THE GENETIC ARCHITECTURE OF ATHEROSCLEROSIS AND THE ATHEROSCLEROSIS-ASSOCIATED METABOLITE TRIMETHYLAMINE-N-OXIDE IN THE DIVERSITY OUTBRED MOUSE POPULATION

Tangi L. Smallwood

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum of Genetics and Molecular Biology.

Chapel Hill
2015

Approved by:

Brian J. Bennett

Daniel Pomp

Tim Wiltshire

Praveen Sethupathy

W. Kimryn Rathmell

ABSTRACT

Tangi L. Smallwood: The genetic architecture of atherosclerosis and the atherosclerosis-associated metabolite trimethylamine-N-oxide in the diversity outbred mouse population (Under the direction of Brian J. Bennett)

Inbred mice exhibit strain-specific variation in susceptibility to complex diseases which renders them useful for dissecting the genetic architecture of these traits. Traditional quantitative trait locus (QTL) mapping studies using inbred strains often identify large genomic regions containing many genes and require extensive follow up studies to identify the causal genes underlying the associations. Alternatively, the use of recently developed multi-parent outbred mice designed to be informative for QTL mapping, such as the Diversity Outbred (DO) mouse population, can expedite the identification of causal genes and variants associated with complex diseases. Here, we used DO mice to study the genetic architecture of atherosclerosis and multiple risk factors associated with this complex disease. We fed 292 female DO mice either a high-fat, cholesterol-containing diet to induce atherosclerosis or a low-fat, high-protein control diet for 18 weeks. We then quantified multiple metabolic phenotypes before and after dietary treatment and measured aortic lesion size in the mice after dietary treatment. We performed linkage mapping for these quantitative traits using a linear regression model based on reconstructed founder haplotypes. The additive haplotype model used produces estimates of the effects of the founder alleles which can be used to inform follow up studies and to identify the causal variants underlying the QTL. Among our highly significant results, we identified QTL associated with atherosclerosis, triglycerides, cholesterol, choline, trimethylamine-N-oxide, blood urea nitrogen, and insulin resistance in this cohort of DO mice. These results demonstrate the value of the DO population to improve mapping resolution and to aid in the identification of potential therapeutic targets for complex diseases.

In loving memory of my father

David Lee Smallwood

1943-2008

ACKNOWLEDGMENTS

I am grateful to all of the people who have helped me along the way through graduate school and in the completion of this dissertation.

I want to sincerely thank my amazing advisor Brian Bennett for his diligent teaching and constant support. He is an exemplary mentor and I could not have been more fortunate to have had the opportunity to learn from him. I would like to thank all of the members of my dissertation committee: Praveen Sethupathy, Tim Wiltshire, Daniel Pomp, and Kim Rathmell. Their support and research advice was essential in the completion of the work contained herein. In particular, I want to thank Daniel Pomp for allowing me the opportunity to work on this amazing project and for his advice along the way. I want to thank Kim Rathmell for her pivotal role on my committee throughout my transition between labs. I want to thank Praveen Sethupathy for his sage advice in both science and life. I want to thank Tim Wiltshire for sharing his expertise in QTL mapping and bioinformatics analysis, without which I could not have completed the studies presented here.

I would also like to thank all of the members of the Bennett and Pomp labs. Specifically, I would like to thank Kuo-Chen Jung, Kunjie Hua, and Li-Yang Zhao for all of the hard work and long hours we spent together caring for and phenotyping the DO mice used in this study. I would like to thank Pamela Quizon and Jody Albright for all of their help with sample processing in Kannapolis. I would also like to thank them for

being amazing lab partners during the long nights of experiments and for providing great music to keep us working hard at the bench. In addition, I would like to thank the labs with which we collaborated on these studies. Particularly, I would like to thank the lab of Fernando Pardo-Manuel de Villena for providing cDNAs for eQTL analysis and the lab of Gary Churchill for the liver microarray data of the DO founder strains.

I would like to give a very special thank you to Dan Gatti, the developer of the DOQTL software used for the mapping studies presented in this dissertation. It was his dedication to my success and the countless hours he spent teaching me about QTL mapping in multi-parent populations that made it possible for me to complete this work.

I would like to sincerely thank all of the people who have supported me and continue to support me in my life-long pursuit of knowledge, starting with my loving parents, David and Frances Smallwood, who always stood by me even in my most outrageous of endeavors. Thank you for always believing in me and what I could accomplish.

I would also like to thank my family and friends without whom I would not be the person I am today. A special thank you to my siblings Shayna Smallwood, Tasha Brenenstuhl, Sherry Teague, and Michael Smallwood for helping me in my earliest of experiments, mostly conducted in the backyard of our Greensboro, NC home. Thank you to my friends Lonna Mollison, Alex Arreola, Stephanie Bellendir, and Christina Trexler for always being there for me throughout graduate school. Each coming from different scientific backgrounds, Lonna, Alex, Stephanie, and Christina helped me broaden my own approaches to many research questions. I thank them for the scientific discussions, both in practice talks and around the fire pit. I would also like to recognize my dogs

Kreuza Tropic and Zero Cool Kleinholz for helping me maintain balance between science in the lab and science in the field. Thank you Kreuza and Zero for being the best dogs anyone could ever ask for.

Last but not least, I would like to think my loving husband Jacob Wang and our amazing son Legend. Legend, you mean the world to me and I want to thank you for reminding me to experience life unfiltered. We have a lifetime of discovery ahead of us and I look forward to every minute of it. Jacob, you have been by my side through the good and the bad and have encouraged me through all of it. You have been a sounding board for me as I thought through many experiments and your perspective as a medical doctor has been instrumental to me keeping my eye on the clinical impact of basic scientific research. I love you and thank you both for everything.

TABLE OF CONTENTS

LIST OF TABLES x

LIST OF FIGURES xi

LIST OF ABBREVIATIONS xiii

CHAPTER 1: INTRODUCTION 1

ATHEROSCLEROSIS 1

COMPLEXITY OF ATHEROSCLEROSISIS RISK FACTORS 2

DISCOVERY OF TMAO AS A RISK FACTOR FOR ATHEROSCLEROSIS 10

DERIVATION OF ADVANCED POPULATION FOR QTL MAPPING IN MICE 13

QTL MAPPING IN THE DIVERSITY OUTBRED MICE 15

SUMMARY OF THE WORK PRESENTED IN THIS DISSERTATION 18

SIGNIFICANCE OF FINDINGS PRESENTED 19

CHAPTER 2: HIGH-RESOLUTION GENETIC MAPING IN THE DIVERSITY OUTBRED MOUSE POPULATION IDENTIFIES *APOBEC1* AS A CANDIDATE GENE FOR ATHEROSCLEROSIS 22

CHAPTER OVERVIEW 22

INTRODUCTION 23

MATERIALS & METHODS 24

RESULTS 29

DISCUSSION 36

CHAPTER 3: IDENTIFICATION OF NUMB AS A POTENTIAL GENETIC REGULATOR OF THE ATHEROSCLEROSIS-ASSOCIATED METABOLITE TRIMETHYLAMINE-N-OXIDE IN THE DIVERSITY OUTBRED MICE 62

CHAPTER OVERVIEW 62

INTRODUCTION 63

MATERIALS & METHODS 64

RESULTS 68

DISCUSSION 75

CHAPTER 4: QUANTITATIVE TRAIT LOCUS ANALYSIS OF METABOLIC PHENOTYPES IN THE DIVERSITY OUTBRED MICE 91

CHAPTER OVERVIEW 91

INTRODUCTION 91

MATERIALS & METHODS 93

RESULTS 97

DISCUSSION 103

CHAPTER 5: SUMMARY, DISCUSSION AND CONCLUSIONS 121

SUMMARY AND SIGNIFICANCE 121

CONCLUSIONS AND DISCUSSION 122

FUTURE DIRECTIONS 132

REFERENCES 145

LIST OF TABLES

Table 1.1: Summary of known QTL regulating TMAO levels in mice and humans 12

Table 1.2: Summary of QTL mapped in the DO mice to date..... 17

Table 2.1: Compositions of the synthetic, high protein, and atherogenic diets used in this study 58

Tabel 2.2: Effects of high protein and high fat, cholic acid diets on clinical markers of cardiovascular disease in the DO mice 60

Table 2.3: Quantitative trait loci for clinical markers of cardiovascular disease in the DO mice at baseline and after dietary treatment 61

Table 3.1: Summary of QTL identified as associated with choline and TMAO in the DO mice 87

Table 3.2: SNP analysis of Chromosome 12 QTL associated with TMAO 88

Table 3.3: SNP analysis of known variants within 10 kb of the candidate gene *Numb* on Chromosome 12 89

Table 3.4: Spearman's correlations between *Numb* expression and cardiovascular disease traits in the DO mice 90

Table 4.1: QTL significantly associated with BUN, HOMA-IR, and change in total cholesterol 117

Table 4.2: Summary of 21 suggestive QTL identified for ALT, AST, BUN, betaine, creatinine, glucose insulin ratio, and HOMA-IR at baseline and after dietary treatment 118

Table 4.3: SNP analysis of the Chromosome 8 QTL associated with HOMA-IR after dietary tretment 119

Table 4.4: Variants within 10 kb of the candidate gene *Fgfr1* 120

LIST OF FIGURES

Figure 2.1: Overall design of the QTL mapping study 42

Figure 2.2: Effects of diet on cardiovascular risk factors in the DO mice 43

Figure 2.3: QTL mapping of clinical markers of cardiovascular disease in 6 week old DO mice at baseline 44

Figure 2.4: High-resolution mapping of significant QTL on Chromosome 9 associated with plasma triglyceride levels 46

Figure 2.5: QTL mapping of total cholesterol after dietary treatment in the DO mice 47

Figure 2.6: Liver expression of candidate genes in the Chromosome 9 QTL region associated with baseline triglyceride levels 49

Figure 2.7: Liver expression of genes previously associated with markers of cardiovascular disease on Chromosome 9 50

Figure 2.8: QTL mapping of atherosclerosis in the DO mice 51

Figure 2.9: A/J preferentially expresses the long isoform of *Apobec1* in response to a high fat, cholic acid diet 54

Figure 2.10: Identification of a *cis*-eQTL on Chromosome 6 for *Apobec1* expression 55

Figure 2.11: *Cis*-eQTL for hepatic *Apobec1* short and long isoforms in the DO mice 56

Figure 2.12: *Apobec1* and ApoB levels are dependent on the genotype of UNC11996440, the Chromosome 6 peak SNP associated with atherosclerosis 57

Figure 3.1: Effects of high fat, cholic acid and high protein diets on choline and TMAO levels in the DO mice 80

Figure 3.2: QTL mapping of choline in DO mice after dietary treatment 81

Figure 3.3: QTL mapping of TMAO in the DO mice at baseline 82

Figure 3.4: QTL mapping of TMAO in the DO mice after dietary treatment. 83

Figure 3.5: Differentially expressed positional candidates within the Chromosome 12 QTL associated with TMAO 84

Figure 3.6: Identification of a *cis*-eQTL regulating *Numb* expression in the DO mice 86

Figure 4.1: Effects of high fat, cholic acid and high protein diets on ALT, AST, and BUN levels in the DO mice 108

Figure 4.2: QTL mapping of blood urea nitrogen levels in DO mice at baseline 110

Figure 4.3: QTL mapping of HOMA-IR in DO mice after dietary treatment. 112

Figure 4.4: Chromosome 9 QTL associated with change in total plasma cholesterol exhibits complex genetic regulation by multiple SNP types 114

LIST OF ABBREVIATIONS

ACF	Apobec1 complementation factor
ALT	Alanine transaminase
ANOVA	Analysis of variance
APOA	Apolipoprotein A
APOB	Apolipoprotein B
APOE	Apolipoprotein E
APOBEC1	ApoB mRNA editing enzyme
ARTLES	Arterial lesions QTL
AST	Aspartate transaminase
ATH1	Atherosclerosis QTL 1
ATH37	Atherosclerosis QTL 37
BGLU13	Blood glucose level QTL 13
BUN	Blood urea nitrogen
CAD	Coronary artery disease
CC	Collaborative cross mouse resource
CD36	Cluster of differentiation 36
CETP	Cholesterylester transfer protein
CON6	Chromosome 6 congenic line
CQ4	Cholesterol QTL 4
CQ5	Cholesterol QTL 5
CVD	Cardiovascular disease
DO	Diversity outbred mouse resource
ELISA	Enzyme-linked immunosorbent assay

ESRRG	Estrogen-related receptor gamma
eQTL	Expression quantitative trait locus
F1	First filial generation
F2	Second filial generation
FGF	Fibroblast growth factor
FGFR	Fibroblast growth factor receptor
FMO3	Flavin monooxygenase family member 3
GDF	Growth differentiation factor
GWAS	Genome-wide association study
HDL	High density lipoprotein
HDMP	Hybrid mouse diversity panel
HFCA	High fat, cholic acid diet
HOMA-IR	Homeostasis model assessment-estimated insulin resistance
HS	Heterogeneous stock mouse population
IDL	Intermediate density lipoproteins
IGHV	Immunoglobulin heavy chain variable region
LDL	Low density lipoprotein
LDLR	Low density lipoprotein receptor
LIPC	Hepatic lipase
LOD	Logarithm of odds ratio
LPL	Lipoprotein lipase
LRP	LDL-receptor related protein
5-LO	5 lipoxygenase
MegaMUGA	Mega mouse universal genotyping array

NPC1L1	Niemann-Pick C1-Like 1
OX40	Cluster of differentiation 134
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
QTL	Quantitative trait locus
RIL	Recombinant inbred line
SLC30A7	Solute carrier family 30, member 7
SNP	Single nucleotide polymorphism
SR-A1	Scavenger receptor A1
TC	Total cholesterol
TMA	Trimethylamine
TMAO	Trimethylamine-N-oxide
TMAU	Trimethylaminurea
TNFSF4	Tumor necrosis factor (ligand) superfamily, member 4
TRIGQ1	Triglyceride QTL 1
TSS	Transcription start site
UTR	Untranslated region
VLDL	Very low density lipoprotein

CHAPTER I. INTRODUCTION

ATHEROSCLEROSIS

Cardiovascular diseases (CVDs) are a group of diseases that affect the heart and blood vessels and it is estimated that CVDs kill more people annually in the United States than any other disease [1]. Based on the most recent census of death certificate data, deaths due to CVDs in the US account for an estimated 26% of all deaths [2, 3]. Additionally, though the standards of quantification by recorded cause of death vary worldwide, is has been estimated that CVDs are the leading cause of morbidity globally as well [4].

The most common cause underlying CVD is atherosclerosis, which accounts for 25-30% of cardiovascular disease-related deaths [5]. It is important to note that atherosclerosis is a progressive disease which is present to some extent in 2 out of every 3 Americans under the age of 35 [6]. While atherosclerosis is a prevalent pathological condition that often leads to fatal adverse cardiovascular events such as heart attack or stroke, it develops asymptomatically in young individuals [7-9]. Therefore, it is important to identify early predictive biomarkers for implementation of behavioral modifications and therapies to prevent the progression of the disease.

