynthesis I	15	28.36	27.41	28.51	7.68	7.50	5	UNCHS039854
Entire mice	superpathway of pyrimidine deoxyribonucleotides de novo biosynthesis (E. coli)	15	28.36	27.41	28.51	7.88	7.53	5	UNCHS039854
Entire mice	superpathway of L-alanine biosynthesis	16	14.25	13.50	17.31	7.82	7.57	39	UNCHS041836
Entire mice	biotin biosynthesis II	16	15.92	14.25	17.39	7.70	7.51	29	UNCHS041877
Entire mice	N10-formyl-tetrahydrofolate biosynthesis	17	64.60	64.52	64.91	7.89	7.61	2	UNC28185137
Entire mice	chorismate biosynthesis I	17	64.60	64.48	64.91	7.72	7.63	2	UNC28185137
Entire mice	superpathway of branched amino acid biosynthesis	17	64.60	64.48	64.91	7.84	7.58	2	UNC28185137
Entire mice	superpathway of aromatic amino acid biosynthesis	17	64.60	64.48	64.91	7.77	7.60	2	UNC28185137
Entire mice	glycogen biosynthesis I (from ADP-D-Glucose)	17	64.60	64.56	64.91	8.16	7.56	2	UNC28185137
Entire mice	L-histidine biosynthesis	17	64.60	64.48	64.91	7.66	7.65	2	UNC28185137
Entire mice	L-isoleucine biosynthesis I (from threonine)	17	64.60	64.25	64.91	7.66	7.61	3	UNC28185137
Entire mice	methylerythritol phosphate pathway I	17	64.60	64.52	64.91	8.15	7.59	2	UNC28185137
Entire mice	pentose phosphate pathway (non-oxidative branch)	17	64.60	64.25	64.91	8.25	7.61	3	UNC28185137
Entire mice	L-lysine biosynthesis VI	17	64.60	64.48	64.91	8.07	7.57	2	UNC28185137
Entire mice	L-isoleucine biosynthesis II	17	64.60	64.48	64.91	7.83	7.55	2	UNC28185137
Entire mice	L-isoleucine biosynthesis III	17	64.60	64.52	64.91	7.74	7.61	2	UNC28185137
Entire mice	L-isoleucine biosynthesis IV	17	64.60	64.48	64.87	8.20	7.53	2	UNC28185137
Entire mice	superpathway of geranylgeranyl diphosphate biosynthesis II (via MEP)	17	64.60	64.50	64.91	7.93	7.76	2	UNC28185137
Entire mice	L-glutamate and L-glutamine biosynthesis	17	64.60	64.56	64.84	8.30	7.50	2	UNC28185137
Entire mice	5-aminoimidazole ribonucleotide biosynthesis I	17	64.60	64.48	64.91	7.67	7.56	2	UNC28185137
Entire mice	5-aminoimidazole ribonucleotide biosynthesis II	17	64.60	64.48	64.91	7.91	7.66	2	UNC28185137
Entire mice	inosine-5'-phosphate biosynthesis I	17	64.60	64.48	64.91	7.72	7.46	2	UNC28185137
Entire mice	superpathway of 5-aminoimidazole ribonucleotide biosynthesis	17	64.60	64.48	64.91	7.91	7.48	2	UNC28185137
Entire mice	starch degradation V	17	64.60	64.52	64.94	7.95	7.66	2	UNC28185137
Entire mice	methylerythritol phosphate pathway II	17	64.60	64.52	64.91	8.15	7.75	2	UNC28185137
Entire mice	L-valine biosynthesis	17	64.60	64.25	64.91	7.66	7.62	3	UNC28185137
Entire mice	purine ribonucleosides degradation	17	64.60	64.25	64.87	7.76	7.65	3	UNC28185137
Entire mice	chorismate biosynthesis from 3-dehydroquinate	17	64.60	64.48	64.91	7.60	7.53	2	UNCHS044716
Entire mice	L-arginine biosynthesis IV (archaebacteria)	17	64.66	64.56	64.91	7.84	7.51	2	UNC28185909
Entire mice	NAD salvage pathway I	17	64.66	63.60	64.91	8.47	7.67	5	UNC28185909
Entire mice	L-arginine biosynthesis II (acetyl cycle)	17	64.69	64.56	64.91	7.97	7.67	2	UNC28186392
Entire mice	L-ornithine biosynthesis	17	64.72	64.56	64.87	8.18	7.53	2	UNC28186937

Significance was considered only when QTL results were $P<0.05$.

[a]95% Bayesian credible interval was calculated as implemented by the bayesint function in R/QTL2.

5.5.5. Co-mapped QTLs identify a novel interaction between aortic lesion area and *Lactococcus* abundance

We searched for QTLs that were associated with both cardiometabolic traits and gut microbial abundance in all mice, female mice, or male mice, respectively. We identified only one case of co-mapped loci for aortic lesion size and *Lactococcus* abundance on chromosome 10 at 26 Mbp ($P<0.1$, **Figure 5.7A and 5.7B**). Furthermore, allelic variation (C/C or A/C) at rs36544414 peak SNP in the aortic lesion area QTL was associated with the lesion size and *Lactococcus* abundance in female mice (**Figure 5.7C**). This overlapped QTL suggests that genetic variants affecting atherosclerosis influence the abundance of the *Lactococcus*, or genetic variation in *Lactococcus* abundance affects atherosclerosis. h^2 of aortic lesion area (0.64, in Chapter 4) and *Lactococcus* abundance (0.87, 9th out of 40 genera) were also relatively high compared to other traits in females. Previous studies have shown the effect of host genetics on atherosclerosis and *Lactococcus* abundance. For example, *Lactococcus* is highly heritable in both mouse and human genetic studies (Davenport et al., 2014; Goodrich et al., 2016; Org et al., 2015), and several studies have found differences in atherosclerosis by host genotype (Bennett et al., 2015; Gaeta, 2007; Smallwood et al., 2014; Watkins and Farrall, 2006). Furthermore, *Lactococcus* abundance was identified as a significantly abundant genus level in the high aortic lesion area group by ANCOM (**Figure 5.4A**) and showed a significant positive correlation with aortic lesion area, plasma total cholesterol, and hepatic total cholesterol in both sexes (**Figure 5.8A-5.8C**). Notably, the aortic lesion area and *Lactococcus* abundance also shared the similar allele effects pattern, where the alleles from NZO strain with the highest level of traits related to obesity and type 2 diabetes (Joost and Schurmann, 2014) showed positive associations, and the alleles from the CAST and PWK strains, wild-derived strains resistant to obesity (Mathes et al., 2011; Reed et

al., 2007; Svenson et al., 2007), showed a negative association with levels of the aortic lesion area and *Lactococcus* abundance at the peak SNP (rs36544414) (**Figure 5.8D and 5.8E**).

In addition, we investigated whether functional profiling associated with *Lactococcus* co-localized with this loci, and six EC gene families were identified ($p<0.1$) (**Figure 5.9A and Table 5.11**). Among them, the QTLs of "Raffinose--raffinose alpha-galactosyltransferase", "Endo-1,4-beta-xylanase" and "Glutamyl aminopeptidase" had the same peak SNPs as *Lactococcus* abundance QTL (**Table 5.11**). All six EC gene families were not only associated with allelic variation (C/C or A/C) at rs36544414 peak SNP in the aortic lesion area QTL (**Figure 5.9B**), but also showed strong positive correlations with three atherosclerotic traits (**Figures 5.10 and Table 5.11**). Furthermore, all of these enzymes are known to be produced by *Lactococcus lactis* subsp. *lactis* (Bolotin et al., 2001; Kelleher et al., 2017; McNulty et al., 2011; Passerini et al., 2013; Siezen et al., 2008). Collectively, the QTL overlap between the aortic lesion area, *Lactococcus* abundance, and EC gene families provides profound evidence that these traits are related and they are responding to the common genetic driver, with the similar allele effects pattern and significant association between traits.

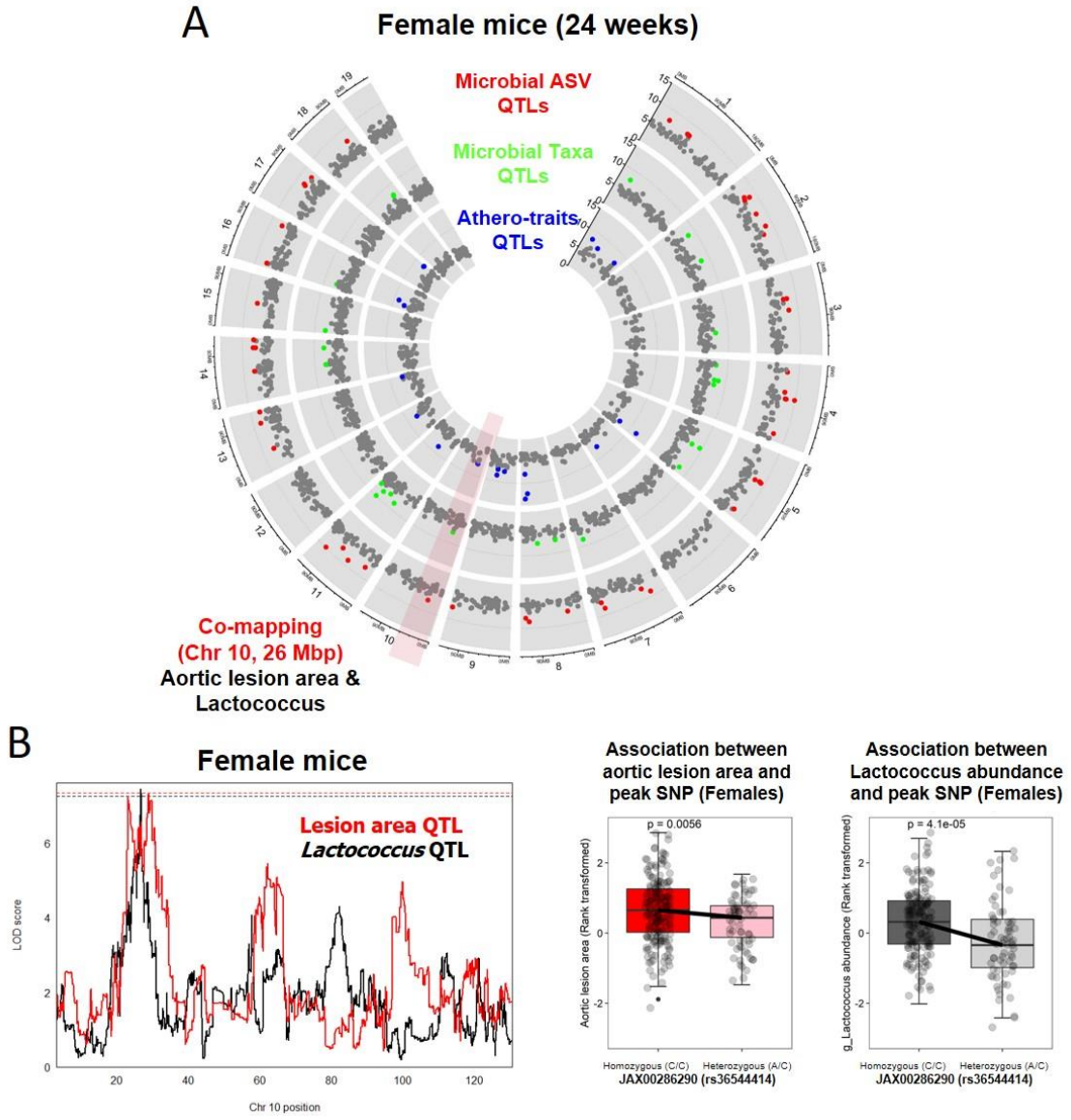


Figure 5.7. Identification of microbial genus levels associated with cardiometabolic traits in DO-F1 mice.

(A) Genetic architecture of QTL for cardio-metabolic traits, microbial taxa abundance, and microbial amplicon sequence variants (ASVs). The outer layer shows the chromosome location. LOD range is shown for each track. Each dot represents a QTL on each chromosome of the mouse genome for a given trait. Grey dots denote QTLs with LOD < 7. Co-mapped loci between aortic lesion area QTL and *Lactococcus* QTL is denoted. (B) LOD profiles on Chromosome 10 highlighting a locus associated with the aortic lesion area and abundance of *Lactococcus* genus level in female mice. Each dashed line (red for aortic lesion area QTL and black for *Lactococcus* abundance QTL) represents suggestive thresholds (P< 0.1). Association between a peak SNP (rs36544414) in aortic lesion area QTL and *Lactococcus* abundance QTL. Heterozygous genotype (A/C) is derived from CAST strain.

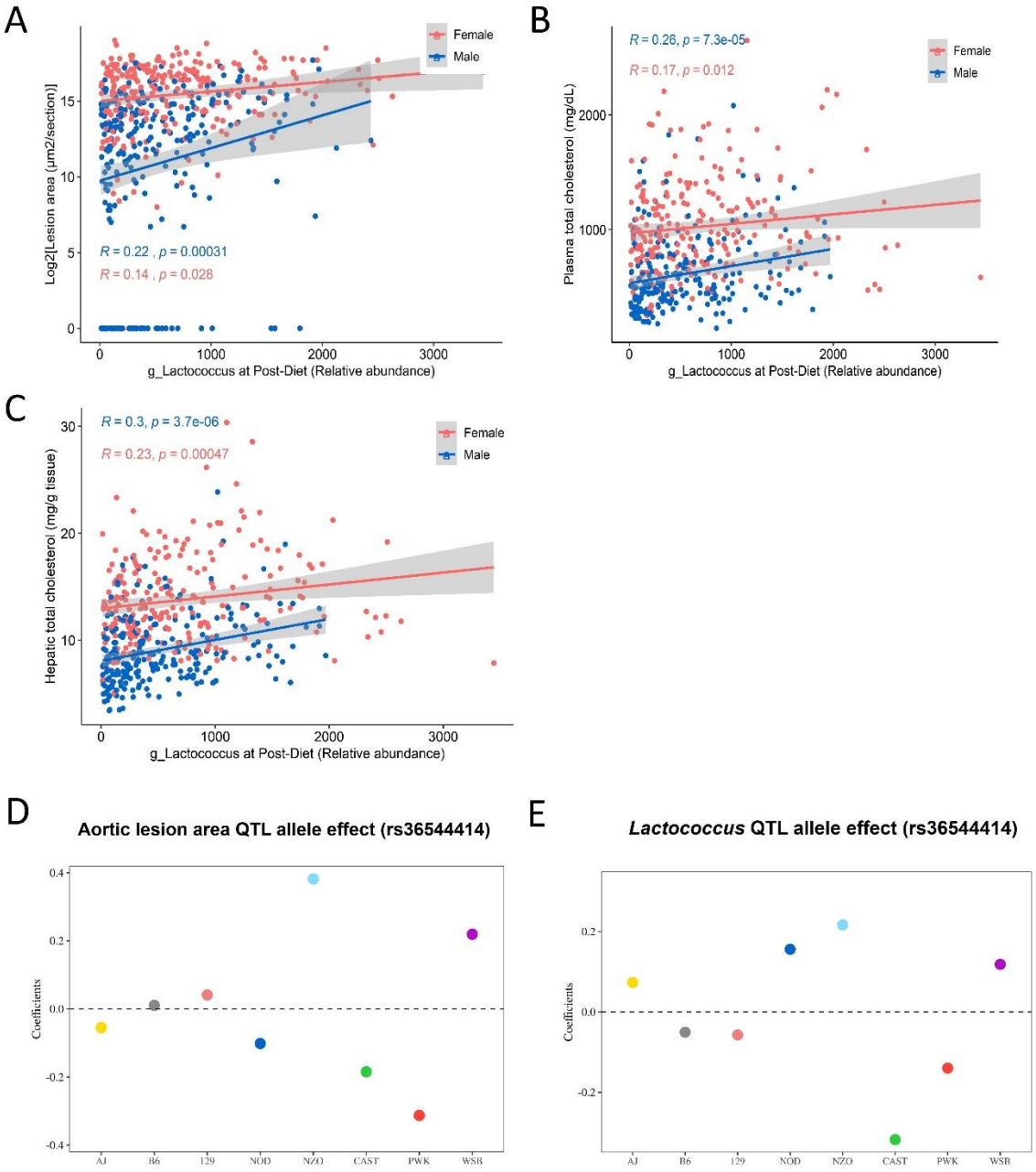


Figure 5.8. Co-mapping of aortic lesion area and *Lactococcus* QTL on chromosome 10 in DO-F1 female mice.

(A) *Lactococcus* abundance is significantly associated with aortic lesion area in females (R = 0.14, p = 0.026) and males (R = 0.22, p = 0.00031). (B) *Lactococcus* abundance is significantly associated with plasma total cholesterol in females (R = 0.17, p = 0.012) and males (R = 0.26, p = 7.3×10^{-5}). (C) *Lactococcus* abundance is significantly associated with hepatic total cholesterol in females (R = 0.23, p = 0.000047) and males (R = 0.26, p = 3.7×10^{-6}). Estimated founder strain levels of aortic lesion area (A) and *Lactococcus* abundance (B) inferred in the DO-F1 population from the founder strain coefficients observed at the corresponding QTL on chr 10.

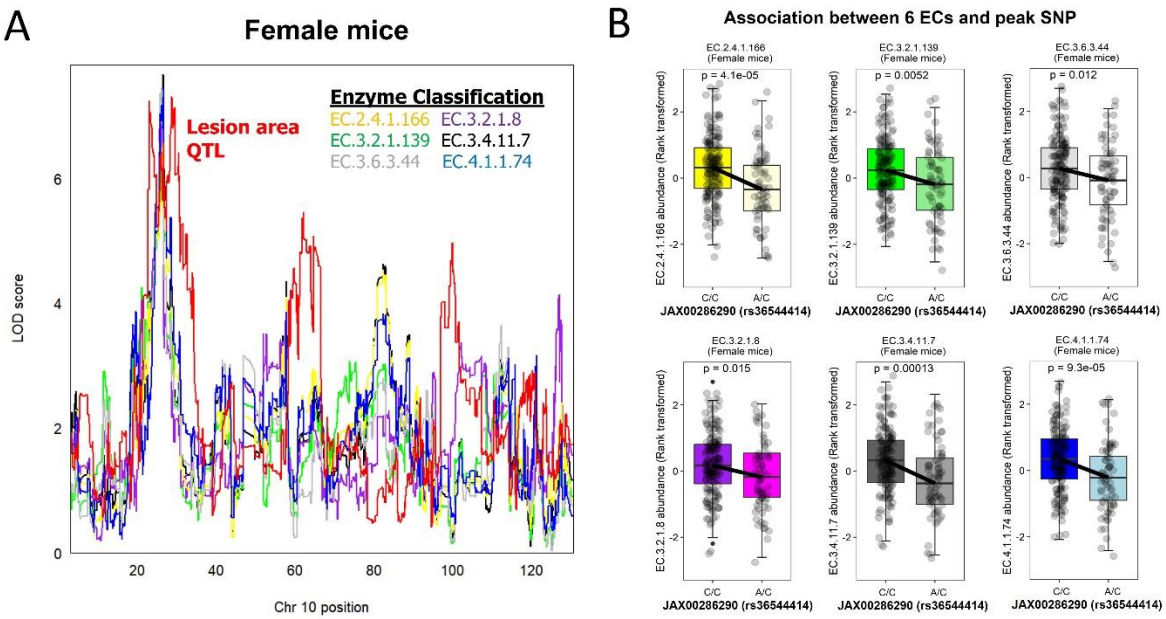


Figure 5.9. Co-mapping of aortic lesion area and six EC gene families associated with *Lactococcus* QTL on chromosome 10 in DO-F1 female mice.
(A) LOD profiles on Chromosome 10 highlighting a locus associated with the aortic lesion area and abundance of 6 EC gene families in female mice. Each dashed line represents suggestive thresholds (P<0.1); red for aortic lesion area QTL, yellow for Raffinose--raffinose alpha-galactosyltransferase (EC.2.4.1.166) QTL, green for Alpha-glucuronidase (EC.3.2.1.139) QTL, purple for Endo-1,4-beta-xylanase (EC.3.2.1.8) QTL, black for Glutamyl aminopeptidase (EC.3.4.11.7) QTL, grey for Xenobiotic-transporting ATPase (EC.3.6.3.44), and blue for Indolepyruvate decarboxylase (EC.4.1.1.74). (B) Association between a peak SNP (rs36544414) in aortic lesion area QTL and 6 EC gene families QTLs. Heterozygous genotype (A/C) is derived from CAST strain.