At a molecular level, atherosclerosis is a result of endothelial dysfunction. It is initiated by localized damage to the endothelial cells lining the arterial wall or by the presence of oxidized LDL inside the vessel wall [10]. Injury to the endothelial cells includes not only loss of cells at the site of injury, but can result from a variety of processes that influence endothelial cell function independent of the loss of endothelial cells [11]. This begins a cascade of events that

ultimately results in macrophage recruitment and aggregation at the site of LDL deposition [12]. As macrophages engulf the accumulating LDL, they grow in size and become lipid-filled foam cells. An atherosclerotic plaque of this type does not impede blood flow through the vessel and is referred to as a fatty streak. In advanced lesions, as LDL continues to accumulate and more foam cells are formed, a fibrous cap forms over the growing plaque that may occlude the vessel lumen and restrict blood flow [13]. Eventually, sheering stress of blood flow or some other force can cause this stable fibrous plaque to rupture and completely restrict blood flow leading to potentially fatal heart attack or stroke.

COMPLEXITY OF ATHEROSCLEROSIS RISK FACTORS

Atherosclerosis is known to be a complex disease resulting from a combination of risk factors, including both environmental and genetic risk factors. Using a forward genetics approach, early family studies identified several inherited autosomal dominant forms of dyslipidemia, the most common of which is known as familial combined hypercholesterolemia, which resulted in abnormally high serum cholesterol levels and an associated increased risk of atherosclerosis developing at an early age in affected families [14, 15]. Subsequent large family studies were performed to identify the genetic regulators of hyperlipidemia in relation to CVD and these lead to the successful identification of several genes regulating plasma cholesterol and triglycerides, including CETP, LPL, and LDLR [16, 17]. Subsequent experimental analyses implicated dysregulation of lipid levels as having a major influence on the progression of atherosclerosis [18]. Taken together, dyslipidemia has been shown to contribute to atherosclerosis under conditions of elevated LDL and TC and decreased HDL and specific guidelines have been proposed to reduce risk of CVD in individuals based on regulating lipid levels [19, 20]. However, dyslipidemia does not account for the overall susceptibility to

atherosclerosis as reported in both animal models and human studies. Recent evidence has implicated a role for inflammatory responses [21-24] and immune function [25] in atherosclerosis progression.

In addition to the identification of genes linked to atherosclerosis susceptibility within families and as a result of variation in known lipoprotein metabolism genes, several human GWAS studies have been performed to identify novel genetic loci that contribute to disease susceptibility [26]. As of 2003, the Human Genome Project mapped millions of common SNPs, representing locations in the genome where a single nucleotide differs between individuals and occurs with a minor allele frequency of at least 5% [27]. By associating common SNPs inherited preferentially with a disease state, researchers can identify genomic regions that are associated with an increased risk of disease susceptibility. This information is then used to inform follow up studies with an end goal of identifying the underlying genetic variants and biology regulating the association. GWAS studies have proven quite successful in identifying large genomic loci that contain genes known to contribute to atherosclerosis and dyslipidemia, such as CETP, LDLR, LIPC, the APOA5-APOA4-APOC3-APOA1 cluster, APOB [28-31]. By studying genome-wide genetic variants, GWAS studies are well suited to generating new hypotheses and the identification of novel genes associated with atherosclerosis and its associated risk factors as well. However, most of the loci identified in GWAS studies tend to be large, have modest effects on the phenotype, and contain hundreds of candidate genes [32]. As such a human GWAS studies require extensive validation of findings in order to identify the underlying genes and causal variants responsible for the association. The identification of genetic loci and subsequent validation of candidate genes is an exciting avenue for the discovery of novel biomarkers of atherosclerosis susceptibility.

Laboratory mice have long been used to study the genetics of human diseases and several mouse models of atherosclerosis have been developed since the domestication of laboratory strains in the early 1900s. The first genetic analyses of atherosclerosis in mice involved the use of high fat diets, because normal mice fed a regular chow diet do not develop atherosclerotic lesions [33]. Similar to the population-specific susceptibility in humans, these early studies identified strain variation in susceptibility to lesions across inbred strains of mice [34]. They also identified strain variation in serum levels of lipids, lipoproteins, and apolipoproteins and documented the correlation of lipid levels with lesion size [33, 35, 36]. Specifically, Bruell et al. identified cholesterol levels ranging from 128 mg/ 100mL in C57BL/6J mice to 208 mg/ mL in C3H/HeJ mice and recognized that male mice exhibited significantly higher cholesterol levels compared to female mice [37]. Further significant strain, sex, and age differences in lipid levels have been observed across unrelated inbred strains of mice [38].

These early studies of atherosclerosis in mice utilized undefined diets that often resulted in an increased mortality rate and increased morbidity, including secondary disease states such as fatty liver and gallstones. In 1985, research by Beverly Paigen identified an atherogenic diet with lower concentrations of cholesterol and fat than the original Roberts and Thompson diet that was capable of producing lesions in mice without an associated increase in mortality [39]. The Paigen diet was a defined mixture of Purina breeder chow and the atherogenic Thomas-Hartroft diet, such that the total mixture contained 15% fat, produced appreciable lesions within 10-14 weeks, and did not cause mortality in treated animals. Using this diet, variation in total plasma cholesterol and strain susceptibility to atherosclerosis was characterized across the A/J, AKR/J, BALB/cJ, C3H/J, C57L/J, C57BL/6J, DBA/2J, SWR/J, NZB/J, and 129/J common inbred strains. They found that BALB/cJ, C3H/J, A/J, SWR/J, and NZB/J mice did not exhibit diet-

induced lesions after 14 weeks; while 129/J, AKR/J, and DBA/2J exhibited lesions after 14 weeks and C57L/J and C57BL/6J were the most susceptible to lesion formation exhibiting lesions as early as 10 weeks on diet [40]. Additionally, although they observed differences in cholesterol levels across the mouse strains, they found that total cholesterol levels were not correlated with susceptibility to lesions indicating underlying pathological differences in lesion formation. Taken together, these results indicate the relevant impact that the choice of inbred strains can have on the interpretation of experimental results.

Considerations of the impact of diet composition on diet-induced models of atherosclerosis led to further advances in defining diets for studies of lesions in mice. In 1990, the Paigen group defined the composition of low and high fat purified, synthetic diets [41]. Such synthetic diets control for the variable components of chow and reduce the potential influence that variable components of standard chow may have on experimental results. By testing different sources of fat and cholesterol, a high fat synthetic diet containing 50% sucrose, 15% cocoa butter, 1% cholesterol, and 0.5% sodium cholate was optimized to produce aortic lesions while reducing the additional pathological burdens of increased ALT, fatty liver, and gallstones [41].

While early studies of diet-induced lesions in mice were quite successful at identifying strain variation in susceptibility to lesions, it was not until the advent of mouse gene targeting technologies that the complexity of the development and progression of atherosclerosis could be fully elucidated. The generation of the *ApoE* knockout mouse (*ApoE*$^{-/-}$) represents the first knockout mouse to be characterized as to its effects on atherosclerotic lesions [42]. ApoE is an apolipoprotein and a structural component of chylomicrons and IDLs. The presence of ApoE on the surface of lipoproteins is recognized by LDLR in the liver, where cholesterol from these

particles is deposited. Previous work that led to the creation of the *ApoE*$^{-/-}$ mouse included evidence that a genetic variant in the human *ApoE* gene, *ApoE-2*, reduces its ability to bind LDLR and results in the early development of atherosclerosis [43, 44].

Importantly, *ApoE*$^{-/-}$ mice represent the first mouse model to develop atherosclerotic lesions with a similar distribution of severity and appearance of that seen clinically in humans, marking a major advance in the ability to characterize mechanisms of atherosclerosis using these mice as a model [45]. In the absence of ApoE, mice have significantly increased total cholesterol levels (~600 mg/dL) and develop small lesions on a normal chow diet [46, 47]. This phenotype is exacerbated by the inclusion of a high fat, high cholesterol "atherogenic" diet containing 15.8% fat, 1.25% cholesterol, and 0.5% sodium cholate [48].

Shortly after the development of the *ApoE*$^{-/-}$ mouse model, homologous recombination was similarly employed to develop a mouse lacking the LDLR known as the *Ldlr* knockout mouse (*Ldlr*$^{-/-}$) [49]. Mutations in LDLR are recognized as one of the underlying genetic causes of familial hypercholesterolemia [50]. The LDL receptor is responsible for recognizing circulating lipoprotein particles, particularly ApoB100- and ApoE-containing lipoproteins, and deficiencies of LDLR in humans have been shown to cause severe hypercholesterolemia and premature atherosclerosis [51, 52]. *Ldlr*$^{-/-}$ mice have only slightly elevated plasma cholesterol levels (~250 mg/dL) on a normal chow diet and develop small lesions comparatively slowly; while *Ldlr*$^{-/-}$ mice fed an high fat diet exhibit extensive atherosclerosis, subcutaneous xanthomas, and plasma cholesterol levels equivalent to those seen in the *ApoE*$^{-/-}$ mice fed an atherogenic diet (~1500 mg/dL) [53].

Extensive use of transgenic and additional knockout mouse models as well as double knockout mice have furthered the understanding of the molecular mechanisms of atherosclerosis

and the interplay between oxidized LDL deposition within plaques and lipoprotein clearance. Construction of *ApoE*$^{-/-}$; *Ldlr*$^{-/-}$ double knockout mice allowed researchers to first identify that ApoE binds the receptor for chylomicron remnants, LPR, in addition to LDLR [54]. *Apobec1*$^{-/-}$; *Ldlr*$^{-/-}$ mice were developed and characterized to more accurately reflect the ratio of ApoB100 to ApoB48 in humans, since mice express the ApoB mRNA editing enzyme *Apobec1* and edit ApoB100 to ApoB48 in both the liver and the intestines, while the ApoB48 isoform is limited to the intestines in humans [55]. Both *Apobec1*$^{-/-}$; *Ldlr*$^{-/-}$ mice and transgenic *Ldlr*$^{-/-}$ mice overexpressing human ApoB100 exhibit severe hyperlipidemia, elevated LDL cholesterol, and the development of spontaneous lesions much more severe than the lesions observed in *Ldlr*$^{-/-}$ mice alone, which is contributed to the elevated levels of atherogenic ApoB100 particles in these mice [56, 57]. These genetic models of atherosclerosis along with diet-induced manipulations of the phenotype have made mice a valuable model for studying the genes and pathways regulating atherosclerosis.

As a complement to the reverse genetics approach of knocking out a specific gene and measuring altered phenotypes to determine the processes the gene is involved in, a commonly used forward genetics approach is to associate naturally occurring alleles with the phenotype of interest to pinpoint genetic loci by mapping the phenotypic traits relative to chromosomal markers. The chromosomal loci identified can then be thoroughly investigated to determine the causal gene and/or variant that is responsible for the measured phenotype [58]. These methods have been successfully employed in the identification of genes regulating rare, Mendelian disorders. However, many common human diseases are polygenic in nature. Additionally, complex diseases such as atherosclerosis, diabetes, and obesity exhibit quantitative variation [59, 60]. In the case of quantitative traits, the phenotype is continuous and not regulated by a single

gene or chromosomal region. In mapping quantitative trait loci (QTLs), the goal is to identify the regions of the genome that affect variation in a quantitative trait, such as aortic lesion size or serum lipid levels. QTL results are often represented by a LOD score, or logarithm of odds ratio, representing the probability of an effect at the locus versus no effect at the locus.

QTL mapping provides an unbiased approach to identify novel genes influencing quantitative traits. Similar to the variation observed in SNPs throughout the human genome and across populations, mice exhibit SNPs at which sequence variation occurs across common inbred strains. Using available genotype data for SNPs, researchers have performed genome-wide linkage studies to identify QTL associated with a multitude of cardiometabolic phenotypes, including diabetes, obesity, metabolic syndrome, and atherosclerosis. With the recent advances in whole genome sequencing, including the development of high density genotyping arrays, genome-wide association studies in mice have contributed to the number of QTLs identified as regulating atherosclerosis and its related phenotypes [61].

Identification of atherosclerosis QTL across inbred strains takes advantage of several crossing strategies including backcross, F2 intercross, or the generation of RILs. The goal of each of these crossing schemes is to introduce variation in the phenotype with knowledge of the pedigree so that defined chromosomal regions can be identified as QTL influencing the trait. Based on the earlier studies that characterized strains as susceptible or resistant to diet-induced atherosclerotic lesion formation, researchers were able to design experiments such that variation in the trait was introduced by outcrossing animals from susceptible strains to animals from resistant strains. A parental cross of an atherosclerosis-susceptible inbred mouse with an atherosclerosis-resistant inbred mouse results in the generation of heterozygous F1 animals that exhibit an intermediate phenotype. In a backcross, the F1 can then be crossed back to a mouse of

one parental genotype to generate an F2 that exhibit recombinations along one chromosome with the other chromosome contributed exclusively by the backcross parent such that the F2 will be approximately 50% homozygous and 50% heterozygous across the genome. In an intercross, the F1 animals are crossed to one another such that chromosomal rearrangements result in recombinations between alleles across the genome. In this way, each F2 animal represents a unique combination of alleles derived from the two parental strains. QTL can be identified in F2 mice susceptible to lesion formation that share chromosomal intervals contributed by the atherosclerosis susceptible parental strain that are not shared by the F2 animals exhibiting resistance to lesion formation.

QTL mapping studies in inbred mice have identified more than 30 loci regulating atherosclerosis. The earliest QTL mapping of atherosclerosis was performed using RILs derived from crosses of atherosclerosis susceptible C57BL/6 mice and either BALB/c mice or C3H/He mice, both of which are resistant to the development of atherosclerotic lesions [62]. RILs represent a stable collection of inbred strains originally derived from carefully selected parental strains that are divergent for the phenotype of interest and bred to isogenicity [63]. Using 17 RILs phenotyped as susceptible or resistant to lesion formation determined by mean number of lesions, researchers identified the *Ath-1* locus on chromosome 1 [62]. The same group went on to identify another locus in RILs derived from C57BL/6J and A/J reciprocal crosses that interacts with *Ath-1* such that resistant alleles at either locus conferred a resistant phenotype to the animal [64]. RIL lines were later used to confirm the identification of *Ath-1* and the additional identification of *Ath-3* [65, 66]. The identification of the *Artles* locus is an example of an F2 intercross to identify atherosclerosis QTL. The locus was identified in an F2 cross between atherosclerosis-susceptible C57BL/6J mice and atherosclerosis-resistant CAST/EiJ mice [67].

Numerous additional studies have identified QTL regulating atherosclerosis using F2 crosses, including 7 crosses of different inbred mouse strains, 5 crosses of *ApoE*$^{-/-}$ sensitized inbred mice, and 2 crosses of *Ldlr*$^{-/-}$ sensitized inbred mice, the results of which have been summarized in recent reviews [68, 69]. Importantly, 63% of atherosclerosis QTLs reported in humans have also been identified in homologous chromosomal regions in the mouse, demonstrating the importance of QTL mapping in mice for furthering our understanding of the biology underlying this detrimental human disease [69]. Ultimately, the goal is to identify the genes at each locus that are involved in atherosclerosis and validate their roles in follow up studies in which candidate genes are tested independently for a functional role related to the phenotype.