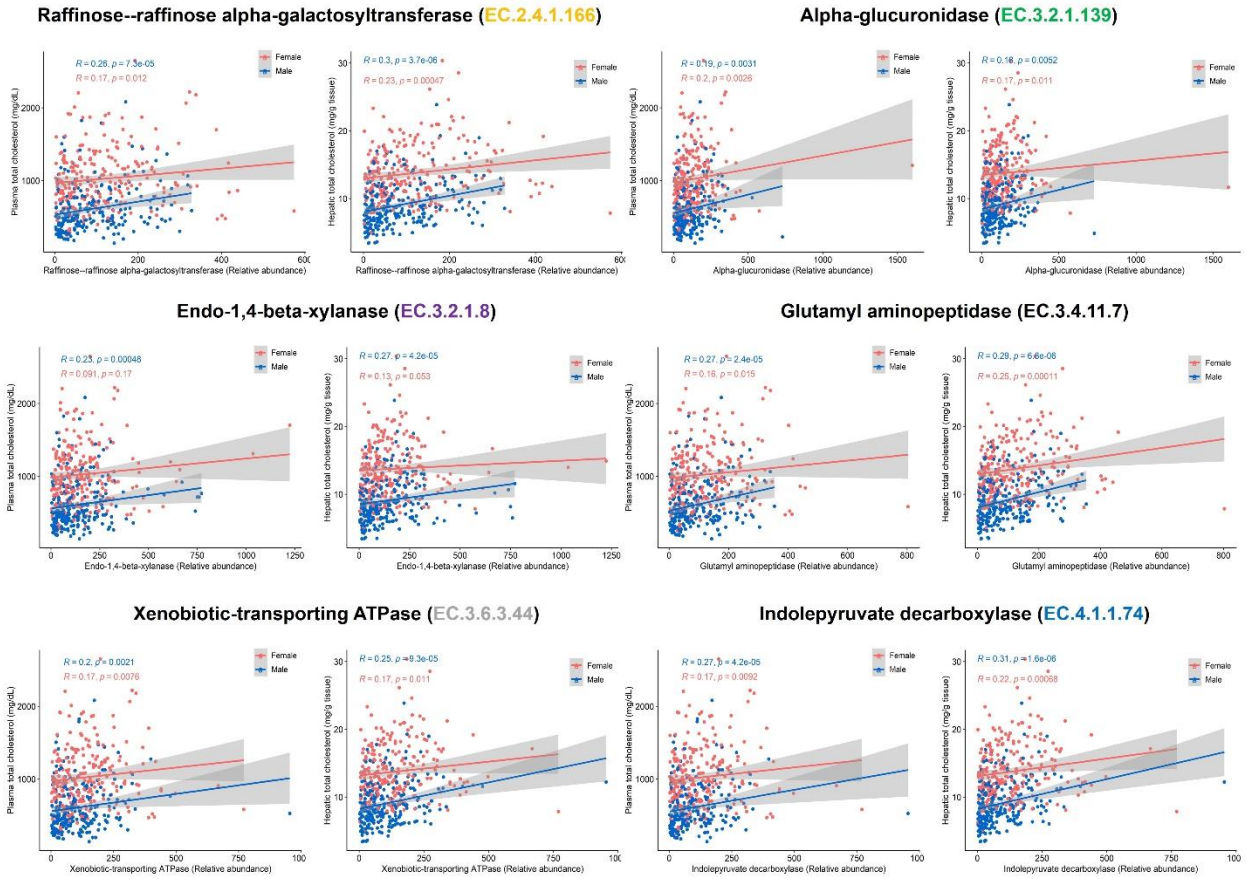


Figure 5.10. Spearman correlation of six EC gene families with plasma total cholesterol or liver total cholesterol.

6 EC gene families are significantly associated with total cholesterol in plasma and liver. Yellow for Raffinose--raffinose alpha-galactosyltransferase (EC.2.4.1.166) QTL, green for Alpha-glucuronidase (EC.3.2.1.139) QTL, purple for Endo-1,4-beta-xylanase (EC.3.2.1.8) QTL, black for Glutamyl aminopeptidase (EC.3.4.11.7) QTL, grey for Xenobiotic-transporting ATPase (EC.3.6.3.44), and blue for Indolepyruvate decarboxylase (EC.4.1.1.74).

5.5.6. Hepatic gene networks are highly correlated with atherosclerosis and gut microbiota

We have identified sexual dimorphisms in liver transcriptome in Chapter 4, suggesting that co-expressed gene clusters may function together to elicit sex-specific responses in clinically relevant traits and gut microbiota. To identify modules of co-expressed genes, we performed weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008). We constructed hepatic co-expression modules in each sex to best estimate associations between modules and cardiometabolic traits or gut microbiota, and performed clustering by the merged dynamic tree cut approach. We identified 9 modules in females and 7 modules in males each assigned an arbitrary color, and the number of hepatic genes contained in each module ranged from 45 to 6,457 in females and 130 to 4,351 in males (**Table 5.11)** (Langfelder and Horvath, 2008).

Highly correlated transcripts are often associated with specific biological pathways (Bennett et al., 2015; Keller et al., 2018; Langfelder and Horvath, 2008). Next, we performed a gene set enrichment analysis on the modules identified in each sex to investigate whether the genes in the module were associated with shared functional annotations. We referred modules with color key identifiers and the modules in each sex were significantly enriched with one or more functional annotations such as GO terms and KEGG pathways (**Table 5.11**). For example, the lightyellow module in females and the black module in males were enriched for sterol and cholesterol biosynthetic pathways. The darkgreen module and darkturquoise module in females and turquoise module in males were associated with immune responses such as type 1 interferon, antigen presentation, and neutrophil-mediated immunity, respectively.

Next, we determined whether the modules were correlated with the cardiometabolic traits or gut microbiota by calculating the module eigengene (ME) to investigate the physiological

aspects of these modules. We confirmed that the ME in purple module enriched for protein ubiquitination in females (Spearman's rho = -0.27, adjusted p-value ≤ 0.05) and the ME in black module enriched for steroid biosynthesis in males (Spearman's rho = -0.43, adjusted p-value ≤ 9.3×10^{-5}) had a negative correlation with aortic lesion area (**Figures 5.11A and 5.12A**), and we further identified that these modules are highly correlated with gut microbiota, such as microbial diversity (**Figures 5.11B and 5.12B**), genera (**Figures 5.11C and 5.12C**), MetaCyc pathways (**Table 5.12**), and EC gene families (**Table 5.12**). In addition, both modules showed a generally negative correlation with taxa, MetaCyc pathways, and EC gene families, which were commonly abundant in the group with high atherosclerotic traits (**Figures 5.11C and 5.12C, Table 5.12**). These analyses demonstrate the ability of co-regulated transcript networks to interrogate sex-biased gene function and probe the interaction of the networks that correlate with cardiometabolic traits and gut microbiota.

Table 5.11. Top 10 Gene Ontology and KEGG pathways results for nine hepatic coexpression gene modules in females and seven hepatic coexpression gene modules in males and in liver in each sex. n = 162 (85 females and 77 males).

Sex	Module (# of genes)	Library	#Terms	Top10 Terms	Overlap (gene count)	Adjusted. P.value (<0.05)
Female	purple (2,905)	GO_Biological_Process_2018	150	protein ubiquitination (GO:0016567)	143/506	5.E-13
Female		GO_Biological_Process_2018		protein modification by small protein conjugation (GO:0032446)	118/398	3.E-12
Female		GO_Biological_Process_2018		ubiquitin-dependent protein catabolic process (GO:0006511)	105/341	4.E-12
Female		GO_Biological_Process_2018		protein modification by small protein removal (GO:0070646)	85/261	4.E-11
Female		GO_Biological_Process_2018		protein deubiquitination (GO:0016579)	84/257	4.E-11
Female		GO_Biological_Process_2018		cellular protein modification process (GO:0006464)	218/1001	2.E-08
Female		GO_Biological_Process_2018		protein polyubiquitination (GO:0000209)	81/283	2.E-07
Female		GO_Biological_Process_2018		proteasome-mediated ubiquitin-dependent protein catabolic process (GO:0043161)	81/291	6.E-07
Female		GO_Biological_Process_2018		modification-dependent protein catabolic process (GO:0019941)	47/141	4.E-06
Female		GO_Biological_Process_2018		regulation of transcription from RNA polymerase II promoter (GO:0006357)	285/1478	1.E-05
Female		KEGG_2019_Mouse	23	Ubiquitin mediated proteolysis	44/138	3.E-05
Female		KEGG_2019_Mouse		Protein processing in endoplasmic reticulum	47/163	2.E-04
Female		KEGG_2019_Mouse		Thermogenesis	59/231	4.E-04
Female		KEGG_2019_Mouse		Autophagy	38/130	6.E-04
Female		KEGG_2019_Mouse		Phosphatidylinositol signaling system	30/98	2.E-03
Female		KEGG_2019_Mouse		Huntington disease	48/192	2.E-03
Female		KEGG_2019_Mouse		Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	12/25	2.E-03
Female		KEGG_2019_Mouse		Alzheimer disease	44/175	3.E-03
Female		KEGG_2019_Mouse		Oxidative phosphorylation	35/134	7.E-03
Female		KEGG_2019_Mouse		mRNA surveillance pathway	27/96	9.E-03
Female	darkgreen (49)	GO_Biological_Process_2018	64	cellular response to type I interferon (GO:0071357)	14/65	3.E-22
Female		GO_Biological_Process_2018		type I interferon signaling pathway (GO:0060337)	14/65	3.E-22
Female		GO_Biological_Process_2018		negative regulation of viral life cycle (GO:1903901)	8/61	2.E-10
Female		GO_Biological_Process_2018		regulation of viral genome replication (GO:0045069)	8/63	2.E-10
Female		GO_Biological_Process_2018		cytokine-mediated signaling pathway (GO:0019221)	15/633	9.E-10
Female		GO_Biological_Process_2018		regulation of type I interferon production (GO:0032479)	8/85	1.E-09
Female		GO_Biological_Process_2018		negative regulation of viral genome replication (GO:0045071)	7/50	1.E-09
Female		GO_Biological_Process_2018		negative regulation of type I interferon production (GO:0032480)	6/44	4.E-08
Female		GO_Biological_Process_2018		response to cytokine (GO:0034097)	8/138	5.E-08
Female		GO_Biological_Process_2018		response to interferon-alpha (GO:0035455)	4/17	2.E-06
Female		KEGG_2019_Mouse	13	Influenza A	11/168	6.E-12
Female		KEGG_2019_Mouse		Measles	10/144	2.E-11
Female		KEGG_2019_Mouse		Hepatitis C	10/160	4.E-11
Female		KEGG_2019_Mouse		Epstein-Barr virus infection	9/229	3.E-08
Female		KEGG_2019_Mouse		Herpes simplex virus 1 infection	10/433	4.E-07
Female		KEGG_2019_Mouse		RIG-I-like receptor signaling pathway	5/68	3.E-06
Female		KEGG_2019_Mouse		NOD-like receptor signaling pathway	6/205	4.E-05
Female		KEGG_2019_Mouse		Human papillomavirus infection	6/360	8.E-04
Female		KEGG_2019_Mouse		Cytosolic DNA-sensing pathway	3/61	1.E-03

Female		KEGG_2019_Mouse	Hepatitis B	4/163	2.E-03
Female	darkgrey (45)	GO_Biological_Process_2018	negative regulation of signal transduction (GO:0009968)	6/283	7.E-03
Female		GO_Biological_Process_2018	regulation of cell migration involved in sprouting angiogenesis (GO:0090049)	3/30	7.E-03
Female		GO_Biological_Process_2018	vasculogenesis (GO:0001570)	3/32	7.E-03
Female		GO_Biological_Process_2018	glomerulus vasculature development (GO:0072012)	2/6	7.E-03
Female		GO_Biological_Process_2018	regulation of kinase activity (GO:0043549)	4/101	7.E-03
Female		GO_Biological_Process_2018	positive regulation of angiogenesis (GO:0045766)	4/103	7.E-03
Female		GO_Biological_Process_2018	positive regulation of vasculature development (GO:1904018)	4/104	7.E-03
Female		GO_Biological_Process_2018	branching morphogenesis of an epithelial tube (GO:0048754)	3/43	8.E-03
Female		GO_Biological_Process_2018	sprouting angiogenesis (GO:0002040)	3/44	8.E-03
Female		GO_Biological_Process_2018	pericardium development (GO:0060039)	2/9	1.E-02
Female		KEGG_2019_Mouse	Cell adhesion molecules (CAMs)	5/170	2.E-03
Female		KEGG_2019_Mouse	Leukocyte transendothelial migration	4/115	3.E-03
Female	darkred (6,457)	GO_Biological_Process_2018	ncRNA processing (GO:0034470)	153/227	8.E-25
Female		GO_Biological_Process_2018	viral process (GO:0016032)	149/220	9.E-25
Female		GO_Biological_Process_2018	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile (GO:0000377)	156/236	2.E-24
Female		GO_Biological_Process_2018	rRNA processing (GO:0006364)	139/202	2.E-24
Female		GO_Biological_Process_2018	mRNA splicing, via spliceosome (GO:0000398)	165/261	7.E-23
Female		GO_Biological_Process_2018	mRNA processing (GO:0006397)	175/283	8.E-23
Female		GO_Biological_Process_2018	neutrophil mediated immunity (GO:0002446)	263/487	5.E-22
Female		GO_Biological_Process_2018	rRNA metabolic process (GO:0016072)	134/200	6.E-22
Female		GO_Biological_Process_2018	neutrophil degranulation (GO:0043312)	259/479	7.E-22
Female		GO_Biological_Process_2018	ribosome biogenesis (GO:0042254)	146/226	1.E-21
Female		KEGG_2019_Mouse	Spliceosome	97/132	2.E-20
Female		KEGG_2019_Mouse	Lysosome	85/124	6.E-15
Female		KEGG_2019_Mouse	Fc gamma R-mediated phagocytosis	60/87	1.E-10
Female		KEGG_2019_Mouse	RNA transport	95/167	1.E-09
Female		KEGG_2019_Mouse	Apoptosis	83/141	2.E-09
Female		KEGG_2019_Mouse	Endocytosis	137/269	2.E-09
Female		KEGG_2019_Mouse	Pathways in cancer	240/535	3.E-09
Female		KEGG_2019_Mouse	Cell cycle	73/123	9.E-09
Female		KEGG_2019_Mouse	NF-kappa B signaling pathway	63/102	1.E-08
Female		KEGG_2019_Mouse	DNA replication	29/35	2.E-08
Female	darkturquoise (47)	GO_Biological_Process_2018	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent (GO:0002479)	6/75	2.E-06
Female		GO_Biological_Process_2018	antigen processing and presentation of exogenous peptide antigen via MHC class I (GO:0042590)	6/78	2.E-06
Female		GO_Biological_Process_2018	regulation of cellular amino acid metabolic process (GO:0006521)	5/51	6.E-06
Female		GO_Biological_Process_2018	regulation of cellular amine metabolic process (GO:0033238)	5/51	6.E-06
Female		GO_Biological_Process_2018	regulation of cellular ketone metabolic process (GO:0010565)	5/61	1.E-05
Female		GO_Biological_Process_2018	negative regulation of G2/M transition of mitotic cell cycle (GO:0010972)	5/62	1.E-05
Female		GO_Biological_Process_2018	tumor necrosis factor-mediated signaling pathway (GO:0033209)	6/123	1.E-05
Female		GO_Biological_Process_2018	negative regulation of cell cycle G2/M phase transition (GO:1902750)	5/69	1.E-05
Female		GO_Biological_Process_2018	negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle (GO:0051436)	5/71	1.E-05
Female		GO_Biological_Process_2018	regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle (GO:0051439)	5/72	1.E-05

Female		KEGG_2019_Mouse	12	Proteasome	5/46	2.E-06
Female		KEGG_2019_Mouse		NOD-like receptor signaling pathway	6/205	9.E-05
Female		KEGG_2019_Mouse		Antigen processing and presentation	4/90	5.E-04
Female		KEGG_2019_Mouse		Toll-like receptor signaling pathway	3/99	1.E-02
Female		KEGG_2019_Mouse		Epstein-Barr virus infection	4/229	1.E-02
Female		KEGG_2019_Mouse		Toxoplasmosis	3/108	1.E-02
Female		KEGG_2019_Mouse		TNF signaling pathway	3/110	1.E-02
Female		KEGG_2019_Mouse		Primary immunodeficiency	2/36	1.E-02
Female		KEGG_2019_Mouse		ABC transporters	2/48	2.E-02
Female		KEGG_2019_Mouse		Cytosolic DNA-sensing pathway	2/61	3.E-02
Female	**greenyellow (224)**	GO_Biological_Process_2018	9	positive regulation of glycogen biosynthetic process (GO:0045725)	4/13	7.E-03
Female		GO_Biological_Process_2018		positive regulation of glycogen metabolic process (GO:0070875)	4/14	7.E-03
Female		GO_Biological_Process_2018		regulation of glycogen biosynthetic process (GO:0005979)	4/20	2.E-02
Female		GO_Biological_Process_2018		positive regulation of glucose transport (GO:0010828)	4/25	4.E-02
Female		GO_Biological_Process_2018		positive regulation of glucose import (GO:0046326)	4/26	4.E-02
Female		GO_Biological_Process_2018		positive regulation of fatty acid oxidation (GO:0046321)	3/11	4.E-02
Female		GO_Biological_Process_2018		monocarboxylic acid transport (GO:0015718)	5/53	4.E-02
Female		GO_Biological_Process_2018		regulation of glucose import (GO:0046324)	4/31	5.E-02
Female		GO_Biological_Process_2018		cellular respiration (GO:0045333)	5/57	5.E-02
Female	**lightyellow (83)**	GO_Biological_Process_2018	77	regulation of alcohol biosynthetic process (GO:1902930)	17/34	6.E-30
Female		GO_Biological_Process_2018		sterol biosynthetic process (GO:0016126)	17/40	8.E-29
Female		GO_Biological_Process_2018		regulation of cholesterol biosynthetic process (GO:0045540)	17/40	8.E-29
Female		GO_Biological_Process_2018		regulation of cholesterol metabolic process (GO:0090181)	17/41	1.E-28
Female		GO_Biological_Process_2018		regulation of steroid biosynthetic process (GO:0050810)	17/44	4.E-28
Female		GO_Biological_Process_2018		cholesterol biosynthetic process (GO:0006695)	16/35	5.E-28
Female		GO_Biological_Process_2018		secondary alcohol biosynthetic process (GO:1902653)	16/36	8.E-28
Female		GO_Biological_Process_2018		cholesterol metabolic process (GO:0008203)	15/68	1.E-20
Female		GO_Biological_Process_2018		fatty-acyl-CoA biosynthetic process (GO:0046949)	9/30	2.E-13
Female		GO_Biological_Process_2018		acetyl-CoA metabolic process (GO:0006084)	6/14	6.E-10
Female		KEGG_2019_Mouse	10	Steroid biosynthesis	10/19	9.E-18
Female		KEGG_2019_Mouse		Terpenoid backbone biosynthesis	7/23	2.E-10
Female		KEGG_2019_Mouse		Biosynthesis of unsaturated fatty acids	7/32	2.E-09
Female		KEGG_2019_Mouse		Pyruvate metabolism	7/38	5.E-09
Female		KEGG_2019_Mouse		Fatty acid biosynthesis	4/14	6.E-06
Female		KEGG_2019_Mouse		Fatty acid elongation	4/29	1.E-04
Female		KEGG_2019_Mouse		Propanoate metabolism	4/31	1.E-04
Female		KEGG_2019_Mouse		PPAR signaling pathway	5/85	4.E-04
Female		KEGG_2019_Mouse		Glycerolipid metabolism	4/61	2.E-03
Female		KEGG_2019_Mouse		Glycolysis / Gluconeogenesis	4/67	2.E-03
Female	**royalblue (68)**	GO_Biological_Process_2018	23	entrainment of circadian clock by photoperiod (GO:0043153)	5/20	2.E-06
Female		GO_Biological_Process_2018		photoperiodism (GO:0009648)	5/21	2.E-06
Female		GO_Biological_Process_2018		negative regulation of glucocorticoid receptor signaling pathway (GO:2000323)	3/6	1.E-04
Female		GO_Biological_Process_2018		regulation of glucocorticoid receptor signaling pathway (GO:2000322)	3/7	2.E-04
Female		GO_Biological_Process_2018		negative regulation of circadian rhythm (GO:0042754)	3/9	3.E-04
Female		GO_Biological_Process_2018		regulation of transcription from RNA polymerase II promoter (GO:0006357)	17/1478	5.E-04
Female		GO_Biological_Process_2018		hexose transport (GO:0008645)	3/16	2.E-03
Female		GO_Biological_Process_2018		glycogen biosynthetic process (GO:0005978)	3/18	2.E-03
Female		GO_Biological_Process_2018		glucan biosynthetic process (GO:0009250)	3/18	2.E-03