DISCOVERY OF TMAO AS A RISK FACTOR FOR ATHEROSCLEROSIS

Even more recently, major roles for the microbiome in human diseases have been discovered. Several papers published in the last two years have indicated that intestinal microbiota metabolism has an impact on cardiovascular disease risk and the development of atherosclerosis [70]. TMAO was identified as a predictive biomarker linking the microbiome and CVD risk. TMAO is an oxidation product of trimethylamine (TMA) and it is found in both humans and mice. TMAO is synthesized as a conversion product from TMA, which is derived from the breakdown of dietary phosphatidylcholine and is characterized by a strong, fishy odor. Classically, studies since the 1970s have indicated an accumulation of gut flora-derived TMA in the occurrence of trimethylaminurea (TMAU), which is a genetic disorder characterized by a pungent foul body odor and caused by rare mutations in the *flavin mono-oxygenase 3 (Fmo3)* gene [71]. Normally, TMA is oxidized to TMAO primarily through the action of FMO3. TMAO was implicated as a potentially proatherogenic metabolite in humans based on a study of human patients with clinical manifestations of cardiovascular disease. Recent studies have shown a

direct link between plasma levels of TMAO, *Fmo3*, and atherosclerosis. Several recent studies have linked dietary intake of choline, phosphatidylcholine, and L-carnitine with increased plasma TMAO levels and increased cardiovascular risk in human subjects [70].

A study by Hazen et al. showed that hepatic *Fmo3* expression and circulating TMAO levels were correlated and that variation in TMAO levels could account for 11% of the variation in atherosclerotic lesion size among a panel of recombinant inbred strains of mice [72]. However, the same study found that treatment of mice with an *Fmo3*-specific antisense oligonucleotide resulted in a 90% reduction in liver *Fmo3* mRNA and a corresponding 47% decline in plasma TMAO levels, suggesting that, while FMO3 plays a major role in regulating TMAO levels, there may be other factors regulating TMAO. Hazen and colleagues then showed that TMAO promotes atherosclerosis in C57BL/6J *ApoE*$^{-/-}$ mice fed a diet supplemented with either choline or TMAO and that plasma TMAO is significantly positively correlated with atherosclerotic lesion size. They showed that when TMAO is added to the diet of atherosclerosis-susceptible C57BL/6J *ApoE*$^{-/-}$ mice, levels of the atherosclerosis-associated inflammatory markers CD36 and SR-A1 were significantly increased on macrophages compared to C57BL/6J *ApoE*$^{-/-}$ mice on a chow diet alone. However, while studies using inbred mouse strains have shown that TMAO increases atherosclerotic lesion size, surface expression of CD36 and SR-A1, and macrophage foam cell formation, it is still unclear what role TMAO plays in promoting atherosclerosis. Further analysis of additional regulators of TMAO and studies of the contribution of TMAO to atherosclerosis susceptibility in additional mouse populations will be a valuable step toward understanding its role in atherosclerosis. **Table 1.1** summarizes QTL studies performed to identify loci regulating TMAO to date.

Organism	Study population	Chromsome (Location)	P value	Candidate Gene
Human	GeneBank	1 (1q23.3)	$2.8x10^{-7}$	*NUF2*
Human	GeneBank	2 (2p12)	$2.8x10^{-7}$	None
Mouse	HDMP	3 (110-115 Mb)	$2.4x10^{-6}$	*Slc30a7*

Table 1.1: Summary of known QTL regulating TMAO levels in mice and humans.

DERIVATION OF ADVANCED POPULATION FOR QTL MAPPING IN MICE

QTL mapping studies in mice have proven quite successful at identifying loci associated with atherosclerosis, with at least 30 atherosclerosis loci identified to date [68]. However, many of these loci are large and encompass several hundred positional candidate genes that require follow up validation to determine the causal associations with the phenotype. Indeed, of the known atherosclerosis QTL in mice, only two of the genes underlying them have been identified.

The gene underlying *Ath-1* has been identified as *Tnfsf4* in studies using mice with targeted mutations in the *Tnfsf4* gene which exhibit smaller lesions than control mice [73]. The *Tnfsf4* gene encodes the OX40 ligand. It is expressed on activated T-cells and promotes survival of T-cells at sites of inflammation and could potential exacerbate atherosclerosis via proinflammatory mechanisms, although the precise mechanisms by which *Tnfsf4* influences atherosclerosis are not fully understood [74, 75].

For the *Artles* QTL on Chromosome 6, originally identified in an F2 cross between atherosclerosis-susceptible C57BL/6J mice and atherosclerosis-resistant CAST/EiJ mice, researchers created a congenic strain (CON6) containing the resistant chromosome 6 region from CAST/EiJ on an otherwise C57BL/6J genetic background and found that this strain exhibited a significant reduction in *5-LO* mRNA and protein levels and went on to show that transplantation of bone marrow from *5-LO*$^{-/-}$ mice to *LDLr*$^{-/-}$ mice markedly reduced atherosclerotic lesion development [76]. 5-LO is a macrophage-expressed enzyme involved in leukotriene biosynthesis that is thought to be involved in atherosclerosis via proinflammatory mechanisms [77]. This same group undertook the extensive process of creating subcongenic strains to further investigate the QTL region and found that there were at least two subregions, termed *Ath37* and *Ath38*, which affected the size of atherosclerotic lesions independently of 5-LO [78]. This example is

one of many that illustrate the potential complexity of interpreting results of QTL mapping studies in inbred strains of mice.

In an effort to improve the resolution of mapping genetic loci responsible for phenotypic variation in mouse models, researchers have begun to use alternative strategies to create more genetically heterogeneous populations of mice to perform mapping studies. In 2002, a population of heterogeneous stock (HS) mice were first described [79]. The HS mice were derived from 8 inbred founder strains (A/J, AKR/J, BALBc/J, CBA/J, C3H/HeJ, C57BL/6 J, DBA/2 J, and LP/J) and were well suited for fine mapping of QTLs with moderately small effects on the phenotype of interest [80]. Indeed, these mice have been used to map 843 QTLs relating to 97 different traits, including a chromosome 14 QTL associated with blood glucose and a chromosome 1 QTL associated with HDL cholesterol [81].

The hybrid mouse diversity panel (HDMP) is a panel of classical inbred strains and recombinant inbred strains that, when used for genome-wide association mapping of complex traits in mice, has been documented to improve mapping resolution [82, 83]. Several studies have utilized HMDP to identify loci of ~1-2 Mb in size [84-91]. Relevant to the study of atherosclerosis, Bennett et al. conducted a pivotal study in which they crossed HDMP mice to human apoB-100 transgenic mice fed a HFCA diet to induce lesion formation enabling them to detect the previously identified *Ath30* locus on Chromosome 1 and refine the region to 1.7Mb containing 1 mRNA and 31 protein-coding genes. Corroboration of mRNA expression data from HMDP and *ApoE*$^{-/-}$ mice enabled them to identify a single gene, *Desmin*, as the candidate responsible for the atherosclerosis-associated phenotype [92]. Additionally, the previously mentioned genome-wide association study for plasma TMAO levels in mice utilized the HDMP to identified a chromosome 3 locus encompassing the positional candidate *Slc30a7* [93].

In 2004, a groundbreaking project to develop a population of recombinant inbred lines constructed from a mosaic of 5 common inbred strains and 3 wild-derived strains was undertaken to advance QTL analysis in mice [94]. The project known as the Collaborative Cross (CC) involved the construction of RILs using a unique funnel crossing strategy of the 8 parental strains of mice, which included A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HiLtJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ. The CC has been proven to be a highly informative population for genome-wide analysis due to the increased genetic diversity achieved by the inclusion of wild-derived strains and the uniform distribution of this variation [95, 96]. To date, the CC contains 12 distributable lines (bred to between 90- 97% homozygosity) and 36 completed lines (bred to 98% homozygosity) available for genetic analyses.

QTL MAPPING IN THE DIVERSITY OUTBRED MICE

The diversity outbred (DO) mice are a population of mice derived from the same eight founder strains used to construct the CC mice [97]. The DO mice were designed to maximize heterozygosity for improved mapping resolution in QTL analyses. Alleles present in the DO mice represent ~90% of the known variation present in laboratory mice due to the inclusion of the wild-derived WSB/EiJ, PWK/PhJ, and CAST/EiJ strains and this variation is distributed throughout the genomes of DO animals. The DO mice are a population of mice that have been strategically outbred to increase the number of recombination events per animal which should allow for finer QTL mapping resolution to smaller chromosomal intervals. While the CC RILs are maintained as inbred lines, the DO mice are consistently maintained as unique individuals outbred randomly every generation for 3 or 4 generations per year. In this way, the population is maintained as ~175 genetically diverse lines by random mating of one male and one female from the first litter of each previous mating pair. The DO mice represent the same set of allelic

variants found in the CC resource, but each DO individual represents a set of uniquely segregating alleles. While individual DO animals are genetically unique and cannot be replicated, the QTL can effectively be reproduced on different genetic backgrounds with the incorporation of follow up studies using the CC mouse resource. As recently reported, the contribution from each of the 8 founder strains ranges from 11.3-13.8% and these mice have levels of heterozygosity closer to that seen in human populations than traditional laboratory mouse strains. Indeed, the most recent generation of DO mice are estimated to have approximately 390 recombination events per animal [98]. Taken together, the DO mice represent a valuable resource for the potential of improving QTL mapping resolution to smaller genetic intervals encompassing fewer genes for follow up validation studies.

Preliminary studies have confirmed that this genetically heterogeneous population has the power to improve mapping resolution of complex traits. As of 2012 this population had been described and a preliminary analysis of plasma lipid parameters had been reported [97, 98]. Specifically, a significant QTL was reported for the variable secondary phenotype change in total plasma cholesterol from 8 to 19 weeks in 91 DO mice. A 2-LOD support interval defined the QTL as a 2 Mb region containing only 11 genes [98]. Several papers have reported QTL in DO mice, including QTL for plasma cholesterol, behavioral phenotypes, and pain sensitivity [99, 100]. **Table 1.2** summarizes the results of the QTL mapping studies conducted to date using DO mice. The recent release of the DOQTL mapping software as a freely available R package should enable many more researchers to take advantage of the improved mapping resolution the DO has to offer [101].

Phenotype	# of Mice	Chr	Interval (Mb)	Significant	# of Genes	Candidate gene	Reference
17-week plasma cholesterol	141	11	60.1-65.3	Suggestive	82	*Dhrs7b*	Churchill, 2012
Change in plasma cholesterol	91	3	50.3-52.3	Significant	11	*Foxo1*	Svenson, 2012
Open field (Time in center slope)	283	4	147.7-149.3	Significant	32	NA	Logan, 2013
Open field (Duration immobile in the center field)	283	2	93.2-100.2	Significant	35	*Hsd17b12*	
	283	6	114.1-115.9	Significant	15	*Slc6a1, Syn2*	
Light-dark box (Duration in light side)	283	11	95.0-96.6	Significant	41	*NA*	
	283	8	107.5-110.4	Significant	96	*Cdh1*	
Visual-cliff (Distance traveled in bottom area)	283	14	21.6-23.2	Suggestive	14	*Myst4*	
Tail suspenstion (Climbing frequency)	283	6	97.8-99.2	Significant	5	*Foxp1*	
Thermal pain sensitivity	261	8	110.4-114.2	Suggestive	44	*Hydin*	Recla, 2014
Neutrophil count	742	1	126.7-132.4	Significant	73	*Cxcr4*	Gatti, 2014
6-week triglycerides	262	9	50.2-51.6	Significant	34	NA	Smallwood, 2014
6-week total cholesterol	277	13	28.7-43.8	Suggestive	181	NA	
6-week glucose	257	5	86.3-99.0	Suggestive	192	NA	
6-week insulin	235	19	22.6-60.8	Suggestive	532	NA	
24-week triglycerides	264	12	49.6-102.3	Suggestive	592	*ApoB*	
24-week total cholesterol	264	9	47.8-70.0	Significant	391	NA	
24-week glucose	267	12	26.7-75.1	Suggestive	401	NA	
24-week insulin	262	13	5.7-10.4	Suggestive	30	*Klf6, Chrm3*	
Diet-induced atherosclerosis	292	6	122.6-122.7	Significant	6	*Apobec1*	
Benzene-induced genotoxicity	144	10	31.8-34.2	Significant	21	*Gm4794, Sult3a1*	French, 2014
Epigallocatechin gallate sensitivity	272	4	142.6-151.8	ND	46	*Mfn2*	Church, 2014

Table 1.2: Summary of QTL mapped in DO mice to date [102, 103].

SUMMARY OF RESEARCH PRESENTED IN THIS DISSERTATION

Atherosclerosis is clearly a complex disease and there is still a large gap in our understanding of how the disease is regulated. Many advances in our understanding of the complexity of this disease have been made over the last few decades and have been outlined in this introduction. Cumulatively, it is well known that atherosclerosis is determined by a combination of both environmental and genetic risk factors. Epidemiological studies in humans and follow up experiments in laboratory mice have identified several major environmental risk factors, including smoking, high blood pressure, sedentary lifestyle, diabetes, obesity, and poor nutrition. Studies of familial inherited forms of dyslipidemias have implicated both lipid-associated and non-lipid-associated genetic factors influencing susceptibility to atherosclerosis. Mouse studies of atherosclerosis, both diet-induced and gene-targeted approaches, have revealed strain-specific influences on the development of atherosclerotic lesions, also indicating a major role for genetic variation in influencing this complex trait.

To date, over 30 atherosclerosis QTL have been identified in mice and the identification of novel QTL and the genes underlying the associations is likely to be expedited with the use of modern advanced mouse mapping populations. In this study, we performed QTL mapping using used the heterogeneous multi-parent DO mouse resource, which contains 37.8 million SNPs and 6.9 million insertions, deletions, and structural variants that are uniformly distributed throughout the genome. The DO mice are designed to be highly informative for QTL mapping due to the introduction of allelic variants from multiple founder strains and the accumulation of meiotic recombination events resulting in small haplotype blocks that provide high mapping resolution. Each DO animal is a mosaic of the eight founder strains used to create the DO mice. We identify

here several QTL associated with atherosclerosis and related traits in this cohort of DO mice.

SIGNIFICANCE OF FINDINGS PRESENTED

In Chapter 2, we present the first study to date to characterize atherosclerosis in the DO mice. We found that DO mice fed a HFCA diet are moderately susceptible to diet-induced lesions and that these mice exhibit a large amount of variation in both lesion size and number of lesions. We were able to identify a highly significant QTL on Chromosome 6 associated with variation in lesion size within the region of the previously identified *Ath37* locus. The QTL we identified was less than 1 Mb and contained only 6 genes, none of which were previously implicated as causal genes underlying the *Ath37* locus. We observed that alleles within the locus contributed by the founder strain A/J are specifically associated with larger lesion size.

Based on recent reports, an A/J-specific isoform of *Apobec1* has been identified that exhibits differential editing efficiency of *Apobec1* targets. Our results suggest that this A/J-specific isoform of *Apobec1* could regulate lesion size through editing of target mRNAs, with ApoB as the most likely functional target. These findings indicate that the DO mice present a valuable resource for studying the complex genetic architecture of atherosclerosis and provide a foundation for the use of DO mice in subsequent studies of atherosclerosis and cardiovascular disease risk factors.