Female		GO_Biological_Process_2018		negative regulation of transcription, DNA-templated (GO:0045892)	11/813	4.E-03
Female		KEGG_2019_Mouse	1	Circadian rhythm	7/30	6.E-10
Female	tan (210)	GO_Biological_Process_2018	108	translational termination (GO:0006415)	16/96	3.E-12
Female		GO_Biological_Process_2018		mitochondrial translational elongation (GO:0070125)	15/87	7.E-12
Female		GO_Biological_Process_2018		mitochondrial translational termination (GO:0070126)	15/89	7.E-12
Female		GO_Biological_Process_2018		translational elongation (GO:0006414)	15/105	7.E-11
Female		GO_Biological_Process_2018		mitochondrial translation (GO:0032543)	15/107	7.E-11
Female		GO_Biological_Process_2018		proteasomal ubiquitin-independent protein catabolic process (GO:0010499)	9/24	2.E-10
Female		GO_Biological_Process_2018		regulation of transcription from RNA polymerase II promoter in response to stress (GO:0043618)	13/93	2.E-09
Female		GO_Biological_Process_2018		regulation of hematopoietic stem cell differentiation (GO:1902036)	11/76	4.E-08
Female		GO_Biological_Process_2018		mitochondrial respiratory chain complex assembly (GO:0033108)	12/97	4.E-08
Female		GO_Biological_Process_2018		regulation of hematopoietic progenitor cell differentiation (GO:1901532)	11/77	4.E-08
Female		KEGG_2019_Mouse	8	Proteasome	9/46	1.E-07
Female		KEGG_2019_Mouse		Oxidative phosphorylation	9/134	5.E-04
Female		KEGG_2019_Mouse		Ribosome	10/170	5.E-04
Female		KEGG_2019_Mouse		Alzheimer disease	10/175	5.E-04
Female		KEGG_2019_Mouse		Thermogenesis	11/231	8.E-04
Female		KEGG_2019_Mouse		Parkinson disease	8/144	3.E-03
Female		KEGG_2019_Mouse		Non-alcoholic fatty liver disease (NAFLD)	8/151	3.E-03
Female		KEGG_2019_Mouse		Huntington disease	9/192	3.E-03
Male	black (130)	GO_Biological_Process_2018	113	sterol biosynthetic process (GO:0016126)	18/40	5.E-27
Male		GO_Biological_Process_2018		secondary alcohol biosynthetic process (GO:1902653)	17/36	3.E-26
Male		GO_Biological_Process_2018		regulation of alcohol biosynthetic process (GO:1902930)	16/34	1.E-24
Male		GO_Biological_Process_2018		cholesterol biosynthetic process (GO:0006695)	16/35	1.E-24
Male		GO_Biological_Process_2018		regulation of cholesterol biosynthetic process (GO:0045540)	16/40	2.E-23
Male		GO_Biological_Process_2018		regulation of cholesterol metabolic process (GO:0090181)	16/41	2.E-23
Male		GO_Biological_Process_2018		regulation of steroid biosynthetic process (GO:0050810)	16/44	8.E-23
Male		GO_Biological_Process_2018		cholesterol metabolic process (GO:0008203)	15/68	1.E-17
Male		GO_Biological_Process_2018		fatty-acyl-CoA biosynthetic process (GO:0046949)	9/30	1.E-11
Male		GO_Biological_Process_2018		lipid biosynthetic process (GO:0008610)	11/72	6.E-11
Male		KEGG_2019_Mouse	49	Steroid biosynthesis	9/19	1.E-13
Male		KEGG_2019_Mouse		Pyruvate metabolism	10/38	2.E-12
Male		KEGG_2019_Mouse		Terpenoid backbone biosynthesis	8/23	3.E-11
Male		KEGG_2019_Mouse		Propanoate metabolism	8/31	4.E-10
Male		KEGG_2019_Mouse		Glycolysis / Gluconeogenesis	8/67	2.E-07
Male		KEGG_2019_Mouse		Biosynthesis of unsaturated fatty acids	6/32	8.E-07
Male		KEGG_2019_Mouse		Fatty acid biosynthesis	4/14	2.E-05
Male		KEGG_2019_Mouse		Citrate cycle (TCA cycle)	5/32	2.E-05
Male		KEGG_2019_Mouse		Fatty acid degradation	5/50	2.E-04
Male		KEGG_2019_Mouse		Oxidative phosphorylation	7/134	2.E-04
Male	blue (3482)	GO_Biological_Process_2018	184	mitotic cell cycle phase transition (GO:0044772)	88/221	3.E-12
Male		GO_Biological_Process_2018		cellular protein modification process (GO:0006464)	252/1001	4.E-08
Male		GO_Biological_Process_2018		extracellular matrix organization (GO:0030198)	79/229	2.E-07
Male		GO_Biological_Process_2018		proteasome-mediated ubiquitin-dependent protein catabolic process (GO:0043161)	93/291	3.E-07
Male		GO_Biological_Process_2018		regulation of mitotic cell cycle phase transition (GO:1901990)	66/184	5.E-07
Male		GO_Biological_Process_2018		G2/M transition of mitotic cell cycle (GO:0000086)	50/126	1.E-06
Male		GO_Biological_Process_2018		cell cycle G2/M phase transition (GO:0044839)	50/127	1.E-06

Male		GO_Biological_Process_2018		post-translational protein modification (GO:0043687)	105/357	2.E-06
Male		GO_Biological_Process_2018		mRNA processing (GO:0006397)	86/283	1.E-05
Male		GO_Biological_Process_2018		G1/S transition of mitotic cell cycle (GO:0000082)	42/105	1.E-05
Male		KEGG_2019_Mouse		Cell cycle	49/123	5.E-07
Male		KEGG_2019_Mouse		Focal adhesion	61/199	2.E-04
Male		KEGG_2019_Mouse		ECM-receptor interaction	32/83	3.E-04
Male		KEGG_2019_Mouse		Protein processing in endoplasmic reticulum	50/163	1.E-03
Male		KEGG_2019_Mouse		p53 signaling pathway	27/71	1.E-03
Male		KEGG_2019_Mouse	37	Pathways in cancer	127/535	2.E-03
Male		KEGG_2019_Mouse		Hepatocellular carcinoma	50/171	2.E-03
Male		KEGG_2019_Mouse		TGF-beta signaling pathway	31/91	2.E-03
Male		KEGG_2019_Mouse		Small cell lung cancer	31/92	3.E-03
Male		KEGG_2019_Mouse		RNA transport	48/167	3.E-03
Male		GO_Biological_Process_2018		protein modification by small protein conjugation (GO:0032446)	52/398	4.E-10
Male		GO_Biological_Process_2018		protein ubiquitination (GO:0016567)	60/506	4.E-10
Male		GO_Biological_Process_2018		protein modification by small protein removal (GO:0070646)	37/261	6.E-08
Male		GO_Biological_Process_2018		ubiquitin-dependent protein catabolic process (GO:0006511)	43/341	6.E-08
Male		GO_Biological_Process_2018		protein deubiquitination (GO:0016579)	36/257	1.E-07
Male		GO_Biological_Process_2018	56	protein ubiquitination involved in ubiquitin-dependent protein catabolic process (GO:0042787)	18/70	2.E-07
Male	**brown (839)**	GO_Biological_Process_2018		protein polyubiquitination (GO:0000209)	34/283	1.E-05
Male		GO_Biological_Process_2018		cellular protein modification process (GO:0006464)	75/1001	2.E-04
Male		GO_Biological_Process_2018		proteasome-mediated ubiquitin-dependent protein catabolic process (GO:0043161)	31/291	5.E-04
Male		GO_Biological_Process_2018		proteasomal protein catabolic process (GO:0010498)	27/237	6.E-04
Male		KEGG_2019_Mouse		Ubiquitin mediated proteolysis	24/138	6.E-07
Male		KEGG_2019_Mouse	4	Peroxisome	13/84	4.E-03
Male		KEGG_2019_Mouse		Mitophagy	11/63	4.E-03
Male		KEGG_2019_Mouse		RNA degradation	11/83	4.E-02
Male		GO_Biological_Process_2018		translation (GO:0006412)	64/232	4.E-45
Male		GO_Biological_Process_2018		protein targeting to ER (GO:0045047)	45/97	4.E-43
Male		GO_Biological_Process_2018		SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	43/89	3.E-42
Male		GO_Biological_Process_2018		nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (GO:0000184)	46/112	2.E-41
Male		GO_Biological_Process_2018	95	cotranslational protein targeting to membrane (GO:0006613)	43/93	2.E-41
Male		GO_Biological_Process_2018		peptide biosynthetic process (GO:0043043)	54/174	2.E-41
Male		GO_Biological_Process_2018		viral transcription (GO:0019083)	45/113	5.E-40
Male		GO_Biological_Process_2018		viral gene expression (GO:0019080)	44/110	3.E-39
Male	**green (509)**	GO_Biological_Process_2018		nuclear-transcribed mRNA catabolic process (GO:0000956)	49/174	2.E-35
Male		GO_Biological_Process_2018		cellular macromolecule biosynthetic process (GO:0034645)	64/367	4.E-33
Male		KEGG_2019_Mouse		Ribosome	57/170	3.E-46
Male		KEGG_2019_Mouse		Parkinson disease	17/144	1.E-05
Male		KEGG_2019_Mouse		Oxidative phosphorylation	16/134	2.E-05
Male		KEGG_2019_Mouse		Huntington disease	19/192	2.E-05
Male		KEGG_2019_Mouse	8	Alzheimer disease	18/175	2.E-05
Male		KEGG_2019_Mouse		Thermogenesis	18/231	8.E-04
Male		KEGG_2019_Mouse		Non-alcoholic fatty liver disease (NAFLD)	13/151	3.E-03
Male		KEGG_2019_Mouse		Retrograde endocannabinoid signaling	12/150	1.E-02
Male		GO_Biological_Process_2018	14	ribosome biogenesis (GO:0042254)	18/226	4.E-07

Male	red (245)	GO_Biological_Process_2018		rRNA processing (GO:0006364)	16/202	2.E-06
Male		GO_Biological_Process_2018		ncRNA processing (GO:0034470)	15/227	5.E-05
Male		GO_Biological_Process_2018		rRNA metabolic process (GO:0016072)	14/200	5.E-05
Male		GO_Biological_Process_2018		maturation of LSU-rRNA (GO:0000470)	6/22	5.E-05
Male		GO_Biological_Process_2018		maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA, LSU-rRNA) (GO:0000463)	4/9	5.E-04
Male		GO_Biological_Process_2018		ribosomal large subunit biogenesis (GO:0042273)	7/63	2.E-03
Male		GO_Biological_Process_2018		maturation of 5.8S rRNA (GO:0000460)	4/17	7.E-03
Male		GO_Biological_Process_2018		adipose tissue development (GO:0060612)	3/9	2.E-02
Male		GO_Biological_Process_2018		regulation of glucose metabolic process (GO:0010906)	4/25	3.E-02
Male		KEGG_2019_Mouse	1	Ribosome biogenesis in eukaryotes	10/115	2.E-04
Male	turquoise (4351)	GO_Biological_Process_2018		neutrophil mediated immunity (GO:0002446)	213/487	1.E-25
Male		GO_Biological_Process_2018		neutrophil degranulation (GO:0043312)	209/479	2.E-25
Male		GO_Biological_Process_2018		neutrophil activation involved in immune response (GO:0002283)	210/483	2.E-25
Male		GO_Biological_Process_2018		cytokine-mediated signaling pathway (GO:0019221)	217/633	2.E-11
Male		GO_Biological_Process_2018		regulation of type I interferon production (GO:0032479)	47/85	7.E-09
Male		GO_Biological_Process_2018	352	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile (GO:0000377)	96/236	8.E-09
Male		GO_Biological_Process_2018		positive regulation of cytokine production (GO:0001819)	89/220	6.E-08
Male		GO_Biological_Process_2018		mRNA splicing, via spliceosome (GO:0000398)	101/261	6.E-08
Male		GO_Biological_Process_2018		vesicle-mediated transport (GO:0016192)	143/410	6.E-08
Male		GO_Biological_Process_2018		protein transport (GO:0015031)	119/326	8.E-08
Male		KEGG_2019_Mouse		Lysosome	81/124	1.E-23
Male		KEGG_2019_Mouse		Osteoclast differentiation	72/128	8.E-16
Male		KEGG_2019_Mouse		Fc gamma R-mediated phagocytosis	50/87	2.E-11
Male		KEGG_2019_Mouse		NF-kappa B signaling pathway	55/102	4.E-11
Male		KEGG_2019_Mouse		Apoptosis	68/141	6.E-11
Male		KEGG_2019_Mouse	107	Chemokine signaling pathway	83/197	2.E-09
Male		KEGG_2019_Mouse		NOD-like receptor signaling pathway	85/205	2.E-09
Male		KEGG_2019_Mouse		C-type lectin receptor signaling pathway	55/112	2.E-09
Male		KEGG_2019_Mouse		Epstein-Barr virus infection	92/229	2.E-09
Male		KEGG_2019_Mouse		Toll-like receptor signaling pathway	50/99	4.E-09
Male		Jensen_DISEASES		inclusion-cell disease	11/14	1.E-02
Male		Jensen_DISEASES		Crohn's disease	81/248	1.E-02
Male		Jensen_DISEASES		Arthritis	64/186	1.E-02
Male		Jensen_DISEASES	6	Amyloidosis	15/27	3.E-02
Male		Jensen_DISEASES		Histoplasmosis	11/17	3.E-02
Male		Jensen_DISEASES		Lung disease	43/119	3.E-02
Male	yellow (576)	GO_Biological_Process_2018		Golgi vesicle transport (GO:0048193)	29/271	3.E-06
Male		GO_Biological_Process_2018		retrograde vesicle-mediated transport, Golgi to ER (GO:0006890)	13/80	3.E-04
Male		GO_Biological_Process_2018		ERBB signaling pathway (GO:0038127)	13/80	3.E-04
Male		GO_Biological_Process_2018		RNA metabolic process (GO:0016070)	19/191	2.E-03
Male		GO_Biological_Process_2018	46	ER to Golgi vesicle-mediated transport (GO:0006888)	18/180	2.E-03
Male		GO_Biological_Process_2018		vascular endothelial growth factor receptor signaling pathway (GO:0048010)	11/70	2.E-03
Male		GO_Biological_Process_2018		negative regulation of sterol transport (GO:0032372)	4/6	3.E-03
Male		GO_Biological_Process_2018		viral life cycle (GO:0019058)	13/107	3.E-03
Male		GO_Biological_Process_2018		regulation of translation (GO:0006417)	19/213	3.E-03

Male		GO_Biological_Process_2018	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile (GO:0000377)	20/236	4.E-03
Male		KEGG_2019_Mouse	Protein processing in endoplasmic reticulum	19/163	5.E-05
Male		KEGG_2019_Mouse	Mitophagy	10/63	1.E-03
Male		KEGG_2019_Mouse	Autophagy	14/130	2.E-03
Male		KEGG_2019_Mouse	Bacterial invasion of epithelial cells	10/74	3.E-03
Male		KEGG_2019_Mouse	Renal cell carcinoma	9/68	6.E-03
Male		KEGG_2019_Mouse	Hepatitis C	14/160	8.E-03
Male		KEGG_2019_Mouse	Cholesterol metabolism	7/49	2.E-02
Male		KEGG_2019_Mouse	ErbB signaling pathway	9/84	2.E-02
Male		KEGG_2019_Mouse	Fluid shear stress and atherosclerosis	12/143	2.E-02
Male		KEGG_2019_Mouse	Pancreatic cancer	8/75	3.E-02

11

[a]Combined score is described as c = log(p) * z , where c = the combined score, p = Fisher exact test p-value, and z = z-score for deviation from expected rank.

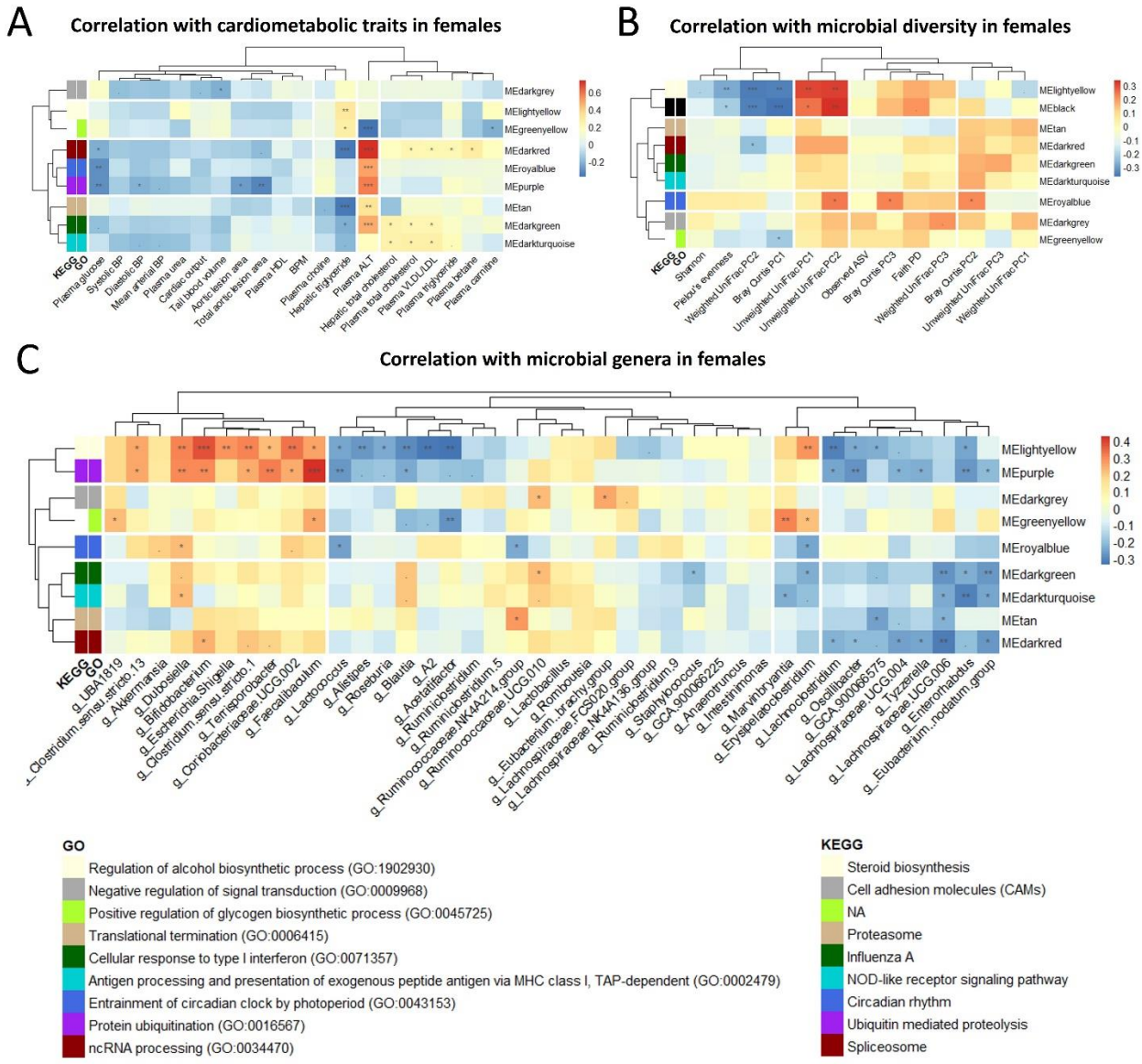


Figure 5.11. Hepatic coexpression gene modules are associated with cardiometabolic traits and gut microbiota in females.

For the WGCNA analysis, log2 transformed 13,094 transcripts greater than 1 TPM measured in 164 DO-F1 liver samples were included. (A-C) Spearman correlation between liver gene modules and cardiometabolic traits (A), microbial diversity (B), genus levels (C) in females. Module names were shown along the right axis and top-enriched GO/KEGG terms in legend. The p-values were adjusted using the Benjamini–Hochberg (BH) FDR procedure. "***"$P < 0.001$, "**"$P < 0.01$, "*"$P < 0.05$, "."$P < 0.10$.