In Chapter 3, we investigate genetic regulation of the atherosclerosis-associated metabolite TMAO in the DO mice. Prior to this study, it was known that hepatic FMO3 oxidizes TMA to TMAO and modulation of *Fmo3* expression in the liver can alter TMAO levels. Additionally, one mouse study identified a QTL on Chromosome 1 in HDMP mice as contributing to variation in TMAO levels across the ~100 inbred strains and RILs used to construct the HDMP, but otherwise, very little is known about the role of host genetics in the regulation of TMAO.

The study we present here represents the first to characterize genetic regulators of TMAO levels in the DO mice. We identified a QTL on Chromosome 12 and a QTL on Chromosome 14 as associated with TMAO in the DO mice. While the allele effects we observed underlying the Chromosome 14 QTL indicated there may be multiple alleles regulating variation in TMAO levels at that locus, the Chromosome 12 QTL clearly indicated that contribution of alleles from the CAST/EiJ and PWK/PhJ strains at the locus were associated with lower TMAO levels. Based on these founder allele effects, we were able to identify one gene, *Numb*, that exhibits expression matching this strain distribution pattern. We measured mRNA expression of *Numb* in the DO mice and identified a highly significant *cis*-eQTL. Follow up studies are needed to determine whether there is a causal relationship between *Numb* and TMAO.

Considering the accumulation of evidence that TMAO levels are predictive of CVD risk, the identification of genetic regulators of TMAO will be important to understand the biological basis of these associations. Although the mechanism has yet to be elucidated, we hypothesize that the candidate gene *Numb* may be regulating TMAO and may represent a novel pathway by which the atherosclerosis-associated metabolite is being regulated by host genetics. While *Numb* has not yet been functionally linked to atherosclerosis, we believe that this may represent a new pathway to the development and progression of atherosclerotic lesions.

In Chapter 4, we highlight additional findings of QTL associated with other metabolic measures in the DO mice. Specifically, we describe a novel locus on Chromosome 1 that we have identified as associated with BUN levels which we were able to resolve at the SNP level. We identified a single intronic SNP in *Esrrg* that matches the founder allele effects at this locus. These results and the known biological role of *Esrrg* in kidney development suggest that this gene is likely to be the gene contributing to variation in BUN levels. The identification of such a

refined interval associated with this phenotype demonstrates the power of the DO mice to expedite candidate gene identification and the determination of the functional significance between *Esrrg* and BUN levels warrants further study.

Among our significant results, we also identified a Chromosome 8 QTL associated with HOMA-IR after dietary treatment that contains 42 positional candidate genes including *Fgfr1* which we identify as a high priority candidate gene based on its known biological function in the maintenance of glucose homeostasis. We report an additional 21 QTL that reached suggestive significance. Phenotyping additional DO mice for traits exhibiting such suggestive QTL that are trending toward genome-wide significance is one approach to assess the validity of suggestive QTL prior to conducting follow up studies.

CHAPTER 2[1]: HIGH-RESOLUTION GENETIC MAPPING IN THE DIVERSITY OUTBRED MOUSE POPULATION IDENTIFIES *APOBEC1* AS A CANDIDATE GENE FOR ATHEROSCLEROSIS

CHAPTER OVERVIEW

Inbred mice exhibit strain-specific variation in susceptibility to atherosclerosis and dyslipidemia which renders them useful in dissecting the genetic architecture of these complex diseases. Traditional quantitative trait locus (QTL) mapping studies using inbred strains often identify large genomic regions, containing many genes, due to limited recombination and/or sample size. This hampers candidate gene identification and translation of these results into possible risk factors and therapeutic targets. An alternative approach is the use of multi-parental outbred lines for genetic mapping, such as the Diversity Outbred (DO) mouse panel, which can be more informative than traditional two-parent crosses and can aid in the identification of causal genes and variants associated with QTL.

We fed 292 female DO mice either a high-fat, cholesterol-containing (HFCA) diet, to induce atherosclerosis, or a low-fat, high-protein diet for 18 weeks and measured plasma lipid levels before and after diet treatment. We measured markers of atherosclerosis in the mice fed the HFCA diet. The mice were genotyped on a high density SNP array and founder haplotypes were reconstructed using a hidden Markov model. The reconstructed haplotypes were then used to perform linkage mapping of atherosclerotic lesion size as well as plasma total cholesterol,

[1] A version of this work was previously published as: Smallwood TL, Gatti DM, Quizon P, Weinstock GM, Jung KC, Zhao L, Hua K, Pomp D, Bennett BJ. High-Resolution Genetic Mapping in the Diversity Outbred Mouse Population Identifies *Apobec1* as a Candidate Gene for Atherosclerosis. G3 (Bethesda). 2014 Oct 23;4(12):2353-63. doi: 10.1534/g3.114.014704.

triglycerides, insulin, and glucose. Among our highly significant QTL we detected a ~100 kb QTL interval for atherosclerosis on Chromosome 6, as well as a 1.4 Mb QTL interval on Chromosome 9 for triglyceride levels at baseline and a coincident 22.2 Mb QTL interval on Chromosome 9 for total cholesterol after dietary treatment. One candidate gene within the Chromosome 6 peak region associated with atherosclerosis is *Apobec1*, the apolipoprotein B (ApoB) mRNA editing enzyme, which plays a role in the regulation of ApoB, a critical component of LDL, by editing ApoB mRNA. This study demonstrates the value of the DO population to improve mapping resolution and to aid in the identification of potential therapeutic targets for cardiovascular disease. Using a DO mouse population fed a HFCA diet we were able to identify an A/J-specific isoform of *Apobec1* that contributes to atherosclerosis.

INTRODUCTION

Identifying novel gene(s) and pathways that regulate susceptibility to atherosclerosis and its underlying risk factors is critically important for biomedical science. Genome-wide association studies (GWAS) have aided in the identification of more than 50 genomic loci associated with atherosclerosis [104] and plasma lipid levels [105, 106]. However, elucidation of the mechanisms underlying novel loci resulting from GWAS studies is challenging and in most cases such loci only explain a fraction of the genetic variance related to the trait [107-109]. For example, GWAS have been successfully used to identify more than 40 loci for coronary artery disease (CAD), but these loci in aggregate explain less than 10% of the phenotypic variance associated with CAD [110].

Mouse models are well suited for studies of complex traits such as atherosclerosis because environmental conditions and other confounding co-factors can be carefully controlled. This reduces the effects of environmental variation on clinical traits and thus increases the

portion of the variance that can be explained by genetics. Recently, a "*Multiparent Advanced Generation Inter-Cross*" (MAGIC) population was developed from 8 inbred strains of mice and is referred to as the Diversity Outbred (DO) mouse population [97]. The DO mice are mosaics of C57BL6/J, A/J, NOD/ShiLtJ, NZO/HiLtJ, WSB/EiJ, CAST/EiJ, PWK/PhJ and 129S1/SvImJ and these mice complement another large endeavor called the Collaborative Cross (CC) [111]. Both the DO and CC populations have tremendous genetic diversity, containing approximately 45 million segregating SNPs [112]. The DO population is maintained by a randomized outbreeding strategy and thus is an ideal resource for high-resolution genetic mapping due to the high allelic diversity and increased number of recombinations represented by this population as compared to typical inbred mouse strains [98].

In the current study, we utilize DO mice to study plasma lipids and atherosclerosis. We identify a significant QTL for atherosclerosis with a sub-megabase resolution within a locus previously identified as associated with atherosclerosis in mice. Using publically available expression data and eQTL analysis, we identify *Apobec1* as a high probability candidate gene. We confirm the eQTL for *Apobec1* in the current study using quantitate real time PCR and quantitate possible co-regulation of circulating apolipoprotein B (ApoB) protein levels at this locus.

MATERIALS AND METHODS

Animals and Diets

Female Diversity Outbred mice (n = 292; J:DO, JAX stock number 009376) were obtained from the Jackson Laboratory (Bar Harbor, ME) as 146 full sibling pairs at 4 weeks of age and at outbreeding generation 11 (G11) (received September, 2012). The mice were group housed (n = 5 mice per cage) with non-irradiated pine bedding and provided with HEPA-filtered

air and free access to sterile water in a climate-controlled facility under a 12 hour light:dark cycle. Mice were maintained on a defined synthetic diet, AIN-76A, until 6 weeks of age to control for differences due to variable components of standard chow (D10001, Research Diets, New Brunswick, NJ); subsequently, 146 mice were transferred to a synthetic high-fat, cholic acid (HFCA) diet, containing 20% fat, 1.25% cholesterol, and 0.5% cholic acid, to induce atherosclerotic lesions and 146 mice were maintained on a high protein diet containing 5% fat and 20.3% protein which is not atherogenic (D12109C and D12083101, respectively, Research Diets, New Brunswick, NJ). One sibling from each of the 146 sibling pairs was randomly assigned to each one of the diets, **Figure 2.1**. The composition of all diets is listed in **Table 2.1.** The source of fat from the diets varied between the baseline diet (corn oil) fed to the mice from 4 to 6 weeks of age and the dietary treatment groups (soybean oil plus cocoa butter) fed to the mice from 6 to 24 weeks of age. All mice were maintained on their respective diets until 24 weeks of age, for a total of 18 weeks. All procedures were approved by the IACUC at UNC Chapel Hill (IACUC Protocol Number 11-299).

Plasma clinical chemistries

In order to ensure that there were no spurious effects due to the potential variable composition of standard laboratory chow, these mice were maintained on a defined synthetic diet for 2 weeks (from 4-6 weeks of age). At 6 weeks of age, the mice were anesthetized using isoflurane and blood was collected after 4 hours of fasting into EDTA-containing microtubes from the retro-orbital sinus. This was repeated at 24 weeks of age, after 18 weeks of dietary treatment. Blood samples were centrifuged for 10 minutes at 10000 rpm at 4°C and stored at -80°C. The plasma levels of total cholesterol, glucose, and triglycerides were quantitated using a Biolis 24i Analyzer (Carolina Liquid Chemistries, Winston-Salem, NC). Insulin was determined

using the Alpco Mouse Ultrasensitive Insulin ELISA assay (Alpco Inc, Salem, NH); samples and controls were run in duplicate and optical densities were measured at 450 nm using a microplate reader and analyzed with Gen5 Data Analysis Software (Bio-Tek, Winooski, VT). Data are presented as mean ± SD and significance was determined using a Student's t-test.

Atherosclerotic lesion size

Hearts, including the proximal aorta, were carefully dissected from 24-week-old mice, perfused with 1X PBS, and stored in 10% formalin at 4˚C. A transverse cut, parallel to the atria, was made to remove the top portion of the hearts which were then embedded in Optimal Cutting Temperature compound (O.C.T.) and stored at -80˚C. Serial sections (10 µm thick) from the aortic sinus were mounted onto slides. Cross sections were arranged 8 to a slide and approximately 10-14 total slides were obtained from each mouse heart, for a total of 80-104 cross sections through 400-560 µm of the aortic sinus per animal. Sections were stained with Oil Red O to measure lipid accumulation. Briefly, slides were fixed with formalin for 15 minutes, rinsed with water then 60% isopropanol, stained for 1 hour in freshly prepared Oil Red O solution, and rinsed with 60% isopropanol then water. Once stained, the slides were imaged using a Zeiss AxioCam MR3 (Zeiss, Munich, Germany) and atherosclerotic lesion area was quantified using Image J software (http://imagej.nih.gov/ij). Data are presented as the average lesion area in μm^2.

***Apobec1* mRNA isoform expression**

QPCR was performed in triplicate using a High Capacity Reverse Transcriptase Kit (Applied Biosystems, Foster City, CA). Following cDNA conversion, 1 µL of sample cDNA, 2 µL KAPA Sybr Fast qPCR mastermix (KK4610), and 0.5 µM primers was added to each well of a 384-well plate. *Apobec1* primers were custom designed and ordered from Eurofins MWG

Operon (Huntsville, AL). The long transcript was amplified using the following primer set 5'-cagcggtgtgactatccaga-3' (left primer) and 5'-ttggccaataagcttcgttt-3' (right primer). This primer set was designed to recognize the known *Apobec1* transcript *Apobec1-001* (ENSMUST00000112586.1). The short transcript was amplified using the following primer set 5'-cccatgagcgttggattc-3' (left primer) and 5'-tcaaccacgggcagtctt-3' (right primer). This primer set was designed to recognize the known *Apobec1* transcript *Apobec1-004* (ENSMUST00000143356.1). A serial dilution of pooled samples was used to create a standard curve. Sterile water was used as a negative control.

Quantification of ApoB in the DO mice

Total ApoB protein levels were measured using the mouse Apolipoprotein B Sandwich-ELISA method in 96-well format using the ApoB ELISA kit from Elabscience (Wuhan, China) in duplicate following the manufacturer's instructions. Optical densities were measured at 450 nm using a microplate reader and analyzed with Gen5 Data Analysis Software (Bio-Tek, Winooski, VT).

Differential expression of genes in peak regions

Data were obtained from The Jackson Laboratory Gene Expression Strain Survey and used for the analysis of differential expression of candidate genes from liver tissue of female C57BL6/J, A/J, NOD/ShiLtJ, NZO/HiLtJ, WSB/EiJ, CAST/EiJ, PWK/PhJ and 129S1/SvImJ mice (http://cgd.jax.org/gem/strainsurvey26/v1). Mice were maintained for 11 weeks on standard chow diet (4% fat content). Genes were identified as differentially expressed between the DO founder strains using analysis of variance and significant between strains differences were calculated using Tukey's Post Hoc test. We used a Bonferroni correction to determine statistical significance and correct for multiple comparisons.

Genotyping

DNA was extracted and purified from tail biopsies taken from 6-week-old mice using Qiagen DNeasy kit according to the manufacturer's instructions. Genotyping was performed using the Mega Mouse Universal Genotyping Array (MegaMUGA) by GeneSeek (Neogen, Lansing, MI) [113]. The call rate exclusion criteria was set at >95% and twelve mice were excluded based on this criteria; the average call rate of the genotyped mice used in the study was 98%. The MegaMUGA array is built on the Illumina Infinium platform and contains 77,808 SNP markers that are distributed throughout the genome at an average spacing of 33 Kb. The MegaMUGA SNPs were subset to include 57,977 informative SNPs that distinguish among the genotypes of the eight founder strains. For the mapping, genomes were reconstructed based on the X and Y allele intensities from the array and founder haplotypes were reconstructed using a hidden Markov model. The founder allele dosages based on the reconstructed haplotypes were then used to perform linkage mapping.

QTL Mapping

QTL mapping was performed using the R package DOQTL version 0.99. Briefly, DOQTL reconstructs the genome in terms of founder haplotypes and performs QTL mapping by regressing the phenotypes on the founder haplotypes with an adjustment for kinship between the mice. Phenotypes were natural log-transformed to satisfy the model assumption of a normal distribution. Diet was included as an additive covariate for post diet measurements, except lesion size which was performed using only the subset of mice fed the HFCA diet (n=146). Candidate genes were identified by position based on the Wellcome Trust Sanger mouse genomes database, www.sanger.ac.uk, release 1303 based on genome assembly GRCm38 [114]. QTL support

intervals were defined by the 95% Bayesian credible interval, calculated by normalizing the area under the QTL curve on a given chromosome [115].