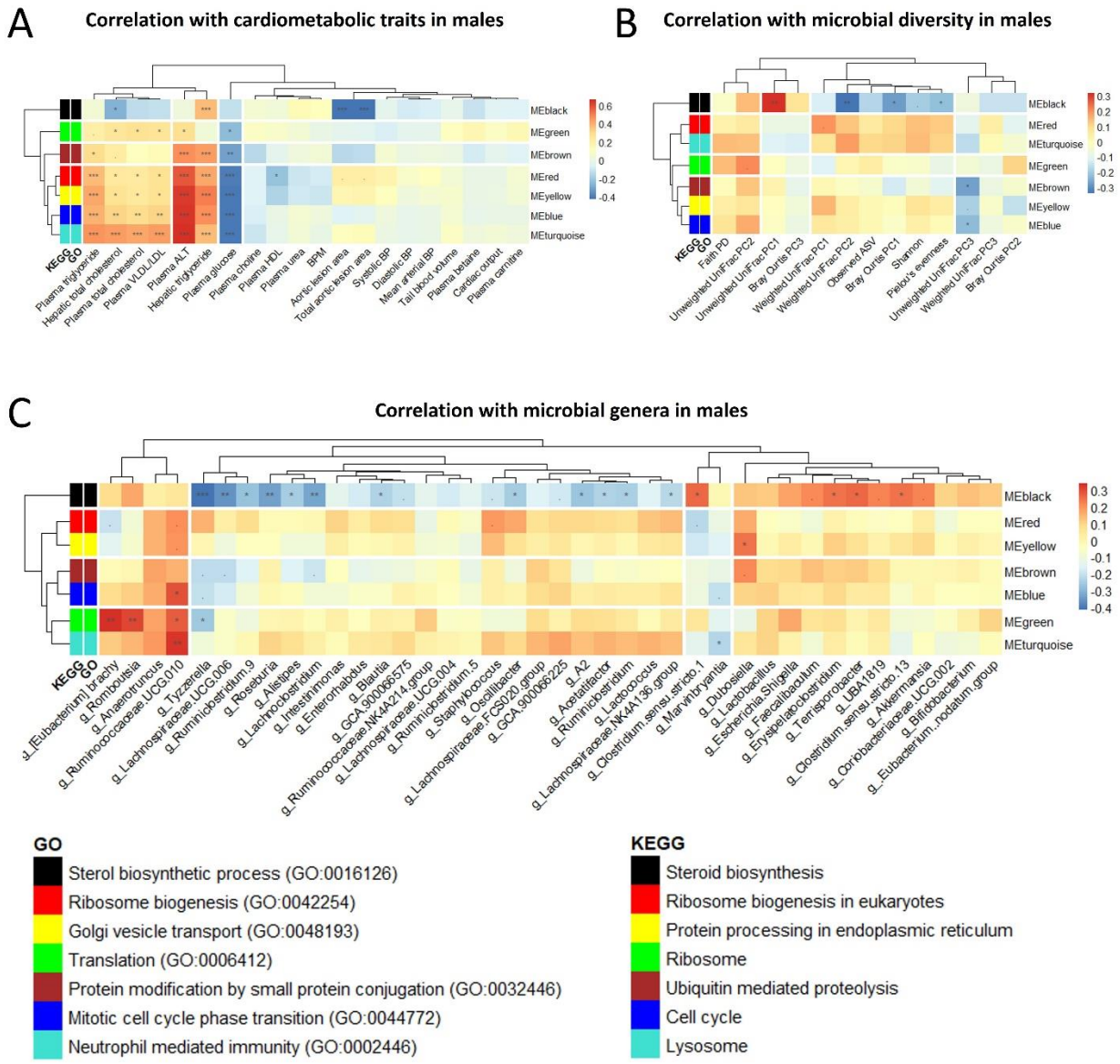


Figure 5.12. Hepatic coexpression gene modules are associated with cardiometabolic traits and gut microbiota in males.

For the WGCNA analysis, log2 transformed 13,094 transcripts greater than 1 TPM measured in 164 DO-F1 liver samples were included. (A-C) Spearman correlation between liver gene modules and cardiometabolic traits (A), microbial diversity (B), genus levels (C) in males. Module names were shown along the right axis and top-enriched GO/KEGG terms in legend. The p-values were adjusted using the Benjamini–Hochberg (BH) FDR procedure. "***"P < 0.001, "**"P < 0.01, "*"P < 0.05, "."P < 0.10.

Table 5.12. The number of significantly correlated MetaCyc pathways and EC gene families with hepatic coexpression gene modules in each sex.

Sex	Module	# Significantly correlated MetaCyc pathways	# Significantly correlated EC gene families
Male mice	MEblack	111	538
	MEred	3	13
	MEyellow	1	7
	MEgreen	12	32
	MEbrown	4	18
	MEblue	0	0
	MEturquoise	2	10
Female mice	MElightyellow	117	638
	MEdarkgrey	0	5
	MEgreenyellow	15	84
	MEtan	2	56
	MEdarkgreen	3	52
	MEdarkturquoise	9	32
	MEroyalblue	34	111
	MEpurple	111	544
	MEdarkred	9	76

5.5.7. Genetic regulation of microbial genus levels and hepatic gene expression

In Chapter 4, we confirmed the effects of host genetics on liver transcriptome in DO-F1 mice, and RNA sequencing of liver tissue yielded significant 2722 *cis*-regulation of the transcriptome (*cis*-eQTLs and ±4 Mb of the gene) in all mice, 1408 *cis*-eQTL in female mice, and 1061 *cis*-eQTL in male mice at a $P<0.05$. In chapter 5, we integrated mbQTL with liver *cis*-eQTLs to identify genetic variants that influence both microbial taxa and liver transcripts. We identified 5 loci in female mice and 1 locus in male mice that co-mapped to both mbQTL and *cis*-eQTL, in which the microbial genus level and liver transcript also correlated in each sex (**Figure 5.13A and 5.13B and Table 5.13**). For example, *Romboutsia* genus and ASV levels on chromosome 5 in female mice were co-mapped with SNPs in the *Klhl5* (kelch-like 5) gene, and a significant positive correlation was observed between liver *Klhl5* and *Romboutsia* genus abundance in female mice. In addition, the *Klhl5* gene was included in the purple module, which correlated with the aortic lesion area and the most gut microbial taxa and functional profiling, suggesting that *Klhl5* may be involved in the regulation of *Romboutsia* abundance in the intestine (**Figure 5.14A and 5.14B and Table 5.13**). Specifically, allelic variation (T/C or C/C) at rs50048053 in the *Klhl5* locus was associated with the abundance of liver *Klhl5* transcript and *Romboutsia* abundance in female mice (**Figure 5.14C**). Furthermore, *Klhl5 cis*-eQTL shared the same allele effect pattern as the *Romboutsia* abundance; both were associated with increased CAST allele and decreased PWK allele (**Figure 5.14D**).

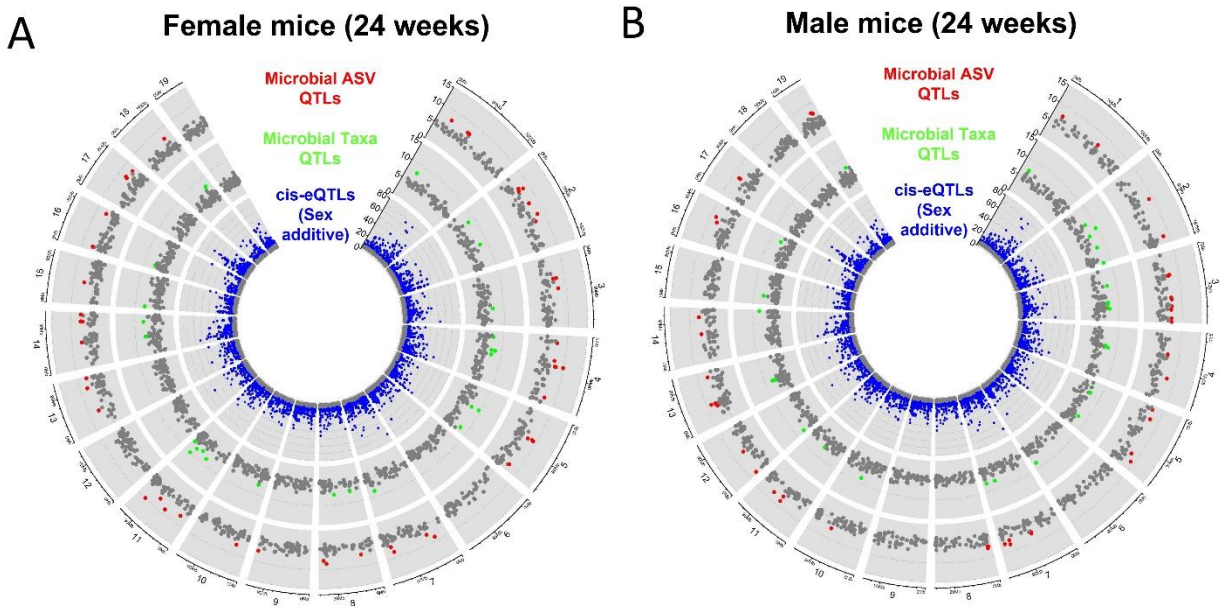


Figure 5.13. Genetic architecture of quantitative trait loci (QTL) for microbial taxa abundance and liver transcriptome in DO-F1 mice.
(A,B) Genetic architecture of QTL for microbial taxa abundance, microbial amplicon sequence variants (ASVs), and *cis*-acting liver transcripts in females (A) and males (B). The outer layer shows the chromosome location. LOD range is shown for each track. Each dot represents a QTL on each chromosome of the mouse genome for a given trait. Grey dots denote QTLs with LOD < 7.

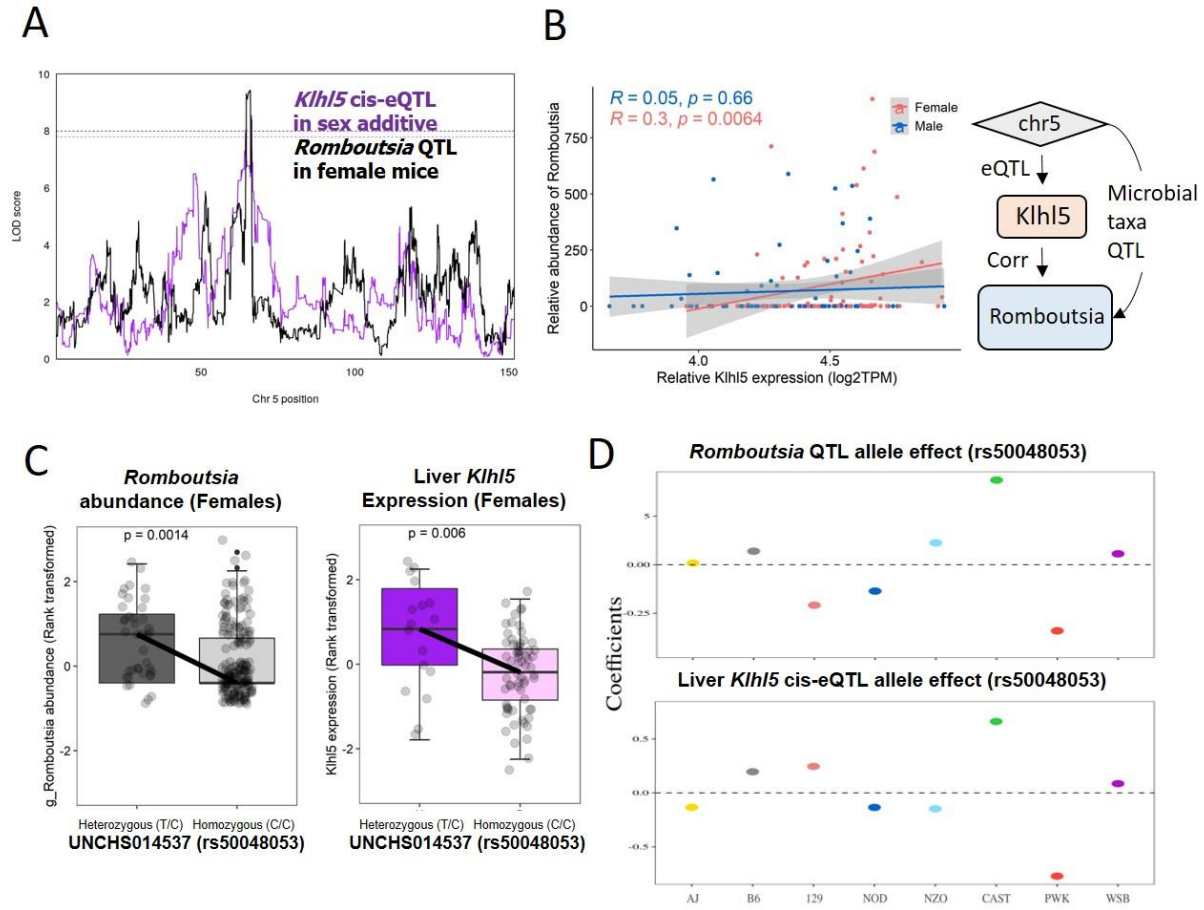


Figure 5.14. Systems genetic analysis of gut microbial genus levels and liver transcriptome in DO-F1 mice.

(A) LOD profiles on Chromosome 5 highlighting a locus significantly associated with the liver *Klhl5* expression and abundance of *Romboutsia* genus level in female mice. Each dashed line (purple for *Klhl5 cis*-eQTL and black for *Romboutsia* abundance QTL) represents significant thresholds (P< 0.05). (B) *Klhl5* gene expression is positively correlated with *Romboutsia* abundance in females (R = 0.3, p = 0.0064), not males (R = 0.05, p = 0.66). Inset on right summarizes the proposed genetic interaction. (C) Association between a SNP (rs50048053) in *Klhl5* gene and *Romboutsia* abundance or liver *Klhl5* gene expression. Heterozygous SNP (T/C) is from wild-type inbred strains (CAST or PWK strains) and homozygous SNP (C/C) is from laboratory inbred strains (A/J, 129, NOD, NZO, or WSB strains). (D) Estimated founder strain levels of *Romboutsia* abundance and liver *Klhl5* gene expression were inferred in the DO-F1 population from the founder strain coefficients observed at the corresponding QTL on chromosome 5. B6, C57BL/6J; 129, 129S1/SvlmJ; NOD, NOD/ShiLtJ; NZO, NZO/HILtJ; CAST, CAST/EiJ; PWJ, PWK/PhJ; WSB, WSB/EiJ.

Table 5.13. Microbial genus QTL, liver cis-eQTL, and a direct correlation between the mapped trait and transcript in three models.

Model	Microbial genus QTL				cis-eQTL				Trait:Transcript correlation (r and p-value)			
	Trait	Chr	Position (Mbp)	LOD	Gene	Chr	Position (Mbp)	LOD	Module color (Females)	Module color (Males)	r	adjusted p-value
Female	g_Romboutsia	2	35.00	8.24	**Cntrl**	2	35.11	7.80	NA	NA	**-0.25**	**0.05**
Female	g_Romboutsia	2	35.00	8.24	Fam129b	2	32.64	13.74	darkred	turquoise	0.16	0.26
Female	g_Romboutsia	2	35.00	8.24	Gsn	2	35.26	19.34	darkred	blue	0.10	0.50
Female	g_Romboutsia	2	35.00	8.24	Hc	2	34.98	18.45	NA	NA	0.05	0.75
Female	g_Romboutsia	2	35.00	8.24	Lrsam1	2	32.93	14.77	darkred	turquoise	0.23	0.07
Female	g_Romboutsia	2	35.00	8.24	Mrrf	2	36.14	9.45	NA	turquoise	-0.04	0.77
Female	g_Romboutsia	2	35.00	8.24	Phf19	2	32.90	11.27	darkred	NA	-0.02	0.89
Female	g_Romboutsia	2	35.00	8.24	Rc3h2	2	36.00	8.69	purple	brown	0.10	0.52
Female	g_Romboutsia	2	35.00	8.24	Rpl12	2	32.96	25.92	NA	NA	0.08	0.58
Female	g_Romboutsia	2	35.00	8.24	Stom	2	35.31	15.42	darkred	yellow	0.18	0.16
Female	g_Romboutsia	2	35.00	8.24	Stxbp1	2	32.79	8.25	darkred	turquoise	0.04	0.81
Female	g_Lachnospiraceae.FCS020	2	110.79	8.05	Aven	2	112.49	8.03	NA	green	-0.09	0.52
Female	g_Lachnospiraceae.FCS020	2	110.79	8.05	Fmn1	2	113.33	9.65	darkred	turquoise	0.14	0.28
Female	g_Lachnospiraceae.FCS020	2	110.79	8.05	Mettl15	2	109.09	10.68	purple	blue	0.06	0.65
Female	g_Lachnospiraceae.FCS020	2	110.79	8.05	Nop10	2	112.26	9.06	darkred	blue	-0.09	0.52
Female	g_GCA.900066575	4	58.16	8.27	Ptgr1	4	58.97	19.50	darkred	turquoise	-0.21	0.10
Female	g_GCA.900066575	4	58.16	8.27	**Susd1**	4	59.31	10.53	NA	NA	**0.35**	**4.E-03**
Female	g_Romboutsia	5	65.91	9.43	**Klhl5**	5	65.11	8.54	purple	blue	**0.30**	**0.01**
Female	g_Romboutsia	5	65.91	9.43	Tlr6	5	64.95	13.45	darkred	turquoise	0.21	0.10
Female	g_Lactococcus	10	26.71	7.45	Akap7	10	25.17	14.64	NA	blue	-0.01	0.95
Female	g_Lactococcus	10	26.71	7.45	**Arhgap18**	10	26.77	11.18	darkred	blue	**-0.25**	**0.05**
Female	g_Lactococcus	10	26.71	7.45	Gm10145	10	28.35	28.33	NA	NA	0.24	0.06
Female	g_Lactococcus	10	26.71	7.45	**Ptprk**	10	28.07	9.20	purple	blue	**-0.45**	**1.E-04**
Female	g_Lactococcus	10	26.71	7.45	Smlr1	10	25.66	9.74	NA	NA	-0.03	0.85
Female	g_Erysipelatoclostridium	11	53.80	8.48	Atox1	11	55.45	17.88	darkred	blue	-0.17	0.21
Female	g_Erysipelatoclostridium	11	53.80	8.48	Gm10447	11	53.05	18.11	NA	NA	-0.13	0.34
Female	g_Erysipelatoclostridium	11	53.80	8.48	Lyrm7	11	54.83	32.22	NA	NA	0.22	0.09
Female	g_Coriobacteriaceae.UCG.002	11	102.36	9.93	Acbd4	11	103.10	23.56	NA	NA	0.03	0.83
Female	g_Coriobacteriaceae.UCG.002	11	102.36	9.93	Dcakd	11	102.99	14.52	darkred	blue	0.08	0.59
Female	g_Coriobacteriaceae.UCG.002	11	102.36	9.93	Dusp3	11	101.97	8.73	darkred	turquoise	0.01	0.96
Female	g_Coriobacteriaceae.UCG.002	11	102.36	9.93	Grn	11	102.43	10.99	darkred	turquoise	0.00	0.98
Female	g_Coriobacteriaceae.UCG.002	11	102.36	9.93	Lsm12	11	102.16	9.93	darkred	blue	-0.01	0.93
Female	g_Coriobacteriaceae.UCG.002	11	102.36	9.93	Tmem101	11	102.15	9.11	darkred	turquoise	0.10	0.49
Female	g_Coriobacteriaceae.UCG.002	11	102.36	9.93	Tmub2	11	102.28	12.50	darkred	turquoise	-0.10	0.47
Male	g_UBA1819	2	77.37	9.74	Ccdc141	2	77.05	18.82	greenyellow	NA	0.11	0.43
Male	g_UBA1819	2	77.37	9.74	E030042O20Rik	2	76.02	26.31	NA	NA	0.02	0.90
Male	g_UBA1819	2	77.37	9.74	Gm13657	2	74.31	11.02	darkred	blue	0.01	0.93
Male	g_UBA1819	2	77.37	9.74	Prkra	2	76.71	11.76	NA	green	0.14	0.28
Male	g_Enterorhabdus	2	181.82	8.08	Col20a1	2	180.90	21.91	purple	NA	-0.13	0.33
Male	g_Enterorhabdus	2	181.82	8.08	Polr3k	2	181.87	9.75	purple	blue	-0.06	0.68
Male	g_Enterorhabdus	2	181.82	8.08	Slc2a4rg-ps	2	181.35	8.48	NA	NA	0.09	0.55
Male	g_Clostridium.sensu.stricto.1	4	62.98	7.84	Alad	4	62.56	31.64	NA	NA	0.06	0.69
Male	g_Clostridium.sensu.stricto.1	4	62.98	7.84	Atp6v1g1	4	63.93	11.96	darkred	turquoise	-0.01	0.94
Male	g_Clostridium.sensu.stricto.1	4	62.98	7.84	Gm12909	4	59.37	16.52	NA	NA	0.07	0.61
Male	g_Clostridium.sensu.stricto.1	4	62.98	7.84	Hdhd3	4	62.12	24.33	NA	NA	0.04	0.79
Male	g_Clostridium.sensu.stricto.1	4	62.98	7.84	Hsdl2	4	62.61	8.43	NA	brown	-0.03	0.82
Male	g_Clostridium.sensu.stricto.1	4	62.98	7.84	Mup1	4	58.80	18.64	darkred	blue	0.02	0.88
Male	g_Clostridium.sensu.stricto.1	4	62.98	7.84	Mup13	4	62.62	18.41	NA	NA	0.09	0.50
Male	g_Clostridium.sensu.stricto.1	4	62.98	7.84	Mup17	4	62.65	10.94	darkred	blue	-0.11	0.42
Male	g_Clostridium.sensu.stricto.1	4	62.98	7.84	Mup2	4	61.86	11.10	darkred	blue	-0.06	0.70
Male	g_Clostridium.sensu.stricto.1	4	62.98	7.84	Mup21	4	64.27	10.04	NA	blue	0.02	0.90
Male	g_Clostridium.sensu.stricto.1	4	62.98	7.84	Mup9	4	62.61	10.24	darkred	blue	-0.10	0.49
Male	g_Clostridium.sensu.stricto.1	4	62.98	7.84	Mup-ps13	4	62.10	14.14	darkred	turquoise	-0.05	0.74
Male	g_Clostridium.sensu.stricto.1	4	62.98	7.84	Mup-ps17	4	62.10	17.18	NA	NA	-0.10	0.48
Male	g_Clostridium.sensu.stricto.1	4	62.98	7.84	Orm1	4	63.11	10.50	darkred	turquoise	-0.15	0.24
Male	g_Clostridium.sensu.stricto.1	4	62.98	7.84	**Orm3**	4	63.07	15.57	NA	turquoise	**-0.25**	**0.05**
Male	g_Clostridium.sensu.stricto.1	4	62.98	7.84	Ptgr1	4	59.01	19.50	darkred	turquoise	-0.24	0.06
Male	g_Clostridium.sensu.stricto.1	4	62.98	7.84	Susd1	4	59.36	10.53	NA	NA	-0.16	0.21
Male	g_Ruminiclostridium.9	13	45.86	8.05	Fam8a1	13	46.63	9.04	purple	blue	0.00	0.99
Male	g_Ruminiclostridium.9	13	45.86	8.05	Nhlrc1	13	46.22	10.10	NA	turquoise	0.20	0.13
Male	g_Ruminiclostridium.9	13	45.86	8.05	Tpmt	13	47.02	45.90	NA	NA	0.04	0.77