The mapping statistic reported is log of the odds ratio (LOD). The significance thresholds were determined by performing 1000 permutations of genome-wide scans by shuffling phenotypic data in relation to individual genotypes. Significant QTL were determined at a genome-wide p-value of < 0.05 and suggestive QTL were determined at a p-value of < 0.63. The latter corresponds to one false positive per genome scan [116].

RESULTS

Effects of Diet on Clinical Markers of Cardiovascular Disease in Diversity Outbred Mice

Our studies using the DO mice were designed to examine nutrigenetic or gene x diet interactions and the overall design is outlined in **Figure 2.1**. Thus, we utilized several diets that we outline here. First, in order to standardize our measurements for this study and our planned future studies, all of the DO mice were placed on a synthetic, chemically-defined diet (AIN-76) upon arrival at the University of North Carolina, Chapel Hill. This diet is henceforth referred to as the baseline diet and was administered for 2 weeks. After two weeks of this defined diet, all 292 DO mice were phenotyped for a variety of clinical traits associated with cardiovascular disease risk. These data were used for QTL mapping. At the end of the two week baseline diet, one mouse from each sibling pair was assigned to one of two diets for subsequent investigation of gene x diet interactions. These diets were also synthetic and chemically defined and included a high-fat, cholic acid diet designed to induce atherosclerosis and a high protein diet not expected to be atherogenic. The DO mice were fed these diets for 18 weeks and then phenotyped at 24 weeks of age.

There was considerable variation among the mice in the clinical traits measured while the mice were on the baseline diet, **Table 2.2**. After diet treatment, there was a robust and statistically significant increase in plasma cholesterol in response to the HFCA diet (199.9 ± 68.6 mg/dl) compared to both the high protein diet (91.7 ± 25.1 mg/dl) and baseline levels (97.6 ±31.5 mg/dl), **Figure 2.2 and Table 2.2**. Conversely, triglyceride levels decreased in response to the HFCA diet (32.3 ± 12.2 mg/dl) compared to both baseline levels (59.3 ± 26.7 mg/dl, $p<0.05$) and in mice on the high protein diet (57.7 ± 30.8 mg/dl, $p<0.05$), **Figure 2.2 and Table 2.2**.

Identifying QTL for Clinical Markers of Cardiovascular Disease in the DO at baseline

We performed QTL mapping for several clinical markers of cardiovascular disease in 6-week-old mice fed the baseline (AIN-76) diet. Significant QTL were determined at a genome-wide p-value of <0.05 and the QTL support interval was defined using the 95% Bayesian credible interval, **Table 2.3**. We identified a highly significant QTL for triglycerides on Chromosome 9 with a peak SNP located at 51.4 Mb (LOD=11.3; n= 262 mice), **Figure 2.3A**. DO mice carrying the CAST/EiJ allele at the Chromosome 9 QTL have higher triglyceride levels, **Figure 2.4A**. There are 34 candidate genes within the QTL interval on Chromosome 9, **Figure 2.4B**. We identified suggestive QTL for total cholesterol at baseline on Chromosome 13 with a peak SNP located at 30.4 Mb (LOD=6.5; n= 277 mice), **Figure 2.3B,** and for glucose at baseline on Chromosomes 5 and 7 with peak SNPs located at 92.6 Mb on Chromosome 5 (LOD=7.0; n= 257 mice) and 27.2 Mb on Chromosome 7 (LOD=6.5; n= 257 mice), **Figure 2.3C**.

Identifying QTL for Clinical Markers of Cardiovascular Disease in the DO after dietary treatment

To assess the effects of diet in the DO, we determined the genetic architecture of several markers of cardiovascular disease by quantifying these phenotypes at 24 weeks of age after dietary treatment. Because diet had significant effects on the phenotypes, it was included in the mapping model as an additive covariate. We observed a significant QTL for plasma cholesterol on Chromosome 9 with a peak LOD score of 7.54 at 48.3 Mb (47.84 Mb- 70.04 Mb), **Figure 2.5A and Table 2.3**. The CAST/EiJ founder haplotype was associated with high levels of cholesterol based on the founder allele contribution in the region, **Figure 2.5B**. This is a relatively large QTL interval of 22.2 Mb based on the Bayesian credible interval (p= 0.95) and the entire interval contains 391 known and predicted genes. The peak SNP is within 3 Mb of the peak SNP identified by mapping of baseline triglyceride levels in this population of mice, suggesting that this genes at this locus are critical in the overall regulation of lipid metabolism.

Differential expression of candidate genes in DO founder strain mice

In order to prioritize candidate genes within the QTL interval we identified as associated with triglycerides, we analyzed hepatic gene expression among the founder strains for the 34 genes in the region from a publically available dataset (http://cgd.jax.org/gem/strainsurvey26). We first identified genes whose expression varies across the progenitor strains using 1-way ANOVA analysis and using a Bonferroni correction based on the number of probes tested (0.05/68 probes representing the 34 genes in the locus). Based on this analysis, we identified 10 probes representing 9 genes that are differentially expressed: *1110032A03Rik, 1810046K07Rik, Alg9, Bco2, Cryab, Gm684, Ppp2r1b, Pts, and Sik2*, **Figure 2.6A-I**.

We can identify the contribution of the founder alleles to the QTL using the founder coefficients from the QTL mapping. For the Chromosome 9 triglyceride QTL, we observed that allelic contribution from the CAST/EiJ founder at this locus is associated with higher triglyceride levels, while comparatively low triglyceride levels are associated with allelic contribution from the other seven founder strains, **Figure 2.4A**. Thus, we next performed pair-wise comparisons using Tukey's HSD for the 9 differentially expressed genes to identify genes which are differentially expressed in CAST/EiJ compared to the other founder strains and thus match the overall allelic effects at the QTL. Of the 9 genes differentially expressed, *Bco2* and *Ppp2r1b* most closely match the allele effects of the QTL and are differentially expressed between CAST/EiJ and the other progenitor strains, **Figure 2.6A and B**. We note that the locus we identified is adjacent to a region harboring genes known to influence plasma lipid levels including: *Apoa1, Apoa4, Apoa5* and *Nnmt*. Both *Apoa4* and *Nnmt* are differentially expressed among the strains, but only *Nnmt* is differentially expressed in CAST/EiJ mice, **Figure 2.7**.

Identification of *Apobec1* as a candidate for regulating lesion size in DO mice

Atherosclerosis is a complex trait and studies using mice have identified numerous QTL for this trait [69] but this trait has not yet been evaluated in the DO mice. Based on previous studies, a high-fat diet containing cholic acid (HFCA) can induce formation of atherosclerotic lesions [62]. Indeed, we found that none of the mice fed the high protein diet exhibited any lesions and our subsequent analyses focus only on the subset of DO mice that were fed the HFCA diet.

We found that 76% of the DO mice were susceptible to atherosclerosis with lesions induced by the HFCA diet, with a range in lesion size from 38 to 33,200 μm^2. Based on this highly variable phenotype, we were able to identify a highly significant QTL on Chromosome 6

(LOD=10.7; 122.6 - 122.7 Mb), **Figure 2.8A**. Although our Chromosome 6 peak SNP, UNC11996440, is within the 95% confidence interval of a previously reported QTL, *Ath37*, the interval is refined to sub-Mb resolution using the DO. The 95% confidence interval reported for this previously defined QTL encompasses an 11.8 Mb interval, while our Chromosome 6 atherosclerosis peak maps to a refined 100 kb region containing 6 genes, **Figure 2.8C**. One candidate gene within the Chromosome 6 peak region is *Apobec1*, the apolipoprotein B mRNA editing enzyme, which plays a role in the regulation of ApoB, a critical component of LDL, by editing *ApoB* mRNA to produce the short ApoB48 isoform. Misregulation of *Apobec1* results in altered *ApoB* isoform editing. For example, *Apobec1*$^{-/-}$ mice have higher levels of ApoB100 compared to the edited isoform, ApoB48 [117] and transgenic rescue of *Apobec1* in *Apobec1*$^{-/-}$ animals has been shown to directly alter chylomicron production [118]. *Apobec1* is an attractive candidate gene for influencing diet-induced lesion size based on the known function of *Apobec1* in lipid homeostasis and a causal allele for *Apobec1* has not been previously identified as associated with atherosclerosis. Additionally, we found that total cholesterol levels after dietary treatment were significantly correlated with both the short and long isoforms of *Apobec1* (r=0.6 and r=0.57, respectively), but not ApoB levels in these mice (r= 0.06).

A/J specifically expresses a long isoform of *Apobec1* and this is induced in response to a high-fat, cholic acid diet

We found that DO mice containing the A/J allele at the Chromosome 6 QTL had larger aortic lesions, **Figure 2.8B**. Among genes in the Chromosome 6 QTL interval, only *Apobec1* has significant differential hepatic expression in the A/J strain, which expresses higher levels of *Apobec1* ($p < 0.0125$), **Figure 2.8D-G.** Interestingly, strain-specific differences in isoform expression of *Apobec1* were recently identified between A/J and C57BL/6J mice, where

C57BL/6J mice express a truncated Apobec1 protein [119]. More importantly, using AXB and BXA recombinant inbred strains they demonstrate genetic regulation of mRNA editing by *Apobec1*, such that animals expressing the A/J-specific isoform of *Apobec1* exhibit higher editing efficiency of Apobec1 targets as compared with C57BL/6J animals.

We quantitated the expression of *Apobec1* in liver tissue from A/J and C57BL/6J mice fed either the HFCA diet or the control synthetic diet and found that on the synthetic diet A/J expresses more of the long isoform of *Apobec1* compared to C57BL/6J, **Figure 2.9B**, while A/J and C57BL/6J express similar levels of the short isoform of *Apobec1* while on a synthetic diet, **Figure 2.9A**. In response to the HFCA diet, expression of both isoforms of *Apobec1* is significantly increased in both A/J and C57BL/6J mice compared to *Apobec1* levels in the mice on the synthetic diet ($p > 0.05$). The short isoform of *Apobec1* is induced 2.5-fold in C57BL/6J animals and 3.5-fold in A/J animals in response to the HFCA diet. The long *Apobec1* isoform is induced 2.5-fold in C57BL/6J animals and 4-fold in A/J animals.

Cis-eQTL for _Apobec1_ exhibit isoform-specific allele effects patterns in the DO mice

Based on the differences in expression of *Apobec1* between A/J and C57BL/6J mice, we hypothesized that the expression of *Apobec1* may be genetically regulated. Thus, we surveyed two publically available databases for expression QTL data to determine if a natural variant near the *Apobec1* gene affects its expression. Specifically, we queried a panel of inbred strains of mice called the hybrid mouse diversity panel (HMDP) and The Jackson Laboratory's Diversity Outbred eQTL viewer, data located at http://systems.genetics.ucla.edu/data/hmdp and http://cgd.jax.org/apps/eqtlviewer-beta/, respectively [83]. *Apobec1* expression varied significantly among the HMDP strains and the major locus regulating expression mapped directly over the *Apobec1* gene, at 122 Mb on Chromosome 6 (data not shown). Similarly, the

Jackson Laboratory's DO eQTL viewer identifies a *cis*-eQTL associated with *Apobec1* expression, **Figure 2.10A.** Similar to our results, these data demonstrate that the A/J allele is associated with higher expression of *Apobec1*, **Supplemental Figure 2.10B**.

Considering our observations of the effect of genetic background on *Apobec1* expression in liver tissue and recent observations that mouse genetic background regulates isoform usage in macrophages, we next asked if there was a difference in the genetic regulation of each of these isoforms among the DO founder strains. We performed QTL mapping of mRNA expression levels of either the short or long isoform of *Apobec1* from RNA isolated from liver tissue from the present DO mouse study population. We identified highly significant *cis*-eQTL for both isoforms on Chromosome 6 with the peak SNP located at 121.8 Mb with a maximum LOD score of 9.9 associated with expression of the short isoform ($p<0.05$, n= 252 mice) and at 123.4 Mb with a maximum LOD score of 15.7 associated with expression of the long isoform ($p<0.05$, n= 251 mice), **Figure 2.11A and 2.11B**, respectively. When we estimate the effect of each founder at each marker along Chromosome 6, we see that CAST/EiJ alleles are associated with higher expression of the short isoform, **Figure 2.11C**, while A/J alleles are associated with higher expression of the long isoform, **Figure 2.11D**. We also mapped the ratio of *Apobec1* isoforms and observe a highly significant *cis*-eQTL with a max LOD score of 41.0 at 122.6 Mb ($p<0.05$, n= 251 mice), **Figure 2.11E and 2.11F**.

***Apobec1* and ApoB levels are dependent on the genotype of UNC11996440, the Chromosome 6 peak SNP associated with atherosclerosis.**

Based on the association of the A/J founder with atherosclerotic lesion size, we hypothesized that a causal allele located in the *Apobec1* sequence would be private to A/J. However, there are no documented SNPs that are private to A/J within the refined peak region of

122.6-122.7 Mb on Chromosome 6. The peak SNP for this locus, UNC11996440, is located distal to the *Apobec1* gene on Chromosome 6 at 122.67 Mb, while *Apobec1* lies between 122.5-122.6 Mb. In addition to the *cis*-eQTL data for *Apobec1* expression, this suggests that a distant SNP affecting *Apobec1* expression may be responsible for the association between the A/J haplotype and atherosclerotic lesion size. However, if the association of the A/J haplotype with increased lesion size in the DO mice is mechanistically related to *Apobec1* expression, then we would expect *Apobec1* expression levels to differ based on the genotype of UNC11996440. We identified T as the major allele (frequency = 0.69) and C as the minor allele (frequency = 0.31) at the tagging SNP UNC11996440. The genotype at this SNP significantly affects *Apobec1* mRNA expression levels ($p > 0.001$), **Figure 2.12A**. We next hypothesized that the association at this locus may be mechanistically related to the role of Apobec1 in editing its primary target ApoB, as elevated plasma ApoB is a known risk factor for atherosclerosis. Therefore, we quantified plasma ApoB levels in duplicate using two mouse-specific ApoB Sandwich-ELISA plates (N=80 mice). Plasma samples from only the high-fat, cholic acid diet-fed mice were used and samples were chosen to represent a range of atherosclerotic lesion sizes. As shown in **Figure 2.12B**, ApoB levels are significantly different between the genotype groups at the tagging SNP UNC11996440 ($p > 0.001$). Therefore, in addition to identifying the A/J-specific isoform of *Apobec1* as associated with atherosclerotic lesion size in the DO mice, these data suggest that this association may be attributable to the role of *Apobec1* in editing ApoB.

DISCUSSION

The DO was designed to be a high-resolution genetic mapping panel. The use of DO mice for mapping offers a number of advantages over classical approaches to mouse genetics, including high mapping resolution, increased heterozygosity, and uniformly distributed genetic

variation across the genome [98]. Using the newly developed DO mouse resource, we were able to refine the positions and identify new candidate genes for previously mapped QTL for atherosclerosis, a highly complex phenotype associated with human disease. We further interrogate this QTL using publically available data, by quantitating mRNA levels for *Apobec1* and by investigating the protein target of Apobec1 to identify a potential mechanism for this QTL. We discuss each of these in detail.