Additive	g_Ruminococcaceae.UCG.010	1	172.48	8.64	Ackr1	1	172.91	8.13	darkred	turquoise	0.12	0.15
Additive	g_Ruminococcaceae.UCG.010	1	172.48	8.64	Apcs	1	172.85	16.03	NA	NA	0.01	0.95
Additive	g_Ruminococcaceae.UCG.010	1	172.48	8.64	Dusp23	1	172.91	14.11	purple	NA	-0.09	0.31
Additive	g_Ruminococcaceae.UCG.010	1	172.48	8.64	Kcnj10	1	172.10	13.76	darkred	turquoise	0.06	0.53
Additive	g_Ruminococcaceae.NK4A214	6	72.73	8.38	Elmod3	6	72.64	16.81	NA	NA	0.03	0.24
Additive	g_Ruminococcaceae.NK4A214	6	72.73	8.38	Gm38832	6	72.64	12.29	NA	NA	0.12	0.18
Additive	g_Ruminococcaceae.NK4A214	6	72.73	8.38	Mat2a	6	72.40	14.39	darkred	turquoise	0.02	0.34
Additive	g_Ruminococcaceae.NK4A214	6	72.73	8.38	Rnf181	6	72.59	35.42	darkred	turquoise	-0.09	0.25
Additive	g_Ruminococcaceae.NK4A214	6	72.73	8.38	Tmem150a	6	72.67	10.53	darkred	green	0.10	0.04
Additive	g_Romboutsia	10	117.97	8.19	Cpm	10	117.96	9.66	NA	NA	0.16	0.07
Additive	g_Terrisporobacter	16	14.44	9.02	Mpv17l	16	13.04	9.73	darkred	NA	-0.03	0.76
Additive	g_Terrisporobacter	16	14.44	9.02	Nde1	16	13.96	8.43	darkred	turquoise	-0.09	0.31
Additive	g_Terrisporobacter	16	14.44	9.02	Ntan1	16	13.96	19.86	darkred	blue	-0.05	0.59
Additive	g_Terrisporobacter	16	14.44	9.02	Pdxdc1	16	14.95	10.88	NA	blue	0.04	0.65
Additive	g_Terrisporobacter	16	14.44	9.02	Top3b	16	12.88	8.95	darkred	NA	-0.05	0.55

Significance was considered when QTL results were $P<0.05$.

[a]cis-eQTLs are the result of the sex additive model.

5.5.8. *Ptprk* is a candidate gene for aortic lesion area and *Lactococcus* abundance

Lastly, we searched candidate genes in the QTL confidence interval via trait-transcript correlation and high-resolution association mapping on chromosome 10 in females identified above. By applying systems genetics using cardiometabolic traits, microbial taxa and functional profiling, and liver transcriptome, we found that only expression of *Ptprk* gene, which encodes for the protein tyrosine phosphatase receptor type K transporter and belongs to the purple module (**Table 5.13**), was significantly correlated with both aortic lesion area (R=-0.25, P≤0.05) and *Lactococcus* abundance (R=-0.45, P≤0.001) in females, not males (**Figure 5.15A and 5.15B**). The purple module to which *Ptprk* gene belongs also had a significant correlation with both aortic lesion area (R=-0.27, P≤0.05) and *Lactococcus* abundance (R=-0.28, P≤0.01) in females (**Figure 5.11A and 5.11C**). In addition, SNPs in *Ptprk* gene were associated with aortic lesion area (17 SNPs), *Lactococcus* abundance (20 SNPs), and *Ptprk* gene expression (12 SNPs) (**Figure 5.15C-5.15E and 5.16**). Especially, wild-derived strains CAST and PWK alleles showed a strong positive association with the *Ptprk* gene expression (**Figure 5.17A**). The 10 common SNPs that associated with aortic lesion area, *Lactococcus* abundance, and *Ptprk* gene expression were attributed to the CAST and PWK haplotypes (**Table 5.14**) and female mice with CAST and PWK haplotypes had significantly lower aortic lesion area and *Lactococcus* abundance, and higher *Ptprk* gene expression than mice with homozygous genotypes (**Figure 5.15C-5.15E**). This is consistent with the result that *Ptprk* gene expression has a negative correlation with both aortic lesion area and *Lactococcus* abundance. Furthermore, *Ptprk* gene expression not only had a strong negative correlation with five candidate EC gene families known to be produced by *Lactococcus lactis* (**Figures 5.18A and Table 5.15),** but also showed a high association with the *Ptprk* SNP in females (**Figure 5.18B**). These data suggest that genetic

variants on the *Ptprk* locus drive changes in hepatic *Ptprk* gene expression, aortic lesion area level, and *Lactococcus* abundance in the feces (**Figure 5.17B**).

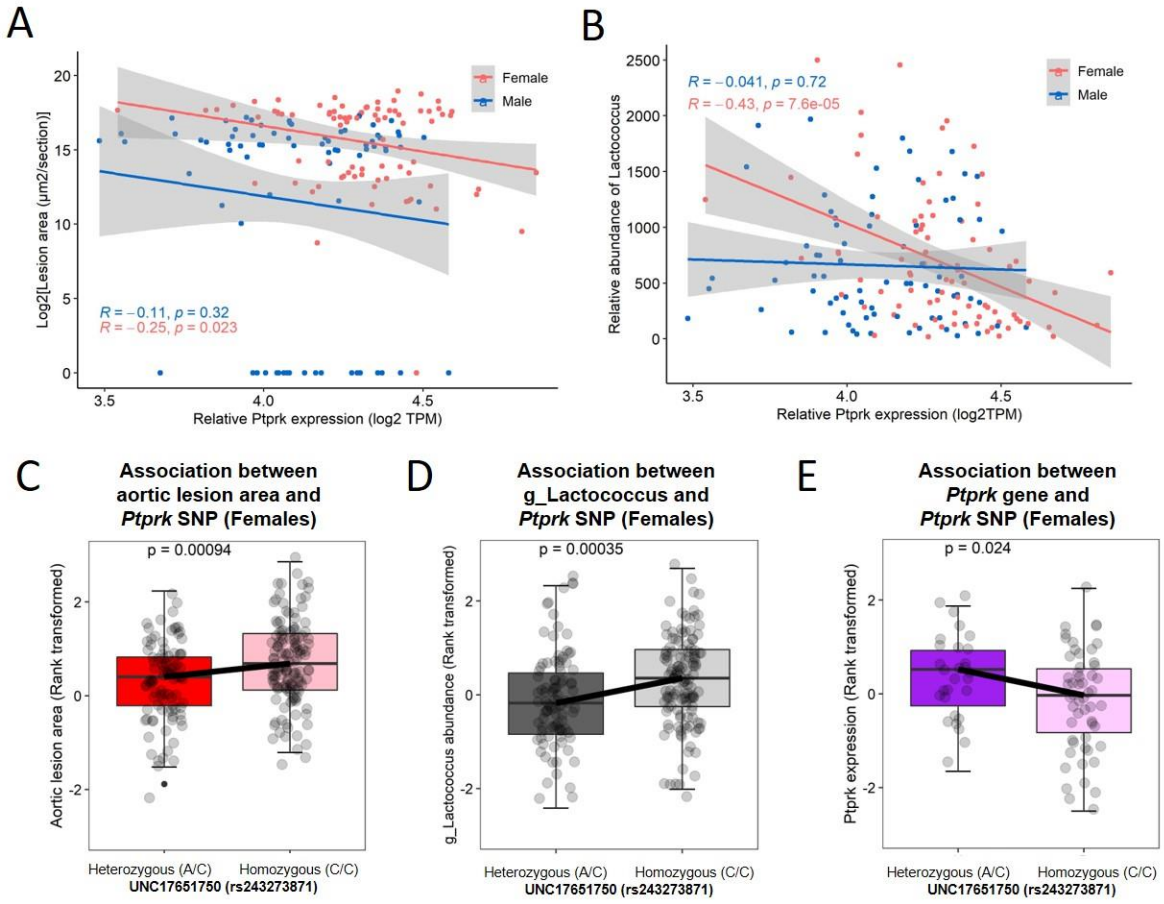
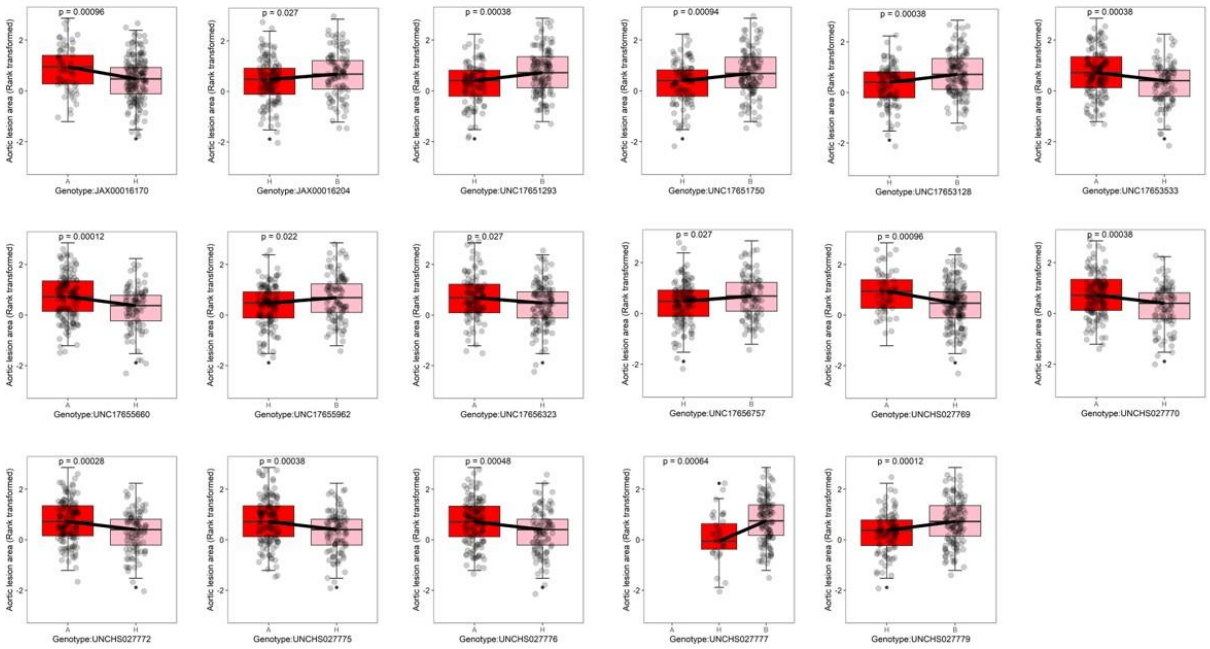
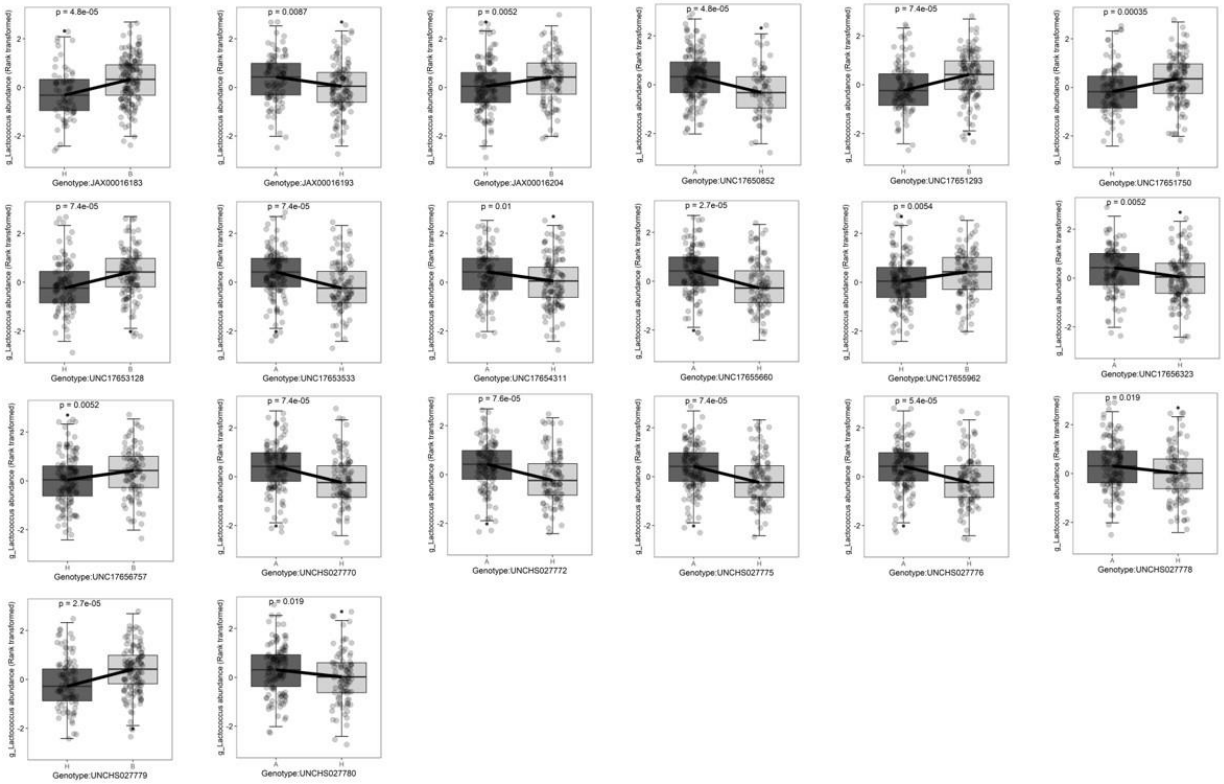


Figure 5.15. Identification of *Ptprk* gene associated with both aortic lesion area and *Lactococcus* abundance in DO-F1 female mice.
(A) *Ptprk* gene expression is negatively associated with aortic lesion area in females (R = -0.25, p = 0.023), not males (R = -0.11, p = 0.32). (B) *Ptprk* gene expression is negatively associated with *Lactococcus* abundance in females (R = -0.43, p = 7.6E-05), not males (R = -0.041, p = 0.72). (C-E) Association between a SNP (rs243273871) in *Ptprk* gene and aortic lesion area (C), *Lactococcus* abundance (D), and *Ptprk* gene expression (E). Heterozygous SNP (A/C) is from wild-type inbred strains (CAST or PWK strains) and homozygous SNP (C/C) is from laboratory inbred strains (A/J, 129, NOD, NZO, or WSB strains).

A **SNP associations between aortic lesion area and 17 SNPs in Ptprk gene (Female mice)**



B **SNP associations between g_Lactococcus and 20 SNPs in Ptprk gene (Female mice)**



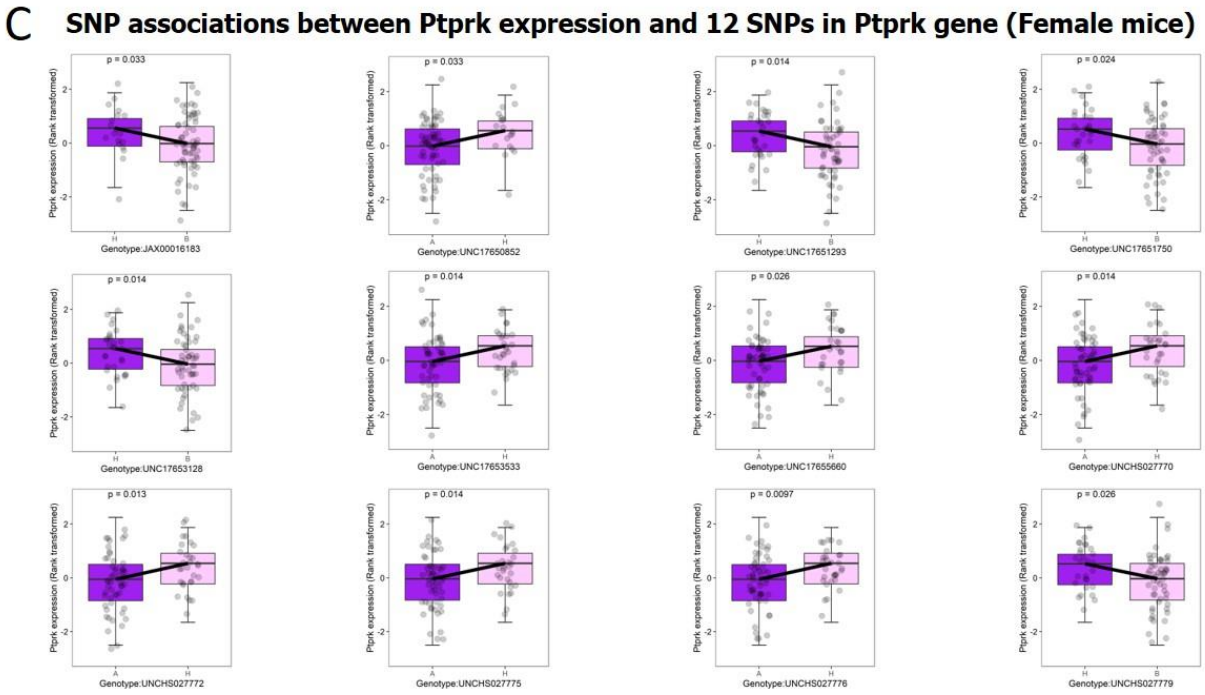


Figure 5.16. Significant associations between SNPs in *Ptprk* gene and aortic lesion area, *Lactococcus* abundance, or *Ptprk* gene expression in DO-F1 female mice.
(A) Significant association between 17 SNPs in *Ptprk* gene and aortic lesion area. (B) A significant association between 20 SNPs in *Ptprk* gene and *Lactococcus* abundance. (C) A significant association between 10 SNPs in *Ptprk* gene and *Ptprk* gene expression. Differences between heterozygous (denoted "H" in x-axis) and homozygous genotype (denoted "A" or "B" in x-axis) were evaluated using an unpaired Wilcoxon signed-rank test.

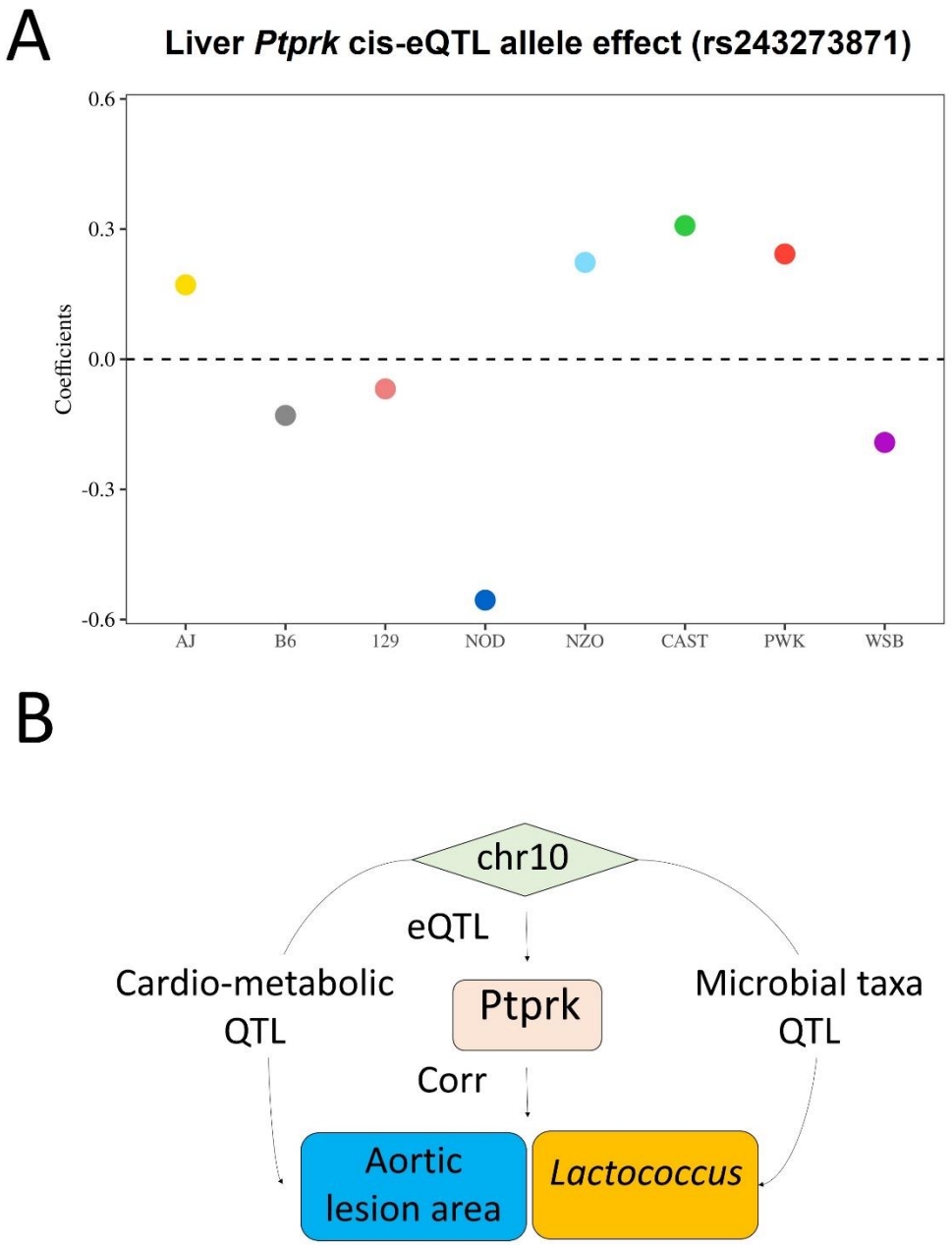


Figure 5.17. Identification of *Ptprk* gene associated with both aortic lesion area and *Lactococcus* abundance in DO-F1 female mice.
(A) Estimated founder strain levels of *Ptprk* gene were inferred in the DO-F1 population from the founder strain coefficients observed at the corresponding QTL on chr 10. B6, C57BL/6J; 129, 129S1/SvlmJ; NOD, NOD/ShiLtJ; NZO, NZO/HILtJ; CAST, CAST/EiJ; PWJ, PWK/PhJ; WSB, WSB/EiJ. (B) Inset summarizes that SNPs on chromosome 10 (chr10) drive variation in *Ptprk* transcript abundance, aortic lesion area, and *Lactococcus* abundance, corroborated by a spearman correlation between hepatic *Ptprk* gene and aortic lesion area or *Lactococcus* abundance.

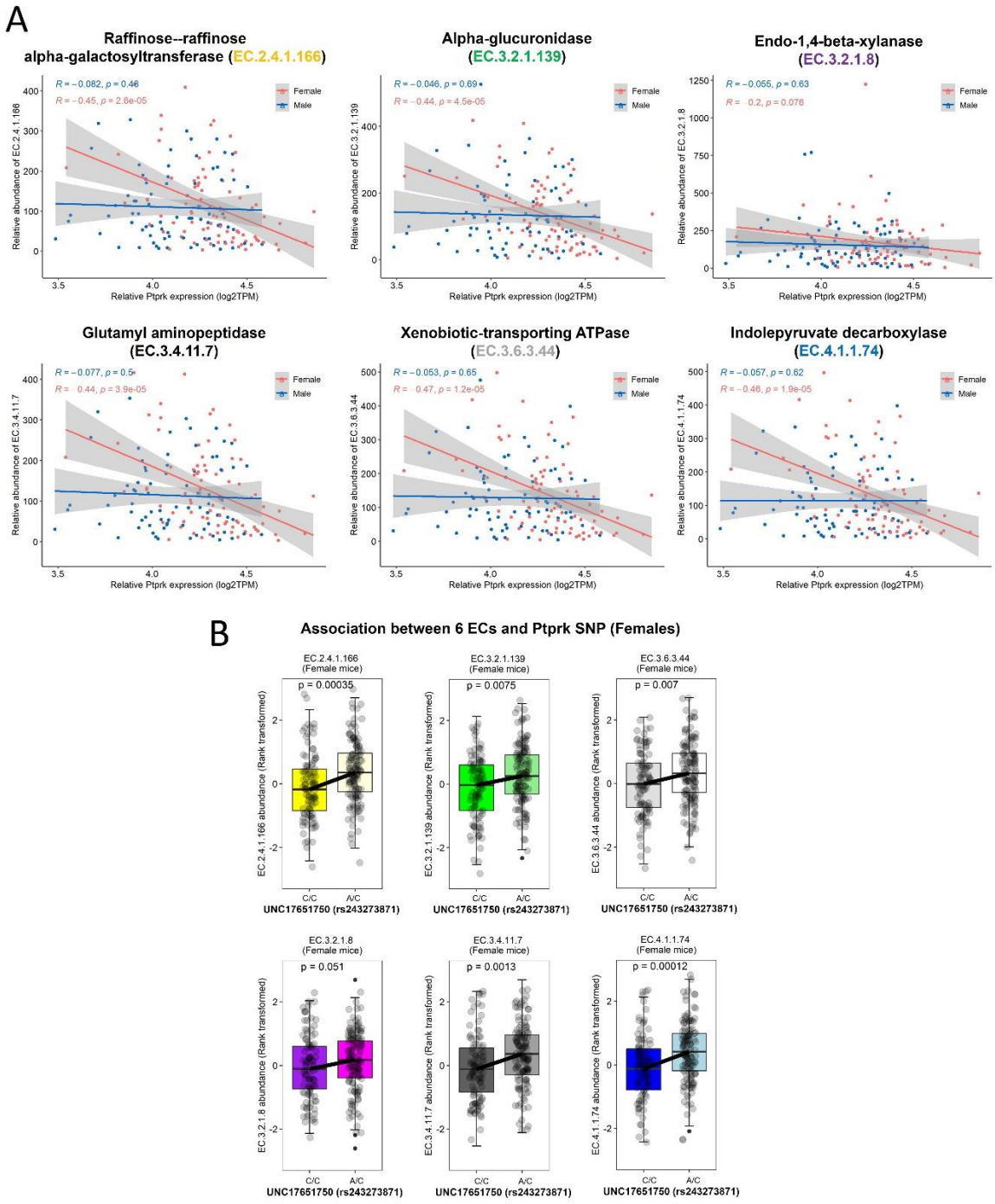


Figure 5.18. Identification of *Ptprk* gene associated with five EC gene families in DO-F1 female mice.

(A) *Ptprk* gene expression is negatively associated with five EC gene families in females, not males. (B) Association between a SNP (rs243273871) in *Ptprk* gene and five EC gene families. Heterozygous SNP (A/C) is from the CAST or PWK strains and homozygous SNP (C/C) is from the A/J, 129, NOD, NZO, or WSB strains.

Table 5.14. List of 10 *Ptprk* SNPs that showed a significant difference between heterozygous and homozygous genotypes.

Maker	Chr	Position (Mbp)	SNP	Strains that have heterozygous genotype	Allele from B6	Allele from strains that have a heterozygous genotype
UNCHS027770	10	28.14	rs29316154	CAST;PWK	C	T
UNC17651750	10	28.22	rs243273871	CAST;PWK	C	A
JAX00016183	10	28.29	rs46801305	CAST	T	G
UNC17653128	10	28.32	rs248898676	CAST;PWK	T	G
UNC17653533	10	28.35	rs52604325	CAST;PWK	G	T
UNCHS027772	10	28.39	rs36371657	CAST;PWK	A	G
UNCHS027775	10	28.40	rs50071536	CAST;PWK;WSB	C	T
UNCHS027776	10	28.41	rs230967592	CAST;PWK	A	G
UNCHS027779	10	28.46	rs238966765	PWK	C	G
UNC17655660	10	28.49	rs225638435	CAST;PWK	C	T

Table 5.15. 6 candidate EC gene families QTLs, liver cis-eQTL, and a direct correlation between the mapped EC and transcript in three models.

Model	EC QTL				cis-eQTL[a]						EC:trait correlation coefficient and p-value			
											All mice		Female mice	
	EC	Chr	Position (Mbp)	LOD	Gene	Chr	Position (Mbp)	LOD	Module color (Females)	Module color (Males)	r	adjusted p-value	r	adjusted p-value
Female	EC.2.4.1.166	10	26.71	7.26	Ptprk	10	28.07	9.20	purple	blue	**-0.24**	**0.00**	**-0.45**	**4.5E-05**
Female	EC.2.4.1.166	10	26.71	7.26	Arhgap18	10	26.77	11.18	darkred	blue	-0.07	0.46	**-0.25**	**0.04**
Female	EC.2.4.1.166	10	26.71	7.26	Akap7	10	25.17	14.64	NA	blue	0.08	0.43	-0.01	0.93
Female	EC.2.4.1.166	10	26.71	7.26	Gm10145	10	28.35	28.33	NA	NA	0.15	0.08	**0.24**	**0.04**
Female	EC.2.4.1.166	10	26.71	7.26	Smlr1	10	25.66	9.74	NA	NA	-0.04	0.68	-0.03	0.84
Female	EC.3.2.1.139	10	26.79	7.39	Ptprk	10	28.07	9.20	purple	blue	**-0.23**	**0.01**	**-0.44**	**7.6E-05**
Female	EC.3.2.1.139	10	26.79	7.39	Arhgap18	10	26.77	11.18	darkred	blue	-0.05	0.58	-0.21	0.09
Female	EC.3.2.1.139	10	26.79	7.39	Akap7	10	25.17	14.64	NA	blue	0.06	0.52	0.01	0.93
Female	EC.3.2.1.139	10	26.79	7.39	Gm10145	10	28.35	28.33	NA	NA	0.15	0.09	0.15	0.22
Female	EC.3.2.1.139	10	26.79	7.39	Smlr1	10	25.66	9.74	NA	NA	-0.07	0.45	-0.05	0.75
Female	EC.3.2.1.8	10	26.71	7.43	Ptprk	10	28.07	9.20	purple	blue	-0.09	0.32	-0.20	0.10
Female	EC.3.2.1.8	10	26.71	7.43	Arhgap18	10	26.77	11.18	darkred	blue	-0.02	0.85	-0.11	0.42
Female	EC.3.2.1.8	10	26.71	7.43	Akap7	10	25.17	14.64	NA	blue	0.06	0.55	-0.07	0.64
Female	EC.3.2.1.8	10	26.71	7.43	Gm10145	10	28.35	28.33	NA	NA	0.10	0.30	0.15	0.22
Female	EC.3.2.1.8	10	26.71	7.43	Smlr1	10	25.66	9.74	NA	NA	0.05	0.60	0.11	0.39
Female	EC.3.4.11.7	10	26.71	7.66	Ptprk	10	28.07	9.20	purple	blue	**-0.23**	**0.01**	**-0.44**	**6.5E-05**
Female	EC.3.4.11.7	10	26.71	7.66	Arhgap18	10	26.77	11.18	darkred	blue	-0.04	0.67	-0.21	0.08
Female	EC.3.4.11.7	10	26.71	7.66	Akap7	10	25.17	14.64	NA	blue	0.12	0.21	0.07	0.64
Female	EC.3.4.11.7	10	26.71	7.66	Gm10145	10	28.35	28.33	NA	NA	0.13	0.16	0.17	0.16
Female	EC.3.4.11.7	10	26.71	7.66	Smlr1	10	25.66	9.74	NA	NA	-0.04	0.68	-0.02	0.93
Female	EC.3.6.3.44	10	26.29	7.45	Ptprk	10	28.07	9.20	purple	blue	**-0.24**	**0.00**	**-0.47**	**2.3E-05**
Female	EC.3.6.3.44	10	26.29	7.45	Arhgap18	10	26.77	11.18	darkred	blue	-0.04	0.65	-0.22	0.06
Female	EC.3.6.3.44	10	26.29	7.45	Akap7	10	25.17	14.64	NA	blue	0.06	0.53	-0.05	0.76
Female	EC.3.6.3.44	10	26.29	7.45	Gm10145	10	28.35	28.33	NA	NA	0.16	0.08	0.20	0.09
Female	EC.3.6.3.44	10	26.29	7.45	Smlr1	10	25.66	9.74	NA	NA	-0.04	0.65	-0.03	0.88
Female	EC.4.1.1.74	10	26.86	7.44	Ptprk	10	28.07	9.20	purple	blue	**-0.22**	**0.01**	**-0.46**	**3.5E-05**
Female	EC.4.1.1.74	10	26.86	7.44	Arhgap18	10	26.77	11.18	darkred	blue	-0.05	0.58	**-0.27**	**0.02**
Female	EC.4.1.1.74	10	26.86	7.44	Akap7	10	25.17	14.64	NA	blue	0.09	0.34	-0.04	0.82
Female	EC.4.1.1.74	10	26.86	7.44	Gm10145	10	28.35	28.33	NA	NA	0.19	0.03	**0.29**	**0.01**
Female	EC.4.1.1.74	10	26.86	7.44	Smlr1	10	25.66	9.74	NA	NA	-0.04	0.67	-0.02	0.92

Significance was considered when QTL results were P<0.05.
[a]cis-eQTLs are the result of the sex additive model.

5.6. Discussion

Here, we characterize the gut microbiota composition and its association with cardiometabolic traits in the hyperlipidemic DO-F1 mice fed a HFHC diet designed to induce atherosclerotic lesions. In the present study, DO-F1 mice were used to generate a systems genetics resource that incorporates cardiometabolic traits, gut microbiome, and liver transcriptome and to assess the effect of genetic variation on trans-omics data. Our study utilizes a unique genetic diversity across the DO-F1 mice with heterozygous genetic architecture, which results in significant variations in the trans-omics data in a set of relatively limited environmental effects. The results have led to several conclusions. First, host genetic factors caused substantial inter-individual variation in the gut microbiota composition in DO-F1 mice. Second, we identified that gut microbial diversity, genera, and functional profiling were associated with key atherosclerotic traits, suggesting that gut microbiota may play a role in atherosclerosis development. Third, we observed sex-specific liver gene networks that were highly associated with atherosclerosis and gut microbiota. Fourth, the integration of trans-omics datasets enabled the investigation of atherosclerosis pathophysiology at the systems level, which has improved our understanding of the complex trait contributing to atherosclerosis development. For example, we identified novel candidate genes for QTL that were previously mapped for atherosclerosis, which is a highly complex trait resulting from the interaction between genetic and environmental factors via high-resolution mapping. We further interrogated this QTL by co-mapping with mbQTLs and liver *cis*-eQTLs, and by investigating the correlation between trans-omics datasets to identify a potential candidate gene for this QTL. Below we discuss each of these points in turn.

The profound effect of genetic background reflects the effectiveness of recent studies showing that gut microbial composition is a complex, polygenic trait. Many studies conducted in DO founder strains have demonstrated the impact of genetic background on gut microbiota (Campbell et al., 2012; Carmody et al., 2015; Kemis et al., 2019). Interestingly, several genetic studies using DO mice or F2 cross mice showed that multiple microbial taxa can be co-localized to a single-genetic locus, suggesting that pleiotropic loci may modulate the abundance of multiple taxa (Benson et al., 2010; Kemis et al., 2019). Our study also revealed significant pleiotropic loci for microbial taxa between Lachnospiraceae NK4A136 group and Lachnospiraceae FCS020 group on chromosome 14 and between *Terrisporobacter* and *Clostridium* sensu stricto 1 on chromosome 16 in entire mice. Furthermore, the pleiotropic loci were more prominent in QTL mapping of MetaCyc pathways and EC gene families predicted from microbial taxa. These results suggest that genetic variation affecting specific microbial taxa influences the gut microbiota composition.

A number of studies in recent years have suggested that certain features in the gut microbiota are associated with metabolic phenotypes (Falony et al., 2016; Manor et al., 2020; Org et al., 2017). However, it is not yet clear whether associations for health status can be seen not only in individual taxa but also in the overall composition of the gut microbiota. Our pairwise correlation and association analysis not only showed a significant association of cardiometabolic traits with the gut microbiota, but also revealed how certain taxa and functional profiling might influence sex differences in atherosclerosis after HFHC diet intake. For example, in DO-F1 mice, *Bifidobacterium*, which was abundant in males, was negatively correlated with atherosclerotic traits (aortic lesion area, plasma TC, and hepatic TC), and female-specific abundant taxa such as *Alistipes* and *Blautia* were negatively correlated with those atherosclerotic

traits at 24 weeks. Several studies have found associations of *Alistipes* and *Blautia* with dysbiosis and gut permeability (Kong et al., 2019; Lam et al., 2012; Parker et al., 2020) which are predominantly manifested in patients with cardiovascular disease (Emoto et al., 2016; Kelly et al., 2016; Koren et al., 2011). In addition, *Blautia* has been reported to be associated with human genetic variants related to obesity and BMI (Bonder et al., 2016). *Bifidobacterium* has also been identified as a sexually dimorphic taxon in many human studies, and is generally more abundant in females with lower levels of metabolic traits (e.g., BMI, LDL-C, TG, insulin, and inflammatory markers) than males (Borgo et al., 2018; Gao et al., 2018; Takagi et al., 2019). Recent studies reveal that *Bifidobacterium* spp. decreases serum TC and LDL-C in hypercholesterolemic patients (Rerksuppaphol and Rerksuppaphol, 2015), prevents bacterial translocation and improves mucosal barrier function in rats by reducing the amount of LPS in the blood (Caplan et al., 1999), and shows a negative correlation with the fibrosis rate in patients with heart disease (Tuomisto et al., 2019), suggesting that it may play a beneficial role in the prevention of coronary atherosclerosis. Therefore, our results, consistent with other recent studies, have shown that microbial taxa are associated with susceptibility to atherosclerosis.

Gene network-based analysis has been used by many groups to uncover the basis of complex diseases (Bennett et al., 2015; Chen et al., 2008; Gargalovic et al., 2006; Miller et al., 2008) and gut microbiota (Fuess et al., 2021; Pan et al., 2021; Pfalzer et al., 2016). Genes with similar expression patterns can form complex pathways or participate in regulatory and signaling circuits (Huang et al., 2007; Ideker et al., 2002). Gene co-expression networks have been extensively studied to describe transcripts in many organisms, including yeast, flies, worms, mice, and humans (Bennett et al., 2015; Stuart et al., 2003). WGCNA is one of the most commonly used analysis tools for this type of analysis (Zhang and Horvath, 2005). WGCNA can

reduce the dimension of large RNA-seq data sets with dozens of modules of genes that share similar expression patterns and the same general function in transcripts. Once identified, the overall behavior and topology of the module can be related to the disease state. Since we observed the sexual dimorphism of cardiometabolic traits and liver transcriptome in Chapter 4, we constructed networks in each sex in order to avoid signal being overwhelmed by sexually dimorphic cardiometabolic traits. Among the sex-specific modules identified in each sex, the purple module enriched for protein ubiquitination in females and the black module enriched for steroid biosynthesis in males not only had negative correlations with aortic lesion area, but also a higher correlation with the microbial diversity, genera, and functional profiling than any other module. For example, *Clostridium sensu* strico 13, abundant in males and the mice with low atherosclerotic traits, showed a positive correlation with both modules, and *Blautia* and *Oscillibacter*, abundant in females and the mice with high atherosclerotic traits, had a negative correlation. These results indicate that co-expressed genes associated with aortic lesion area and gut microbiota showed similar directionalities in both sexes, but were sex-specifically enriched in functional pathways.