Atherosclerosis in the DO

Atherosclerotic lesion development is the most common cause of cardiovascular disease. Here, we characterize for the first time the development of diet-induced atherosclerotic lesions in the newly developed DO mouse population. In the present study, 292 mice on either a high-fat, cholic acid diet designed to induce atherosclerotic lesions or a non-atherogenic, low-fat, high protein diet were phenotyped for atherosclerosis. We found that none of the mice fed the high protein diet exhibited any lesions. We found that 76% of the DO mice were susceptible to lesions induced by the HFCA diet, with a range in lesion size from 38 to 33,200 μm^2. We detected one locus on Chromosome 6 for atherosclerosis that maps within the 95% CI of a previously reported QTL, *Ath37*, [78] which was identified in studies using sub-congenic mice between C57BL/6J and CAST/EiJ. The locus in the current study contains *Apobec1,* which was originally described as the enzyme responsible for deamination of a cytosine in mature ApoB mRNA, resulting in a premature stop codon and production of the truncated protein ApoB48. Lipoproteins containing ApoB48 are more efficiently cleared from the circulation as demonstrated by adenoviral overexpression studies that reduce plasma lipid levels [120]; in contrast, *Apobec1*$^{-/-}$ mice synthesize only apoB-100 and have increased atherosclerosis when crossed to *Ldlr*$^{-/-}$ mice [55].

Closer examination of the locus we identified here indicates that A/J mice contain a susceptible allele, which is unexpected as A/J has classically been defined as an atherosclerosis-resistant mouse strain [40]. However, it is possible that if *Apobec1* editing is affected by an allele carried by the A/J strain, this effect could be masked by transgressive variation or epistasis in inbred A/J animals. When we queried the Sanger SNP database for alleles private to A/J within 10 kb of the *Apobec1* gene sequence (122,567,890- 122,612,426 Mb), we found that there are no documented SNPs that are private to A/J. However, the association of distant SNPs influencing gene expression of genes as far as 300 kb away is not uncommon [121]. While our data suggests that a distant SNP affecting *Apobec1* expression may be responsible for the association between the A/J haplotype and atherosclerotic lesion size, we have not yet ruled out an association between lesion size and the other five genes in the locus, nor have we ruled out nearby genes with A/J-specific alleles. Additionally, in our study we did not find lesion size to be correlated with *Apobec1* expression ($r = -0.02$) although total cholesterol was correlated with this phenotype ($r = 0.60$).

Recently, Hassan and colleagues have reported that genetic differences between C57BL/6J and A/J at the *Apobec1* gene affect global RNA editing patterns between these strains [119]. Indeed, this study identified an A/J-specific isoform of *Apobec1* in murine macrophages that increased the editing efficiency of this enzyme for multiple *Apobec1* targets [119, 122]. Thus, differences in *Apobec1* structure or expression could affect editing or perhaps overall levels of ApoB or another of Apobec1's target genes. Our initial studies, in a subset of the DO mice used in this study, indicate that total ApoB levels may be influenced by this locus.

Mapping Clinical Markers of Cardiovascular Disease in the DO

Measures of cholesterol, triglycerides, and glucose are commonly used markers of cardiovascular disease. Therefore, we were interested in investigating the genetic architecture of these traits in the DO mice. We identified highly significant QTL for triglycerides in mice at baseline. The significant QTL on Chromosome 9 for baseline triglycerides is 1.4 Mb in size and is coincident with a ~30 Mb QTL previously identified as associated with triglyceride levels in C57BL/6J x KK-Ay/a F2 mice, *Trigq1*. *Trigq1* maps to 61 Mb on Chromosome 9 with a peak LOD score of 4.2 at the marker D9Mit163 [123]. The large *Trigq1* locus includes multiple gene candidates known to regulate lipid levels such as the *ApoA5-ApoA4-ApoA3-ApoA1* gene cluster and *Lipc*. While the *ApoA5-ApoA4-ApoC3-ApoA1* gene cluster is located at 46.2 Mb, just proximal to the QTL interval we identified here and *Lipc* is located 20 Mb distal to our peak region, these genes would seem to be excluded based on our results. However, at least one of these genes, *ApoA5*, has been well characterized as associated with triglyceride levels [124]. In humans, a mutation in *ApoA5*, also called the Delhi gene, is linked to extremely elevated triglyceride levels and increased risk of cardiovascular disease [125, 126]. Hepatic lipase, encoded by the *Lipc* gene, hydrolyzes triglycerides and phospholipids in lipoprotein particles and is therefore also likely to be functionally associated with heart disease. We utilized publically available gene expression data to prioritize genes whose expression may be genetically regulated. These analyses are not able to identify candidate genes that have a functional variant that affects protein function, protein stability, or binding. Further study to confirm that one of the genes in the triglyceride QTL identified in this study is a causal gene remains to be performed.

This region of Chromosome 9 clearly contains multiple genes that are important for the regulation of cardiovascular disease risk factors. We also identified the same region as associated

with total cholesterol after dietary treatment in our DO mouse population. Additionally, the human syntenic region of the Chromosome 9 QTL we identified as associated with both baseline triglycerides and total cholesterol after dietary treatment has been identified in several human GWAS studies as associated with multiple atherosclerosis risk factors across multiple human populations [127-129]. When one QTL contains more than one causal variant, the Bayesian interval calculation may fail to differentiate genetic signals from the closely linked genes. Several studies have suggested the use of 2-LOD support intervals for generating conservative estimates of genes for candidate testing. Indeed, in our study the 95% Bayesian credible intervals represent approximately 1-LOD support intervals for baseline triglycerides and for total cholesterol after diet treatment and if we broaden these intervals to 2-LOD support intervals, the *ApoA5-ApoA4-ApoA3-ApoA1* gene cluster and *Lipc* gene falls within the more conservative interval estimate.

We identified a suggestive QTL for total cholesterol on Chromosome 13 with a peak SNP located at 30.4 Mb (LOD=6.5; n= 277 mice) and for glucose on Chromosome 5 with a peak SNP located at 92.6 Mb (LOD=7.0; n= 257 mice). These suggestive QTL co-localize with previously reported QTL. The Chromosome 13 locus associated with baseline cholesterol levels is just proximal to *Lipq2*, which was identified in backcross between MOLF/EiJ and C57BL/6J mice on with an $Ldlr^{-/-}$ mutation [130]. The suggestive QTL associated with blood glucose on Chromosome 5 is less than 2 Mb from the peak SNP reported for the QTL, *Bglu13*, which was identified in an F2 intercross cross of mutant C3H/HeJ and C57BL/6J carrying the $ApoE^{-/-}$ mutation [131].

In summary, we demonstrate here the use of the Diversity Outbred mice for high resolution mapping of traits related to atherosclerosis. We identify several candidate genes for

triglycerides in mice on a synthetically defined diet. Perhaps the most interesting result is the identification of an atherosclerosis QTL on Chromosome 6 that is 100 kb in size. This locus contains the candidate gene *Apobec1* which is known to alter circulating lipoprotein composition, but also has widespread RNA editing capabilities, perhaps indicating an additional mechanism by which susceptibility to atherosclerosis is regulated.

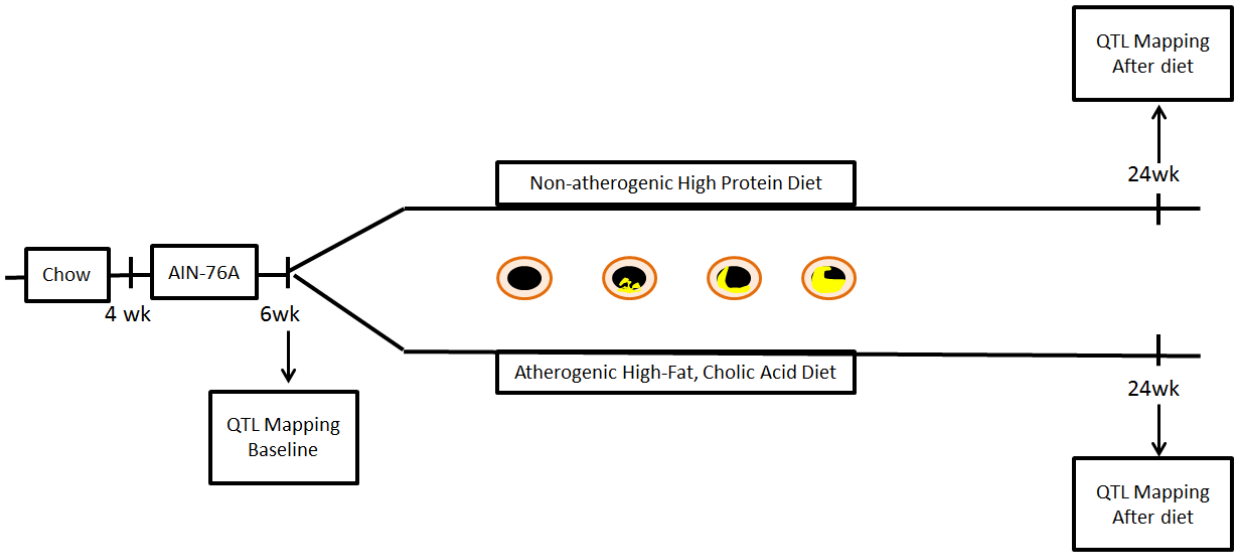


Figure 2.1- Overall design of the QTL mapping study. Mice obtained from the Jackson Laboratory were fed a chow diet of variable composition prior to arriving at the UNC Mouse Facilities at 4 weeks of age. Mice were transferred to a controlled synthetic diet from 4 to 6 weeks of age (AIN-76A). At 6 weeks of age, clinical markers of cardiovascular disease were measured and baseline QTL mapping was performed. The mice were then transferred to one of two diet groups such that one sibling of each sib pair was randomly assigned to either the high-fat, cholic acid diet group designed to induce atherosclerosis or the high protein diet group expected to be non-atherogenic. QTL mapping was then performed in 24-week-old mice after diet exposure. Quantification of atherosclerotic lesions was performed in the mice at 24 weeks of age.

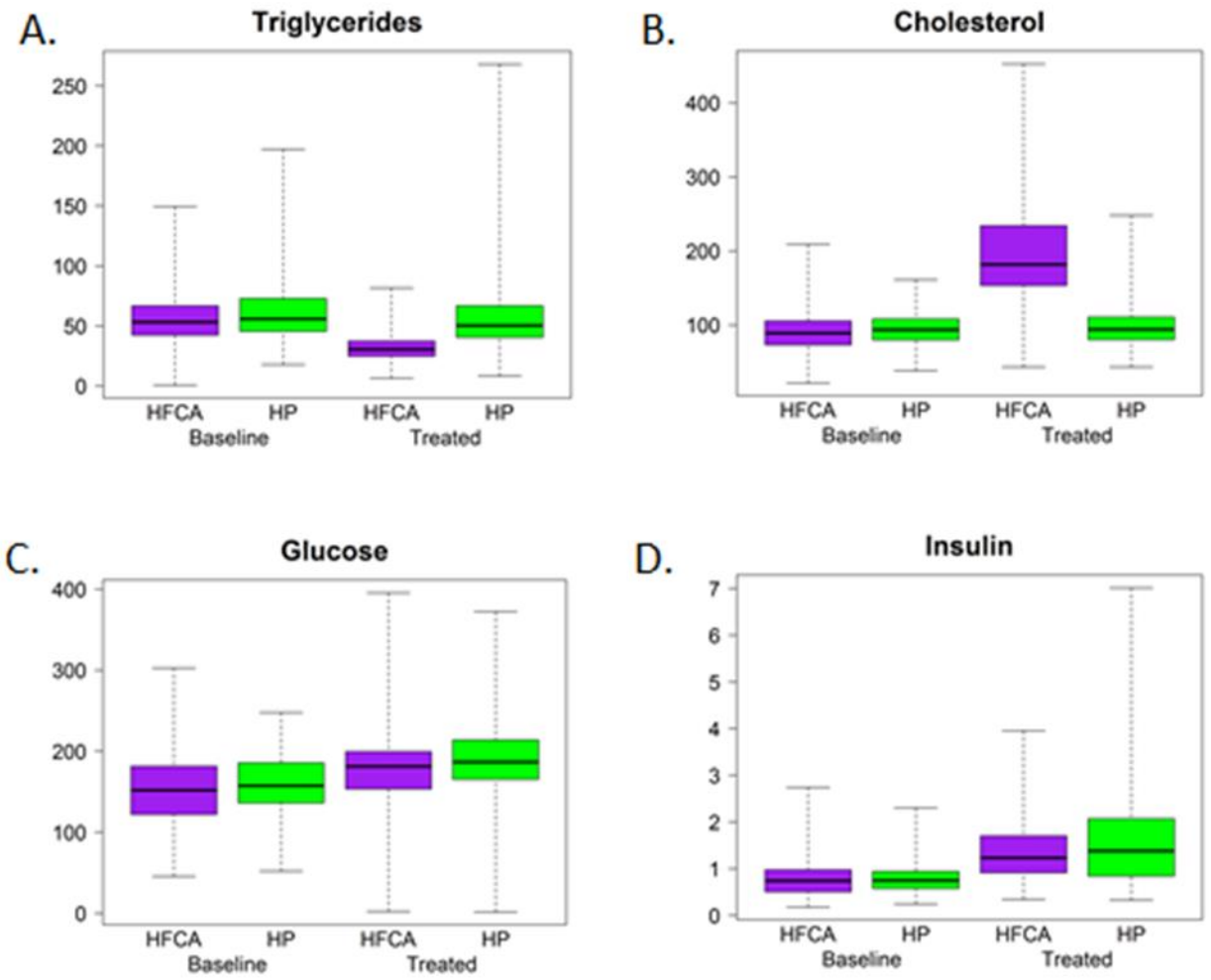
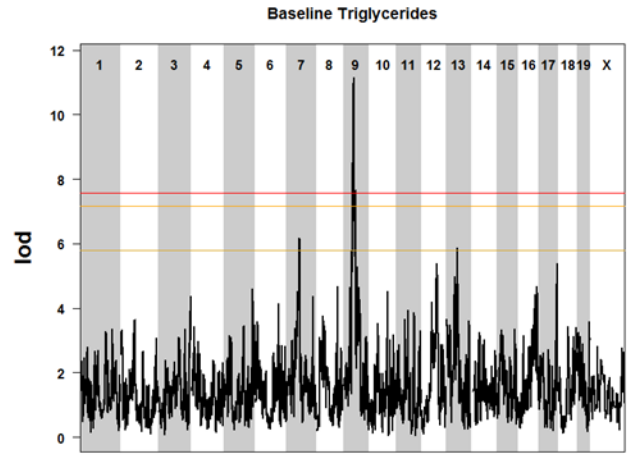
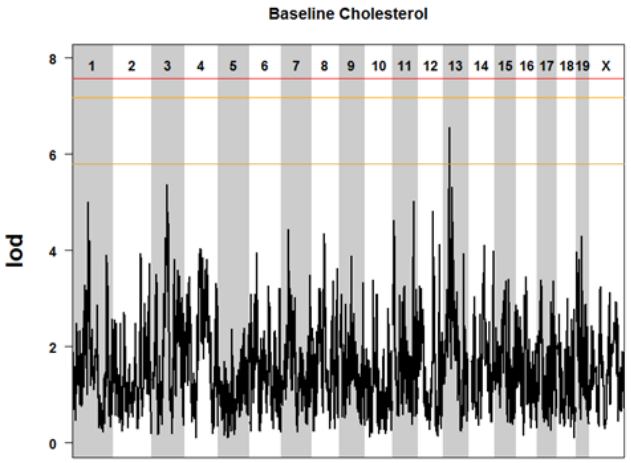


Figure 2.2- Effects of Diet on Cardiovascular Risk Factors in Diversity Outbred Mice. Mice were maintained on a synthetic diet for two weeks, fasted for four hours, and then phenotyped for plasma clinical chemistries at 6 weeks of age (Baseline). Following two weeks of synthetic diet, mice were transferred to either a high protein diet (HP) or an atherogenic diet (HFCA). Plasma was taken from 24-week-old mice after 18 weeks on their respective diets, and with four hours fasting, and then phenotyped for plasma clinical chemistries after diet treatment (Treated).

A.


Baseline Triglycerides
lod
0
2
4
6
8
10
12
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 X

B.


Baseline Cholesterol
lod
0
2
4
6
8
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 X

C.

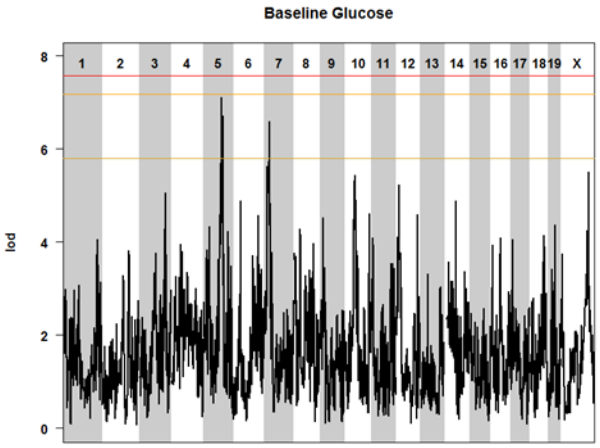


Figure 2.3- QTL mapping of clinical markers of cardiovascular disease in 6-week-old DO mice at baseline. Genome-wide QTL scans for loci affecting plasma levels of triglycerides (A.), total cholesterol (B.), and glucose (C.) in the DO population at baseline. Chromosomes 1 through X are represented numerically on the *x*-axis and the *y*-axis represents the LOD score. The relative width of the space allotted for each Chromosome reflects the relative length of each Chromosome. Mice were maintained on a synthetic diet for 2 weeks and then phenotyped for plasma clinical chemistries at 6 weeks of age. Colored lines show permutation-derived significance thresholds (N=1000) at P = 0.05 (LOD=7.57, shown in red), P = 0.10 (LOD=7.17, shown in orange), and P = 0.63 (LOD=5.79, shown in yellow).

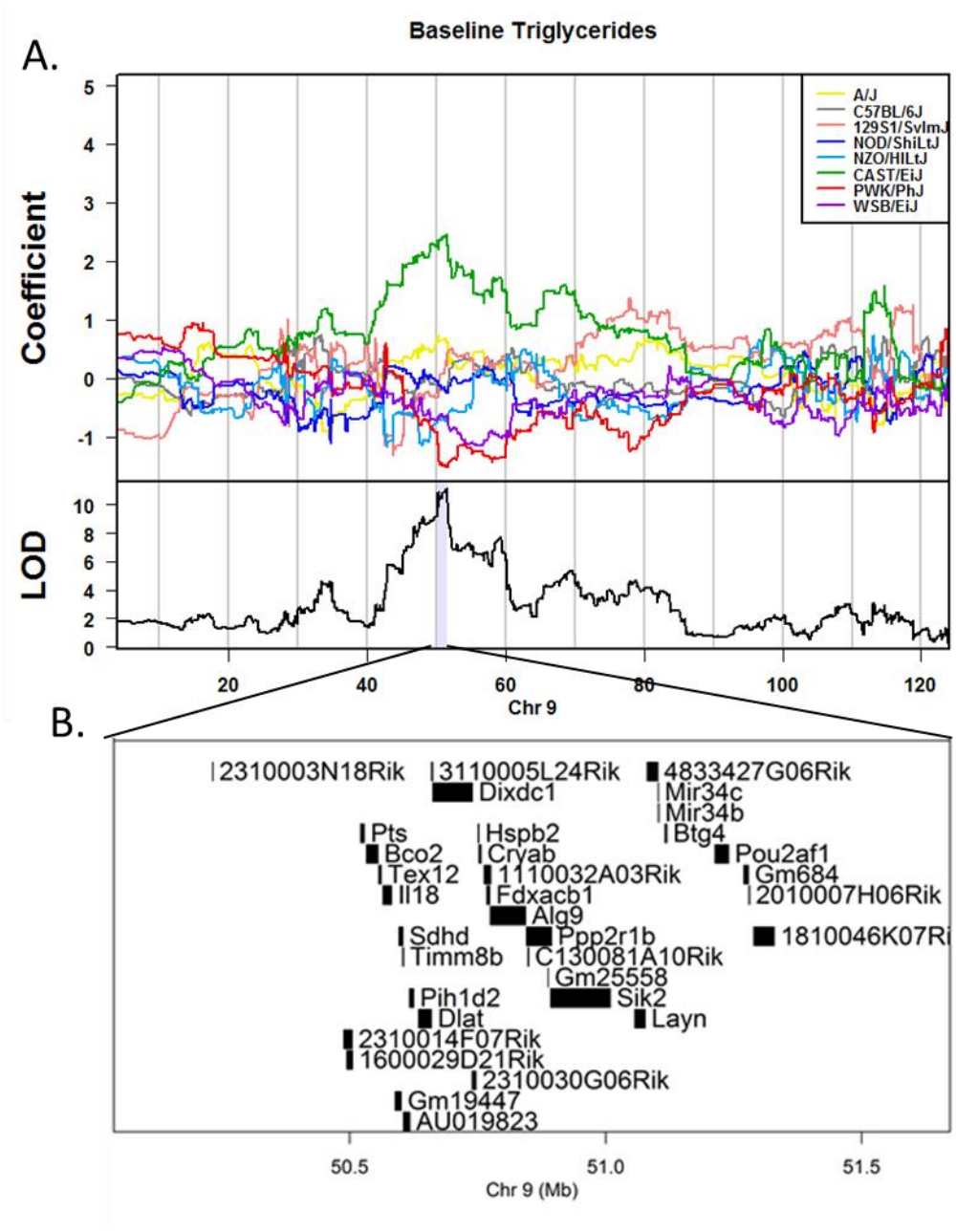


Figure 2.4- High-resolution mapping of significant QTL on Chromosome 9 associated with plasma triglyceride levels. The eight coefficients of the QTL model show the effect of each founder haplotype on the phenotype. The model coefficients for the mapping of baseline triglycerides are plotted for each founder allele at each marker along Chromosome 9 and shading identifies the 95% Bayesian estimated interval around the peak (A). There are 36 potential candidate genes within the Chromosome 9 locus associated with plasma triglycerides at baseline (B).

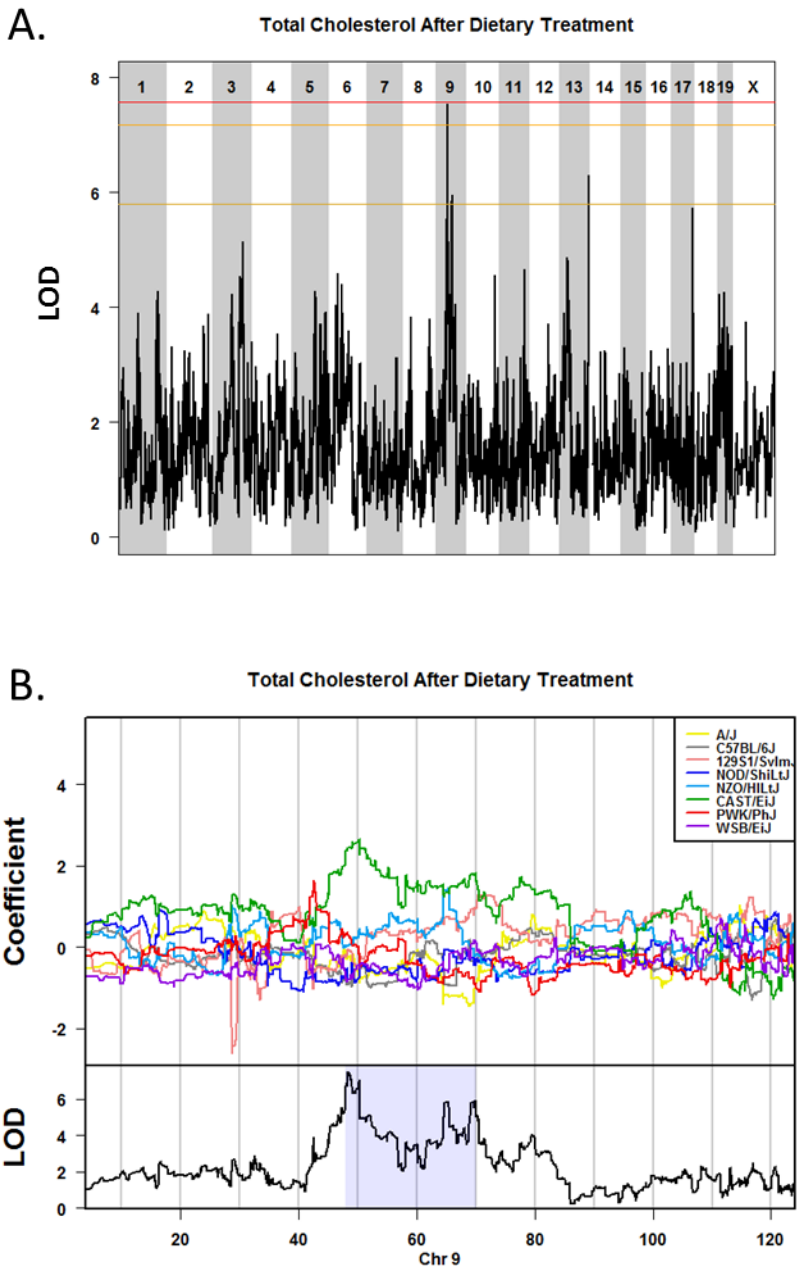


Figure 2.5- QTL mapping of total cholesterol after dietary treatment in the DO mice. Genome-wide QTL scan for loci affecting plasma levels of total cholesterol after 18 weeks of dietary treatment (A). Chromosomes 1 through X are represented numerically on the *x*-axis and the *y*-axis represents the LOD score. The relative width of the space allotted for each chromosome reflects the relative length of each chromosome. Plasma was taken from 24-week-old mice after 18 weeks of dietary treatment. Colored lines show permutation-derived significance thresholds (N=1000) at $P = 0.05$ (LOD=7.57, shown in red), $P = 0.10$ (LOD=7.17,

shown in orange), and P = 0.63 (LOD=5.79, shown in yellow). The eight coefficients of the QTL model show the effect of each founder haplotype on the phenotype. Shading identifies the 95% Bayesian credible interval around the peak (B.).

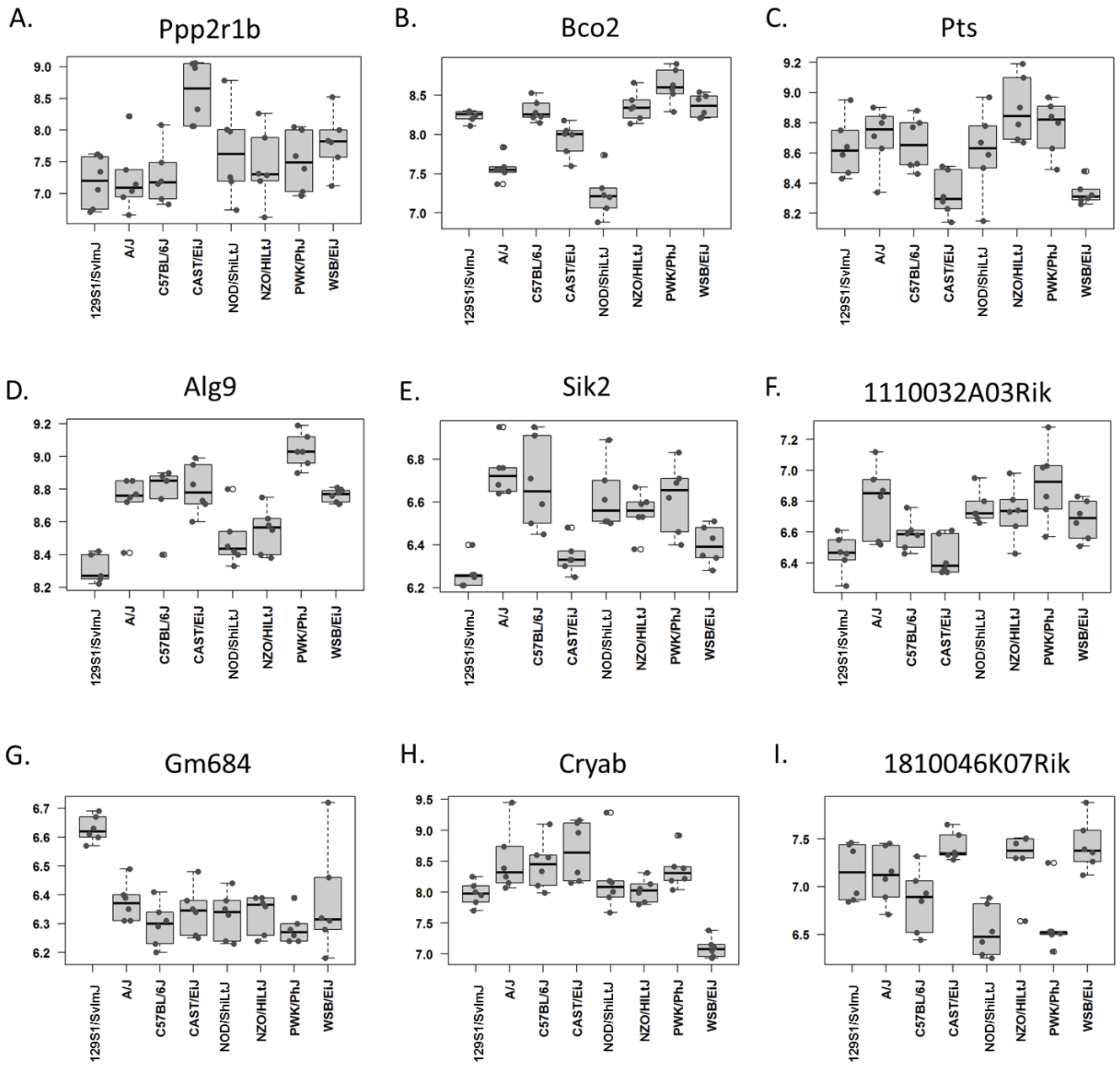


Figure 2.6- Liver expression of candidate genes in the Chromosome 9 peak region associated with baseline triglyceride levels. Gene expression data was obtained from livers of female C57BL6/J, A/J, NOD/ShiLtJ, NZO/HiLtJ, WSB/EiJ, CAST/EiJ, PWK/PhJ and 129S1/SvImJ mice (http://cgd.jax.org/gem/strainsurvey26). Hepatic gene expression of 68 probes representing the 34 genes at the locus were compared. Differential expression of genes across the founder strains was determined using a 1-way ANOVA analysis and a Bonferroni correction for multiple tests. We identified 9 genes as differentially expressed among the founder strains: *Ppp2r1b, Bco2, Pts, Alg9, Sik2, 1110032A03Rik, Gm684, Cryab,* and *1810046K07Rik* (A.–I.). *Bco2* and *Ppp2r1b* are differentially expressed between CAST/EiJ and the other progenitor strains and, therefore, most closely match the allele effects of the QTL.