To discover individual candidate genes within genomic mapping intervals of trans-omics data, we assessed three additional data sources: correlation, allele effects, and eQTL. We analyzed the correlation between each omics data (traits, microbial taxa, transcripts) and examined whether there is a significant association between each dataset. Since 8 alleles were segregated from the DO-F1 mice, the contribution of each allele to each omics data was assessed. Furthermore, hepatic *cis*-eQTL was screened in which the allele effect pattern is consistent with that of the cardiometabolic trait QTL or mbQTL. For example, this study confirmed that the QTL of *Lactococcus* abundance was co-localized with the aortic lesion area

QTL on chromosome 10. In addition, gene families of six enzymes reported to be produced in *Lactococcus lactis* subsp. *lactis* were also co-localized to the same region (Bolotin et al., 2001; Kelleher et al., 2017; McNulty et al., 2011; Passerini et al., 2013; Siezen et al., 2008). In order to find candidate genes that have a regulatory function for the aortic lesion area and *Lactococcus* abundance, we considered a significant correlation with the aortic lesion area, *Lactococcus* abundance, six EC gene families, hepatic gene modules, and *cis*-eQTL for candidate genes located in these loci, and only *Ptprk* meets all statistical criteria. This gene encodes protein tyrosine phosphatases (PTPs) and PTPs are important modulators of cellular processes such as migration, proliferation, and differentiation, which are involved in pathological vascular wall function (Kappert et al., 2005). *Ptprk* has previously been shown to associate with colorectal cancer (Yan et al., 2020), angiogenesis (Sun et al., 2017), and cell-cell contacts in epithelial cells (Fearnley et al., 2019); however, other biological functions are unclear. Human genome-wide association studies have shown that *Ptprk* SNPs are associated with athletic performance and risk of Celiac disease (Bondar et al., 2014; Pickering et al., 2019). Furthermore, data from the Phenotype-Genotype Integrator in NCBI (https://www.ncbi.nlm.nih.gov/gap/phegeni) demonstrate that *Ptprk* is associated with LDL-C, platelet function tests, and inflammatory bowel disease across approximately 500,000 individuals (**Table 5.16**).

In conclusion, the integration of trans-omics datasets in a hyperlipidemic DO-F1 mouse population provides a powerful tool for the identification of microbial taxa, genes, and biological pathways involved in complex atherosclerotic traits. Specifically, we identified *Ptprk* gene as a regulator of atherosclerosis and *Lactococcus* abundance. Future validation studies are required to investigate how a perturbation of *Ptprk* expression affects aortic lesion area and the consequences on *Lactococcus* colonization in the hyperlipidemic mouse model. Finally, our

study shows the power of systems genetics to identify novel interactions between cardiometabolic traits and gut microbial taxa, and trans-omics datasets and methodologies may help complement and validate existing and future discovery resources.

Table 5.16. Phenome wide association (PheWAS) with Ptprk gene that identified in aortic lesion area and Lactococcus abundance QTLs in DO-F1 female mice.

Gene (Mouse)	Chr (Mouse)	Position (bp, Mouse)	Phenotypes (Human)	CVD or gut microbiota related phenotypes	SNP_rs (Human)	Gene (Human)	Chr (Human)	Position (bp, Human)	P_Value ($p<1\times10^{-7}$)
Ptprk	10	28074820	Celiac Disease	Yes	802734	PTPRK	6	127957653	3.00E-14
Ptprk	10	28074820	Multiple Sclerosis	No	802734	PTPRK	6	127957653	6.00E-09
Ptprk	10	28074820	Crohn Disease	Yes	9491891	PTPRK	6	127956006	3.00E-16
Ptprk	10	28074820	Crohn Disease	Yes	13204742	PTPRK	6	127924620	8.00E-15
Ptprk	10	28074820	Inflammatory Bowel Diseases	Yes	13204742	PTPRK	6	127924620	5.00E-10
Ptprk	10	28074820	Menarche	No	6938574	PTPRK	6	128069835	2.00E-09
Ptprk	10	28074820	Cholesterol, LDL	Yes	7755473	PTPRK	6	128232071	4.75E-08
Ptprk	10	28074820	Platelet Function Tests	Yes	9375562	PTPRK	6	128354763	5.81E-09
Ptprk	10	28074820	Platelet Function Tests	Yes	17055564	PTPRK	6	128290066	1.82E-14

5.7. Limitations of Study

Our study has several limitations. The transcriptomic analysis was limited to liver tissue and some major tissues (e.g. aorta and adipose tissue) in atherosclerosis development have not been tested. The genes responsible for atherosclerosis may not be expressed in the liver tissue we studied, or they may not show any expression differences or eQTLs. Secondly, the amplicon-based prediction of the function of gut microbial taxa using PICRUSt2 may have a high correlation with the functional profile based on shotgun metagenomics sequencing data, but the actual functions found to be significantly different between groups can substantially differ. Lastly, a functional validation study targeting putative gene or taxon was missing.

5.8. Acknowledgments

Mice carrying the human ApoE3 Leiden variant were kindly provided by Dr. Lusis, University of California, Los Angeles, CA, USA. This research was supported in part by NIH grant 5R01HL128572 (B.J.B), USDA project 2032-51530-025-00D (B.J.B).

5.9. Declaration of Interests

The authors declare no conflicts of interest.

5.10. References

AlSiraj, Y., Chen, X., Thatcher, S.E., Temel, R.E., Cai, L., Blalock, E., Katz, W., Ali, H.M., Petriello, M., Deng, P., et al. (2019). XX sex chromosome complement promotes atherosclerosis in mice. Nat Commun *10*, 2631.

Barbera, P., Kozlov, A.M., Czech, L., Morel, B., Darriba, D., Flouri, T., and Stamatakis, A. (2019). EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences. Syst Biol *68*, 365-369.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - A Practical and Powerful Approach to Multiple Testing. J. R. Statst. Soc. B *57*, 289-300.

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

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

Bis, J.C., Kavousi, M., Franceschini, N., Isaacs, A., Abecasis, G.R., Schminke, U., Post, W.S., Smith, A.V., Cupples, L.A., Markus, H.S., et al. (2011). Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. Nat Genet *43*, 940-947.

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

Bolotin, A., Wincker, P., Mauger, S., Jaillon, O., Malarme, K., Weissenbach, J., Ehrlich, S.D., and Sorokin, A. (2001). The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403. Genome Res *11*, 731-753.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., and Asnicar, F. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature biotechnology *37*, 852-857.

Bondar, C., Plaza-Izurieta, L., Fernandez-Jimenez, N., Irastorza, I., Withoff, S., Wijmenga, C., Chirdo, F., Bilbao, J.R., and Cegec (2014). THEMIS and PTPRK in celiac intestinal mucosa: coexpression in disease and after in vitro gliadin challenge. Eur J Hum Genet *22*, 358-362.

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

Borgo, F., Garbossa, S., Riva, A., Severgnini, M., Luigiano, C., Benetti, A., Pontiroli, A.E., Morace, G., and Borghi, E. (2018). Body Mass Index and Sex Affect Diverse Microbial Niches within the Gut. Front Microbiol *9*, 213.

Broman, K.W. (2012a). Genotype probabilities at intermediate generations in the construction of recombinant inbred lines. Genetics *190*, 403-412.

Broman, K.W. (2012b). Haplotype probabilities in advanced intercross populations. G3 (Bethesda) *2*, 199-202.

Broman, K.W., Gatti, D.M., Simecek, P., Furlotte, N.A., Prins, P., Sen, Ś., Yandell, B.S., and Churchill, G.A. (2019). R/qtl2 - Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multi-parent Populations. Genetics *211*, 495-502.

Brown, J.M., and Hazen, S.L. (2018). Microbial modulation of cardiovascular disease. Nat Rev Microbiol *16*, 171-181.

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

Campbell, J.H., Foster, C.M., Vishnivetskaya, T., Campbell, A.G., Yang, Z.K., Wymore, A., Palumbo, A.V., Chesler, E.J., and Podar, M. (2012). Host genetic and environmental effects on mouse intestinal microbiota. ISME J *6*, 2033-2044.

Caplan, M.S., Miller-Catchpole, R., Kaup, S., Russell, T., Lickerman, M., Amer, M., Xiao, Y., and Thomson, R.J. (1999). Bifidobacterial supplementation reduces the incidence of necrotizing enterocolitis in a neonatal rat model. Gastroenterology *117*, 577-583.

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

Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C.A., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A., et al. (2016). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res *44*, D471-480.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr - interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *15*, 128.

Chen, Y., Zhu, J., Lum, P.Y., Yang, X., Pinto, S., MacNeil, D.J., Zhang, C., Lamb, J., Edwards, S., Sieberts, S.K., et al. (2008). Variations in DNA elucidate molecular networks that cause disease. Nature *452*, 429-435.

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

Czech, L., and Stamatakis, A. (2019). Scalable methods for analyzing and visualizing phylogenetic placement of metagenomic samples. PLoS One *14*, e0217050.

Davenport, E.R., Mizrahi-Man, O., Michelini, K., Barreiro, L.B., Ober, C., and Gilad, Y. (2014). Seasonal variation in human gut microbiome composition. PLoS One *9*, e90731.

Doerge, R.W., and Churchill, G.A. (1996). Permutation tests for multiple loci affecting a quantitative character. Genetics *142*, 285-294.

Douglas, G.M., Maffei, V.J., Zaneveld, J.R., Yurgel, S.N., Brown, J.R., Taylor, C.M., Huttenhower, C., and Langille, M.G.I. (2020). PICRUSt2 for prediction of metagenome functions. Nat Biotechnol *38*, 685-688.

Emoto, T., Yamashita, T., Sasaki, N., Hirota, Y., Hayashi, T., So, A., Kasahara, K., Yodoi, K., Matsumoto, T., Mizoguchi, T., et al. (2016). Analysis of Gut Microbiota in Coronary Artery Disease Patients: a Possible Link between Gut Microbiota and Coronary Artery Disease. J Atheroscler Thromb *23*, 908-921.

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

Fan, Y., and Pedersen, O. (2021). Gut microbiota in human metabolic health and disease. Nat Rev Microbiol *19*, 55-71.

Fearnley, G.W., Young, K.A., Edgar, J.R., Antrobus, R., Hay, I.M., Liang, W.C., Martinez-Martin, N., Lin, W., Deane, J.E., and Sharpe, H.J. (2019). The homophilic receptor PTPRK selectively dephosphorylates multiple junctional regulators to promote cell-cell adhesion. Elife *8*.

Franceschini, N., Giambartolomei, C., de Vries, P.S., Finan, C., Bis, J.C., Huntley, R.P., Lovering, R.C., Tajuddin, S.M., Winkler, T.W., Graff, M., et al. (2018). GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. Nat Commun *9*, 5141.

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

Fuess, L.E., den Haan, S., Ling, F., Weber, J.N., Steinel, N.C., and Bolnick, D.I. (2021). Immune Gene Expression Covaries with Gut Microbiome Composition in Stickleback. mBio *12*.

Gaeta, G. (2007). Heritability of atherosclerosis. Am J Hypertens *20*, 255.

Gao, X., Zhang, M., Xue, J., Huang, J., Zhuang, R., Zhou, X., Zhang, H., Fu, Q., and Hao, Y. (2018). Body Mass Index Differences in the Gut Microbiota Are Gender Specific. Front Microbiol *9*, 1250.

Gargalovic, P.S., Imura, M., Zhang, B., Gharavi, N.M., Clark, M.J., Pagnon, J., Yang, W.P., He, A., Truong, A., Patel, S., et al. (2006). Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids. Proc Natl Acad Sci U S A. *103*, 12741-12746.

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

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

Hartiala, J., Schwartzman, W.S., Gabbay, J., Ghazalpour, A., Bennett, B.J., and Allayee, H. (2017). The Genetic Architecture of Coronary Artery Disease: Current Knowledge and Future Opportunities. Curr Atheroscler Rep *19*, 6.

Hiratsuka, T., Furihata, K., Ishikawa, J., Yamashita, H., Itoh, N., Seto, H., and Dairi, T. (2008). An alternative menaquinone biosynthetic pathway operating in microorganisms. Science *321*, 1670-1673.

Huang, Y., Li, H., Hu, H., Yan, X., Waterman, M.S., Huang, H., and Zhou, X.J. (2007). Systematic discovery of functional modules and context-specific functional annotation of human genome. Bioinformatics *23*, i222-229.

Ideker, T., Ozier, O., Schwikowski, B., and Siegel, A.F. (2002). Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics. *18*, S233-240.

Jonsson, A.L., and Backhed, F. (2017). Role of gut microbiota in atherosclerosis. Nat Rev Cardiol *14*, 79-87.

Joost, H.G., and Schurmann, A. (2014). The genetic basis of obesity-associated type 2 diabetes (diabesity) in polygenic mouse models. Mamm Genome *25*, 401-412.

Kappert, K., Peters, K.G., Bohmer, F.D., and Ostman, A. (2005). Tyrosine phosphatases in vessel wall signaling. Cardiovasc Res *65*, 587-598.
Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol *30*, 772-780.
Kaul, A., Mandal, S., Davidov, O., and Peddada, S.D. (2017). Analysis of Microbiome Data in the Presence of Excess Zeros. Front Microbiol *8*, 2114.
Kelleher, P., Bottacini, F., Mahony, J., Kilcawley, K.N., and van Sinderen, D. (2017). Comparative and functional genomics of the Lactococcus lactis taxon; insights into evolution and niche adaptation. BMC Genomics *18*, 267.
Keller, M.P., Gatti, D.M., Schueler, K.L., Rabaglia, M.E., Stapleton, D.S., Simecek, P., Vincent, M., Allen, S., Broman, A.T., Bacher, R., et al. (2018). Genetic Drivers of Pancreatic Islet Function. Genetics *209*, 335-356.
Kelly, T.N., Bazzano, L.A., Ajami, N.J., He, H., Zhao, J., Petrosino, J.F., Correa, A., and He, J. (2016). Gut Microbiome Associates With Lifetime Cardiovascular Disease Risk Profile Among Bogalusa Heart Study Participants. Circ Res *119*, 956-964.
Kemis, J.H., Linke, V., Barrett, K.L., Boehm, F.J., Traeger, L.L., Keller, M.P., Rabaglia, M.E., Schueler, K.L., Stapleton, D.S., Gatti, D.M., et al. (2019). Genetic determinants of gut microbiota composition and bile acid profiles in mice. PLoS Genet *15*, e1008073.
Khera, A.V., and Kathiresan, S. (2017). Genetics of coronary artery disease: discovery, biology and clinical translation. Nat Rev Genet *18*, 331-344.
Khyzha, N., Alizada, A., Wilson, M.D., and Fish, J.E. (2017). Epigenetics of Atherosclerosis: Emerging Mechanisms and Methods. Trends Mol Med *23*, 332-347.
Kim, M., Huda, M.N., O'Connor, A., Albright, J., Durbin-Johnson, B., and Bennett, B.J. (2021). Hepatic transcriptional profile reveals the role of diet and genetic backgrounds on metabolic traits in female progenitor strains of the Collaborative Cross. Physiol Genomics *53*, 173-192.
Kolde, R., Franzosa, E.A., Rahnavard, G., Hall, A.B., Vlamakis, H., Stevens, C., Daly, M.J., Xavier, R.J., and Huttenhower, C. (2018). Host genetic variation and its microbiome interactions within the Human Microbiome Project. Genome Med *10*, 6.
Kong, C., Gao, R., Yan, X., Huang, L., and Qin, H. (2019). Probiotics improve gut microbiota dysbiosis in obese mice fed a high-fat or high-sucrose diet. Nutrition *60*, 175-184.
Koren, O., Spor, A., Felin, J., Fak, F., Stombaugh, J., Tremaroli, V., Behre, C.J., Knight, R., Fagerberg, B., Ley, R.E., et al. (2011). Human oral, gut, and plaque microbiota in patients with atherosclerosis. Proc Natl Acad Sci U S A *108 Suppl 1*, 4592-4598.
Kurilshikov, A., Wijmenga, C., Fu, J., and Zhernakova, A. (2017). Host Genetics and Gut Microbiome: Challenges and Perspectives. Trends Immunol *38*, 633-647.
Lam, Y.Y., Ha, C.W., Campbell, C.R., Mitchell, A.J., Dinudom, A., Oscarsson, J., Cook, D.I., Hunt, N.H., Caterson, I.D., Holmes, A.J., et al. (2012). Increased gut permeability and microbiota change associate with mesenteric fat inflammation and metabolic dysfunction in diet-induced obese mice. PLoS One *7*, e34233.
Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.
Li, J., Zhao, F., Wang, Y., Chen, J., Tao, J., Tian, G., Wu, S., Liu, W., Cui, Q., Geng, B., et al. (2017). Gut microbiota dysbiosis contributes to the development of hypertension. Microbiome *5*, 14.

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

Luedde, M., Winkler, T., Heinsen, F.A., Ruhlemann, M.C., Spehlmann, M.E., Bajrovic, A., Lieb, W., Franke, A., Ott, S.J., and Frey, N. (2017). Heart failure is associated with depletion of core intestinal microbiota. ESC Heart Fail *4*, 282-290.

Lusis, A.J., Seldin, M.M., Allayee, H., Bennett, B.J., Civelek, M., Davis, R.C., Eskin, E., Farber, C.R., Hui, S., Mehrabian, M., et al. (2016). The Hybrid Mouse Diversity Panel: a resource for systems genetics analyses of metabolic and cardiovascular traits. J Lipid Res *57*, 925-942.

Lynn, D.J., Benson, S.C., Lynn, M.A., and Pulendran, B. (2021). Modulation of immune responses to vaccination by the microbiota: implications and potential mechanisms. Nat Rev Immunol.

Manor, O., Dai, C.L., Kornilov, S.A., Smith, B., Price, N.D., Lovejoy, J.C., Gibbons, S.M., and Magis, A.T. (2020). Health and disease markers correlate with gut microbiome composition across thousands of people. Nat Commun *11*, 5206.

Marenberg, M.E., Risch, N., Berkman, L.F., Floderus, B., and de Faire, U. (1994). Genetic susceptibility to death from coronary heart disease in a study of twins. N Engl J Med. *330*, 1041-1046.

Marian, A.J. (2012). The enigma of genetics etiology of atherosclerosis in the post-GWAS era. Curr Atheroscler Rep *14*, 295-299.

Mathes, W.F., Aylor, D.L., Miller, D.R., Churchill, G.A., Chesler, E.J., de Villena, F.P., Threadgill, D.W., and Pomp, D. (2011). Architecture of energy balance traits in emerging lines of the Collaborative Cross. Am J Physiol Endocrinol Metab *300*, E1124-1134.

Mayer, B., Erdmann, J., and Schunkert, H. (2007). Genetics and heritability of coronary artery disease and myocardial infarction. Clin Res Cardiol *96*, 1-7.

McArdle, B.H., and Anderson, M.J. (2001). FITTING MULTIVARIATE MODELS TO COMMUNITY DATA - A COMMENT ON DISTANCE-BASED REDUNDANCY ANALYSIS. Ecology *82*, 290-297.

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

McNulty, N.P., Yatsunenko, T., Hsiao, A., Faith, J.J., Muegge, B.D., Goodman, A.L., Henrissat, B., Oozeer, R., Cools-Portier, S., Gobert, G., et al. (2011). The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. Sci Transl Med. *3*, 106ra106.

Miller, J.A., Oldham, M.C., and Geschwind, D.H. (2008). A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. J Neurosci *28*, 1410-1420.

Nazmul Huda, M., Winnike, J.H., Crowell, J.M., O'Connor, A., and Bennett, B.J. (2020). Microbial modulation of host body composition and plasma metabolic profile. Sci Rep *10*, 6545.

Needham, B.D., Kaddurah-Daouk, R., and Mazmanian, S.K. (2020). Gut microbial molecules in behavioural and neurodegenerative conditions. Nat Rev Neurosci *21*, 717-731.

Nelson, C.P., Goel, A., Butterworth, A.S., Kanoni, S., Webb, T.R., Marouli, E., Zeng, L., Ntalla, I., Lai, F.Y., Hopewell, J.C., et al. (2017). Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nat Genet *49*, 1385-1391.

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

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., et al. (2018). vegan: Community Ecology Package. R package version 2.5-2. https://CRAN.R-project.org/package=vegan.

Org, E., Blum, Y., Kasela, S., Mehrabian, M., Kuusisto, J., Kangas, A.J., Soininen, P., Wang, Z., Ala-Korpela, M., Hazen, S.L., et al. (2017). Relationships between gut microbiota, plasma metabolites, and metabolic syndrome traits in the METSIM cohort. Genome Biol *18*, 70.

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

Pan, X., Li, Z., Li, B., Zhao, C., Wang, Y., Chen, Y., and Jiang, Y. (2021). Dynamics of rumen gene expression, microbiome colonization, and their interplay in goats. BMC Genomics *22*, 288.

Parker, B.J., Wearsch, P.A., Veloo, A.C.M., and Rodriguez-Palacios, A. (2020). The Genus Alistipes: Gut Bacteria With Emerging Implications to Inflammation, Cancer, and Mental Health. Front Immunol *11*, 906.

Passerini, D., Coddeville, M., Le Bourgeois, P., Loubiere, P., Ritzenthaler, P., Fontagne-Faucher, C., Daveran-Mingot, M.L., and Cocaign-Bousquet, M. (2013). The carbohydrate metabolism signature of lactococcus lactis strain A12 reveals its sourdough ecosystem origin. Appl Environ Microbiol *79*, 5844-5852.

Pfalzer, A.C., Kamanu, F.K., Parnell, L.D., Tai, A.K., Liu, Z., Mason, J.B., and Crott, J.W. (2016). Interactions between the colonic transcriptome, metabolome, and microbiome in mouse models of obesity-induced intestinal cancer. Physiol Genomics *48*, 545-553.

Pickering, C., Suraci, B., Semenova, E.A., Boulygina, E.A., Kostryukova, E.S., Kulemin, N.A., Borisov, O.V., Khabibova, S.A., Larin, A.K., Pavlenko, A.V., et al. (2019). A Genome-Wide Association Study of Sprint Performance in Elite Youth Football Players. J Strength Cond Res. *33*, 2344-2351.

Pletscher-Frankild, S., Palleja, A., Tsafou, K., Binder, J.X., and Jensen, L.J. (2015). DISEASES: text mining and data integration of disease-gene associations. Methods *74*, 83-89.

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

Reed, D.R., Bachmanov, A.A., and Tordoff, M.G. (2007). Forty mouse strain survey of body composition. Physiol Behav *91*, 593-600.

Rerksuppaphol, S., and Rerksuppaphol, L. (2015). A Randomized Double-blind Controlled Trial of Lactobacillus acidophilus Plus Bifidobacterium bifidum versus Placebo in Patients with Hypercholesterolemia. J Clin Diagn Res *9*, KC01-04.

Sasidhar, M.V., Reddy, S., Naik, A., and Naik, S. (2014). Genetics of coronary artery disease - a clinician's perspective. Indian Heart J *66*, 663-671.

Saul, M.C., Philip, V.M., Reinholdt, L.G., Center for Systems Neurogenetics of, A., and Chesler, E.J. (2019). High-Diversity Mouse Populations for Complex Traits. Trends Genet *35*, 501-514.

Sen, S., and Churchill, G.A. (2001). A statistical framework for quantitative trait mapping. Genetics *159*, 371-387.

Siezen, R.J., Starrenburg, M.J., Boekhorst, J., Renckens, B., Molenaar, D., and van Hylckama Vlieg, J.E. (2008). Genome-scale genotype-phenotype matching of two Lactococcus lactis isolates from plants identifies mechanisms of adaptation to the plant niche. Appl Environ Microbiol *74*, 424-436.

Smallwood, T.L., Gatti, D.M., Quizon, P., Weinstock, G.M., Jung, K.C., Zhao, L., Hua, K., Pomp, D., and Bennett, B.J. (2014). High-resolution genetic mapping in the diversity outbred mouse population identifies Apobec1 as a candidate gene for atherosclerosis. G3 (Bethesda) *4*, 2353-2363.

Stuart, J.M., Segal, E., Koller, D., and Kim, S.K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. Science *302*, 249-255.

Sun, P.H., Chen, G., Mason, M., Jiang, W.G., and Ye, L. (2017). Dual roles of protein tyrosine phosphatase kappa in coordinating angiogenesis induced by pro-angiogenic factors. Int J Oncol *50*, 1127-1135.

Svenson, K.L., Von Smith, R., Magnani, P.A., Suetin, H.R., Paigen, B., Naggert, J.K., Li, R., Churchill, G.A., and Peters, L.L. (2007). Multiple trait measurements in 43 inbred mouse strains capture the phenotypic diversity characteristic of human populations. J Appl Physiol (1985) *102*, 2369-2378.

Tabrett, A., and Horton, M.W. (2020). The influence of host genetics on the microbiome. F1000Res *9*.

Takagi, T., Naito, Y., Inoue, R., Kashiwagi, S., Uchiyama, K., Mizushima, K., Tsuchiya, S., Dohi, O., Yoshida, N., Kamada, K., et al. (2019). Differences in gut microbiota associated with age, sex, and stool consistency in healthy Japanese subjects. J Gastroenterol *54*, 53-63.

Tuomisto, S., Huhtala, H., Martiskainen, M., Goebeler, S., Lehtimaki, T., and Karhunen, P.J. (2019). Age-dependent association of gut bacteria with coronary atherosclerosis: Tampere Sudden Death Study. PLoS One *14*, e0221345.

Turpin, W., Espin-Garcia, O., Xu, W., Silverberg, M.S., Kevans, D., Smith, M.I., Guttman, D.S., Griffiths, A., Panaccione, R., Otley, A., et al. (2016). Association of host genome with intestinal microbial composition in a large healthy cohort. Nat Genet *48*, 1413-1417.

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

Watkins, H., and Farrall, M. (2006). Genetic susceptibility to coronary artery disease: from promise to progress. Nat Rev Genet *7*, 163-173.

Wu, J.H.Y., Micha, R., and Mozaffarian, D. (2019). Dietary fats and cardiometabolic disease: mechanisms and effects on risk factors and outcomes. Nat Rev Cardiol *16*, 581-601.

Yalcin, B., Wong, K., Agam, A., Goodson, M., Keane, T.M., Gan, X., Nellaker, C., Goodstadt, L., Nicod, J., Bhomra, A., et al. (2011). Sequence-based characterization of structural variation in the mouse genome. Nature *477*, 326-329.

Yan, H.H.N., Siu, H.C., Ho, S.L., Yue, S.S.K., Gao, Y., Tsui, W.Y., Chan, D., Chan, A.S., Wong, J.W.H., Man, A.H.Y., et al. (2020). Organoid cultures of early-onset colorectal cancers reveal distinct and rare genetic profiles. Gut.

Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M., and Price, A.L. (2014). Advantages and pitfalls in the application of mixed-model association methods. Nat Genet *46*, 100-106.
Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., and Glöckner, F.O. (2014). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Research *42*, D643-D648.
Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol *4*, Article17.
Ziyatdinov, A., Vazquez-Santiago, M., Brunel, H., Martinez-Perez, A., Aschard, H., and Soria, J.M. (2018). lme4qtl: linear mixed models with flexible covariance structure for genetic studies of related individuals. BMC Bioinformatics *19*, 68.

CHAPTER 6.

Conclusion and Future Direction

This dissertation focuses on improving the current understanding of the potential mechanisms and therapeutic targets on the underlying genetic process of atherosclerosis, the effect of liver transcriptome on metabolic dysfunctions such as atherosclerosis and dyslipidemia depending on diet and genetic backgrounds, and the relationship between gut microbiota and atherosclerosis. In Chapter 2, we examined the associations of gut microbiota and microbial metabolites with atherosclerosis, coronary artery disease, blood pressure, and type 2 diabetes in human and rodent models. In Chapter 3, we assessed the effects of atherogenic diet and host genetics on the liver transcriptome, and focused on the *Nox4* gene, which was associated with TMAO and liver triglyceride, which are risk factors of atherosclerosis and dyslipidemia. In Chapter 4, we utilized state-of-the-art mouse resources and bioinformatics tools to uncover sex-specific liver genetic pathways, gene networks, and genetic loci associated with atherosclerosis in a population of hyperlipidemic DO-F1 mice. Finally, in Chapter 5, fecal microbial taxa associated with atherosclerotic traits were identified, and trans-omics analysis confirmed that aortic lesion area and *Lactococcus* abundance colocalized at the same loci.

The studies presented in this dissertation suggest that complex traits related to atherosclerosis and CVD susceptibility are associated with the liver transcriptome and gut microbiota, and reveal novel findings of genes and microbial taxa that contribute to the genetic mapping through integrative genetic analysis with those omics data. This is the first study to characterize atherosclerosis with multi-omics approaches in DO-F1 mice. Follow-up studies

could be performed to reliably identify the biological relevance of genes and genetic variants identified by co-mapping between cardiometabolic traits, liver transcriptome, or microbial taxa. I will discuss the proposed experiments in turn below.

6.1. Functional Significance of the CAST/EiJ and PWK/PhJ Alleles of *Ptprk* and *Pten* Genes in Atherosclerosis in a Sex-specific Manner

We revealed sex-specific loci associated with atherosclerosis on chromosome 10 (LOD: 7.3, BCI: 22.90-30.75 Mbp, 33 genes within the BCI) and chromosome 19 (LOD: 7.92, BCI: 32.00-40.23 Mbp, 58 genes within the BCI) in DO-F1 mice. Among the genes located within the BCI of each aortic lesion area QTL, *Ptprk* (females) and *Pten* (males) genes that showed significant *cis*-eQTLs and correlated with the aortic lesion area were identified in each sex. In both aortic lesion area QTLs and *cis*-eQTLs for both genes, alleles from CAST/EiJ and PWK/PhJ were found to be associated with lesion size. This suggests that differential expression of these genes in each sex may contribute to differences in lesion size. Based on these results, we can compare the differences of expression in the *Ptprk* and *Pten* genes in each sex between hyperlipidemic CAST/EiJ, PWK/PhJ, and C57BL/6J or from mice with CAST/EiJ and PWK/PhJ allele contributions in the aortic lesion area QTLs compared to those that have contribution from the other founders. To do this, we can validate our findings with the Collaborative Cross (CC) mouse model. DO and CC mice are complementary resources and each DO mouse has a unique heterozygous genome, whereas CC mice that have a homozygous genome are maintained as reproducible strains. Therefore, it is possible to identify the difference between the two genes and aortic lesion area in each sex by obtaining CC lines with CAST/EiJ and PWK/PhJ alleles on the aortic lesion area QTLs identified in DO-F1 mice. Based on the results of our study, the hyperlipidemic CC line with the CAST/EiJ and PWK/PhJ alleles in *Ptprk cis*-eQTL may show

small lesions in females in response to the atherogenic diet, and the hyperlipidemic CC line with the CAST/EiJ and PWK/PhJ alleles in *Pten cis*-eQTL could show large lesions in males.

If the results of these proposed studies suggest that CAST/EiJ and PWK/PhJ mice have causal genetic variants in the *Ptprk* and *Pten* genes that regulate atherosclerosis susceptibility, we will conduct *in vitro* and *in vivo* studies to determine the functional role of causal mutations in both genes. First, we can confirm the alterations in expression of two genes by inducing inflammatory conditions in macrophages and hepatocytes, or measure intracellular lipid accumulation, inflammatory cytokines, and cholesterol efflux after knockdown or overexpression of each gene. In *in vivo* models, we can also compare the levels of atherosclerotic traits in C57BL/6J, $Ptprk^{-/-}$, and $Pten^{-/-}$ mice treated with an atherogenic diet in both sexes to assess the sex-specific effect of the two genes on atherosclerosis. In addition, CRISPR can be used to affect protein function by targeting missense variants or SNPs that were associated with atherosclerosis. Alternatively, the levels of atherosclerotic traits between the CAST, PWK, and C57BL/6J can be compared for each sex after knockdown or overexpression of each gene by tail vein injection of an adenovirus construct. The results of the above study will certainly confirm the functional roles of the CAST/EiJ and PWK/PhJ allelic contributions to candidate genes and provide a basis for understanding the biological mechanisms by which these strains may contribute to atherosclerosis susceptibility.

6.2. Functional Validation of *Lactococcus* as a Candidate Taxon Regulating Atherosclerosis

In the current study, we used four different models for QTL mapping such as sex additive (accounting for sex as an additive covariate), sex interactive (accounting for sex as an interactive covariate), females, and males to identify genetically regulated co-localized loci between atherosclerosis and microbial taxa. We found that aortic lesion area (LOD: 7.3, BCI: 22.90-30.75

Mbp, 33 genes within the BCI) and *Lactococcus* abundance (LOD: 7.45, BCI: 25.13-28.06 Mbp, 8 genes within the BCI) QTLs were significantly co-mapped on chromosome 10 in DO-F1 female mice. In addition, CAST/EiJ and PWK/PhJ allele effects were associated with low aortic lesion area and *Lactococcus* abundance levels at the locus. Consistent with these results, *Lactococcus* abundance showed a high positive correlation with plasma total cholesterol (R=0.26, P=7.3E-05 in males; R=0.17, P=0.012 in females) and hepatic total cholesterol (R=0.3, P=3.7E-06 in males; R=0.23, P=0.00047 in females) as well as aortic lesion area (R=0.22, P=0.00031 in males; R=0.14, P=0.028 in females) in both sexes. Based on these results, we assume that *Lactococcus* may play a functional role in the development of atherosclerosis. In order to validate the association between *Lactococcus* and atherosclerosis, CC lines also can be utilized. We plan to measure *Lactococcus* abundance and aortic lesion area levels after challenge with an atherogenic diet in 6-week-old CC mice with CAST/EiJ and PWK/PhJ allelic contribution at the chromosome 10 locus. CC lines with CAST/EiJ and PWK/PhJ allele contributions at chromosome 10 locus may exhibit lower *Lactococcus* abundance and aortic lesion area levels than other CC lines.

After confirming this trend in a preliminary study, we can directly test the effect of *Lactococcus* administration on atherosclerosis in mice. For example, we can compare the development of atherosclerosis and gut microbiota changes after treatment with *Lactococcus* between atherosclerosis susceptible C57BL/6J and atherosclerosis resistant CAST/EiJ. In order to identify the species level of *Lactococcus* found in DO-F1 mice, it is necessary to directly isolate it from DO-F1 fecal samples and then perform metagenomic sequencing. Alternatively, the *Lacococcus* strain identified up to the species level can be purchased from the American Type Culture Collection (ATCC) or DSM companies, and then tested in an atherosclerotic

mouse model. We plan to set 2 of the 4 groups as controls for each C57BL/6J and CAST/EiJ strain and administrate *Lactococcus* strain orally (10^9 CFU/mouse in PBS) to the other 2 groups. All mice will be first treated with the antibiotics ciprofloxacin (0.2 g/L) and metronidazole (1 g/L) once and challenged with an atherogenic diet for 16 weeks. Higher atherosclerosis-related phenotypic levels may be observed in mice administered *Lactococcus* strain. Additionally, in order to verify the effect of the *Lactococcus* strain on atherosclerosis in germ free model, we can confirm that the development of atherosclerosis is accelerated after co-administration of *Lactococcus* strain and core species commonly found in human microbiota to germ free C57BL/6J ApoE$^{-/-}$ mice.

6.3. Future Directions

The use of the DO mouse model with a heterozygous genetic architecture similar to the human genome is just beginning, and it is expected to increase worldwide to discover disease-related targets through multi-omics analysis in the future. The advantages of using this mouse model in the future are as follows. First, it will bring the advancement of systems biology analysis by sharing a multi-omics integrated analysis platform. The DO mouse model has a relatively low barrier to entry because it is easier to control the environmental factors and collect samples than human multi-omics studies. In the future, the data generation cost of multi-omics will continue to decrease, and the high-resolution multi-omics integrated analysis platform, which will be built with the participation of many researchers, will greatly contribute to the advancement of metabolic disease research and academic development for researchers around the world. Second, the use of this mouse model will promote the dissemination of personalized efficacy evaluation techniques that take into account genetic factors. Previously, validation of the target has been performed only in one inbred strain which is susceptible to metabolic diseases

such as the C57BL/6J. However, it will be possible to evaluate the effectiveness of the disease-related target considering genetic factors by validating the target in disease-susceptible (i.e. C57BL/6J) and resistant mouse strains (i.e. CAST/EiJ) that have the contribution of alleles on the target QTL. Third, this mouse model will facilitate the discovery of new targets associated with disease. In the case of human multi-omics data, uncontrolled confounding factors (race, age, sex, medication use, and underlying disease) for each cohort make it difficult to integrate with the omics data from other cohorts. However, the use of DO mice whose confounding factors are controlled and the sharing of rich multi-omics data obtained from them facilitate integration with other omics data from this mouse model, discovery of new disease-related targets, and development of disease diagnosis biomarkers. The biomarkers discovered in this way can be utilized for non-invasive analysis of metabolic organs relating to conditions such as cardiovascular disease, steatohepatitis, and kidney disease in the future.

In summary, this is the first comprehensive study of diet, sex, and genetic effects on atherosclerosis, liver transcriptome, and gut microbiota in eight DO founder strains and DO-F1 mice. These results increase our understanding of diet, sex, and genetic effects on atherosclerosis at the molecular and microbial taxa level. We will investigate the causal relationship between atherosclerosis and disease-related targets through functional validation of genes and microbial taxa that are expected to control atherosclerosis derived in this study according to the method proposed above. However, it is necessary to discover a novel biomarker that is highly related to atherosclerosis through integration with not only the liver transcriptome and gut microbiome we used, but also other omics data that can act as intermediate phenotypes. For example, our study focused on the relationship between the liver transcriptome, gut microbiota, and atherosclerosis, but the multi-omics-derived markers in metabolic organs (aorta and adipose tissue) that are

directly affected by atherosclerosis cannot be overlooked. In particular, single-cell RNA-sequencing technology, which provides transcriptomic profiling for each cell type in tissue, will be able to identify novel atherosclerosis-related markers at the single-cell level by determining which expressed genes in specific cell types are associated with atherosclerosis (Williams et al., 2020). However, since single-cell RNA-sequencing is 10 to 200 times more expensive and offers a limited read depth that can be detected per cell 10 times lower than bulk RNA-sequencing approaches, there are still limitations to applying this technique to hundreds of samples for QTL mapping. In addition, it is important to discover biomarkers for diagnosing atherosclerosis by identifying host or microbial metabolites that are genetically regulated at the same locus with the atherosclerosis via integration with other omics data such as metabolomics and lipidomics. Recently, novel genes related to lipid metabolism were identified via genetic mapping analysis of plasma and liver lipidomics in HMDP and DO mice with natural genetic variation (Linke et al., 2020; Parker et al., 2019). To this end, we plan to perform metabolomic analysis on fecal and/or plasma samples collected from hyperlipidemic DO-F1 mice. Since the pathogenesis of atherosclerosis is associated with host and microbial metabolites, performing multi-omics analysis, genetic mapping, and correlation analysis on these samples from DO mice could provide greater insight into the regulatory mechanisms of atherosclerosis. As a result, the studies presented here were conducted using computational analysis of experimental data from mice and provides sufficient association and convincing data to expand. These results may contribute to the understanding of the regulatory mechanisms of atherosclerosis and serve as a strong basis for future integrated analysis with multiple omics data.

6.4. References

Linke, V., Overmyer, K.A., Miller, I.J., Brademan, D.R., Hutchins, P.D., Trujillo, E.A., Reddy, T.R., Russell, J.D., Cushing, E.M., Schueler, K.L., *et al.* (2020). A large-scale genome-lipid association map guides lipid identification. Nat Metab *2*, 1149-1162.

Parker, B.L., Calkin, A.C., Seldin, M.M., Keating, M.F., Tarling, E.J., Yang, P., Moody, S.C., Liu, Y., Zerenturk, E.J., Needham, E.J., *et al.* (2019). An integrative systems genetic analysis of mammalian lipid metabolism. Nature *567*, 187-193.

Williams, J.W., Winkels, H., Durant, C.P., Zaitsev, K., Ghosheh, Y., and Ley, K. (2020). Single Cell RNA Sequencing in Atherosclerosis Research. Circ Res *126*, 1112-1126.