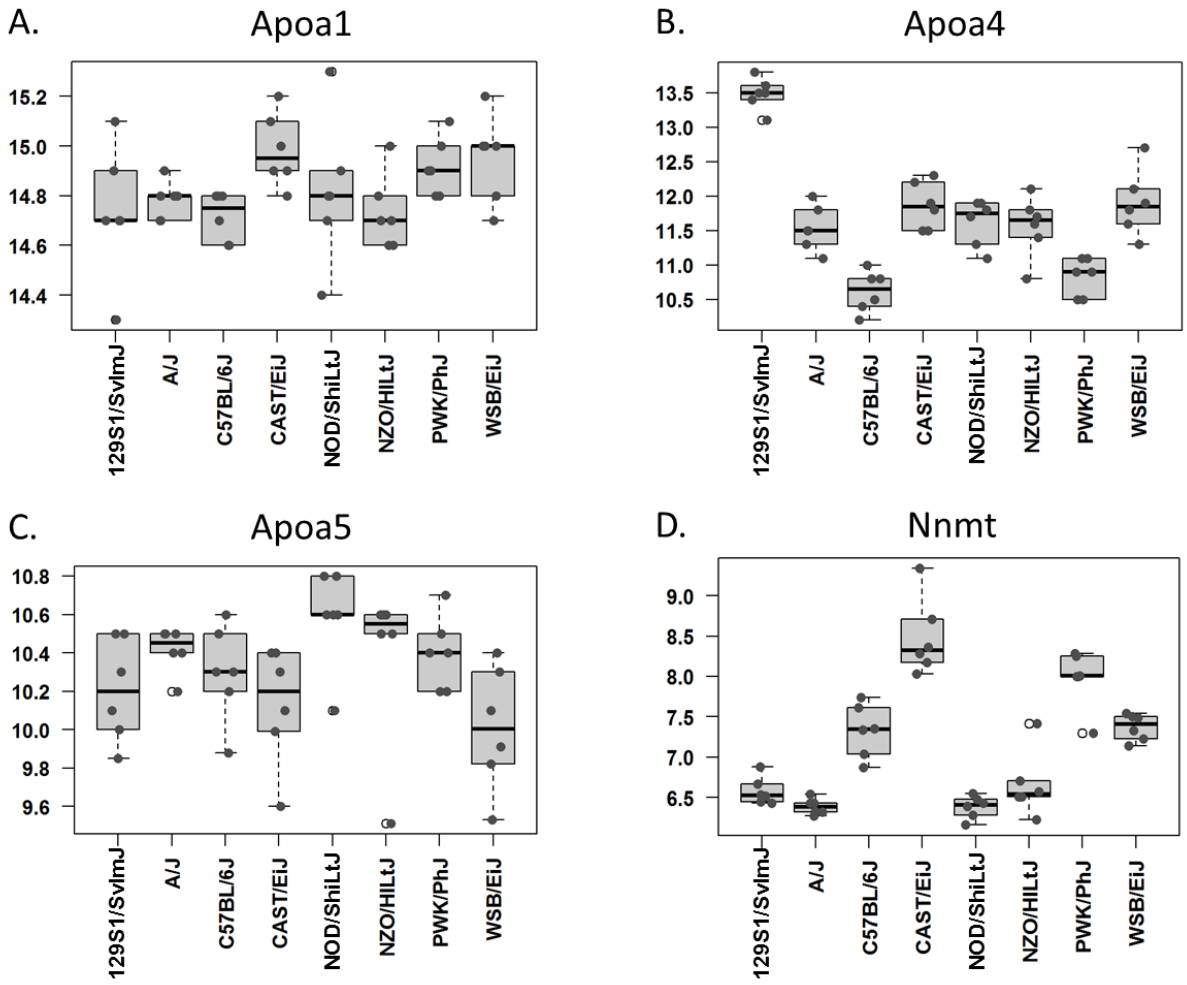
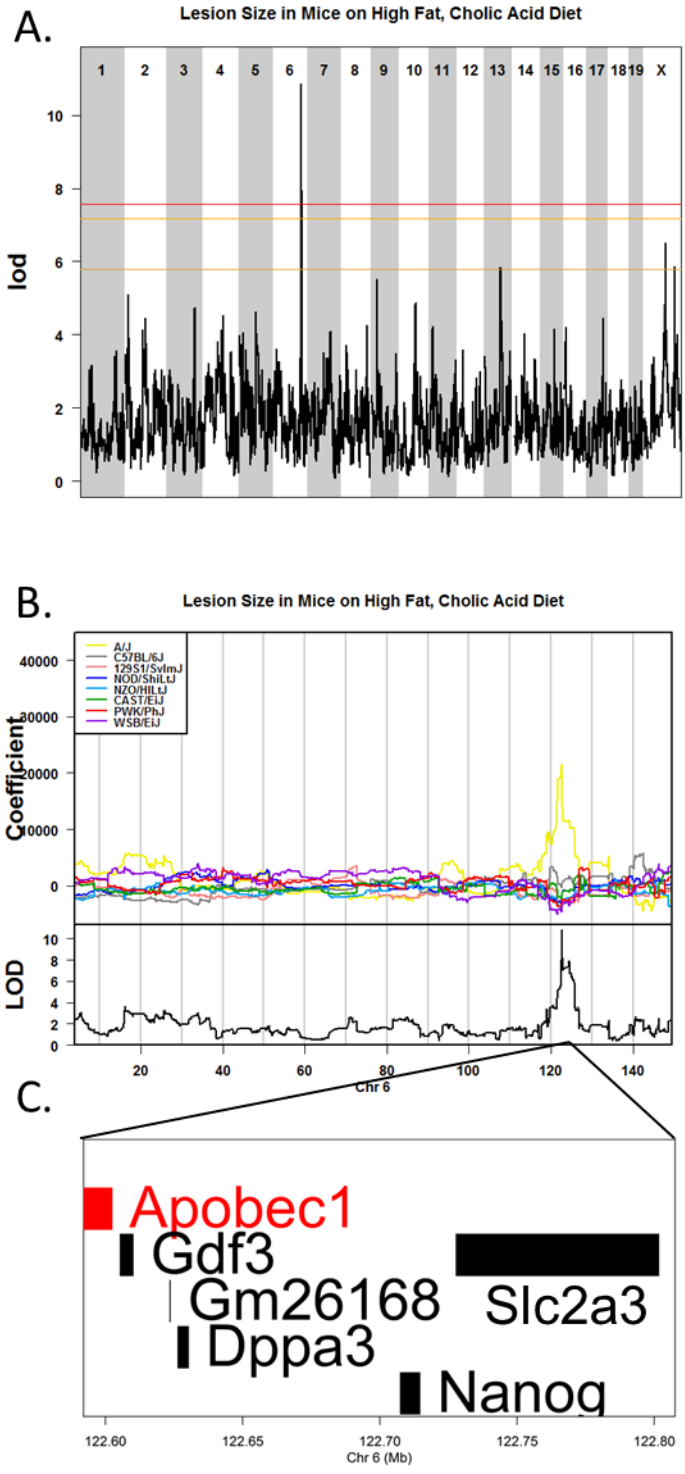


Figure 2.7- Liver expression of genes previously associated with markers of cardiovascular disease on Chromosome 9. Gene expression data was obtained from livers of female C57BL6/J, A/J, NOD/ShiLtJ, NZO/HiLtJ, WSB/EiJ, CAST/EiJ, PWK/PhJ and 129S1/SvImJ mice (http://cgd.jax.org/gem/strainsurvey26). Hepatic gene expression of Chromosome 9 genes previously identified as associated with triglyceride levels were compared. Differential expression of genes across the founder strains was determined using a 1-way ANOVA analysis and a Bonferroni correction for multiple tests. *Apoa1 and Apoa5* were not differentially expressed across the founder strains (A. and C.). *Apoa4* was differentially expressed among the founder strains, but only *Nnmt* was differentially expressed in CAST/EiJ compared to all of the other founder strains (B. and D.).


A.
Lesion Size in Mice on High Fat, Cholic Acid Diet
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 X
lod
0 2 4 6 8 10
B.
Lesion Size in Mice on High Fat, Cholic Acid Diet
A/J
C57BL/6J
129S1/SvImJ
NOD/ShiLtJ
NZO/HlLtJ
CAST/EiJ
PWK/PhJ
WSB/EiJ
Coefficient
40000
20000
0
-20000
-40000
LOD
0 2 4 6 8 10
20 40 60 80 100 120 140
Chr 6
C.
Apobec1
Gdf3
Gm26168
Slc2a3
Dppa3
Nanog
122.60 122.65 122.70 122.75 122.80
Chr 6 (Mb)

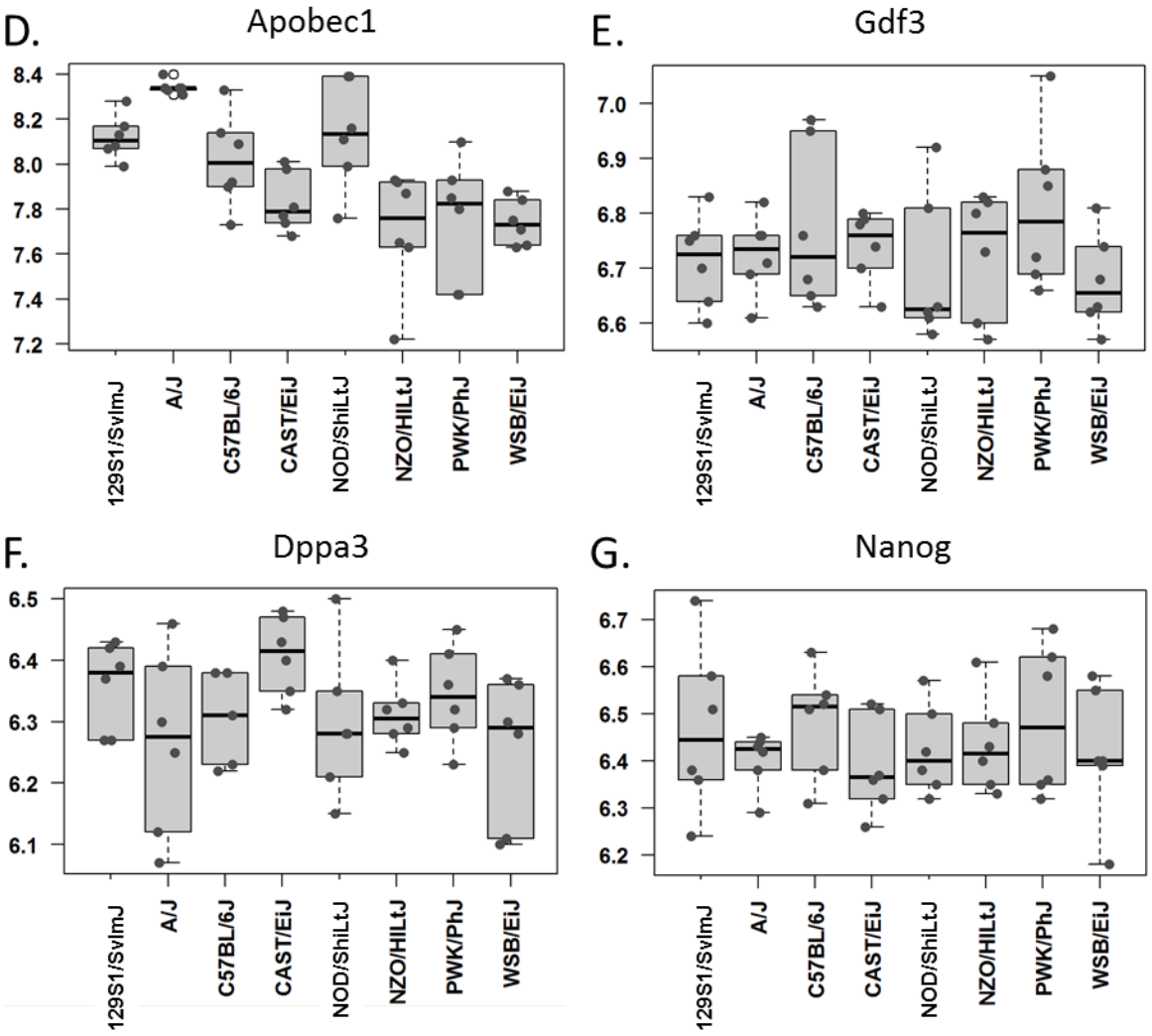


Figure 2.8- QTL mapping of atherosclerosis in the DO mice. Genome-wide QTL scan for loci affecting atherosclerotic lesion size in mice fed a high-fat, cholic acid diet (A.). Chromosomes 1 through X are represented numerically on the *x*-axis and the *y*-axis represents the LOD score. The relative width of the space allotted for each chromosome reflects the relative length of each chromosome. Hearts were harvested from 146 mice after 18 weeks on a high-fat, cholic acid diet. Colored lines show permutation-derived significance thresholds (N=1000) at $P = 0.05$ (LOD=7.57, shown in red), $P = 0.10$ (LOD=7.17, shown in orange), and $P = 0.63$ (LOD=5.79, shown in yellow). The eight coefficients of the QTL model show the effect of each founder haplotype on the phenotype. A/J founder alleles are associated with larger lesion size in the DO mice (B.). There are 6 candidate genes within the 100,000kb QTL interval on Chromosome 6, *Apobec1*, *Gdf3*, *Dppa3*, *Nanog*, *Slc2a3*, and the predicted gene *Gm26168* (C.). Gene expression data was obtained from livers from female C57BL6/J, A/J, NOD/ShiLtJ, NZO/HiLtJ, WSB/EiJ, CAST/EiJ, PWK/PhJ and 129S1/SvImJ mice (http://cgd.jax.org/gem/strainsurvey26). There are six candidate genes in the QTL interval on Chromosome 6, *Apobec1*, *Gdf3*, *Dppa3*, *Nanog*, *Slc2a3*, and the predicted gene *Gm26168*, of which four, *Apobec1* (D.), *Gdf3* (E.), *Dppa3* (F.),

and *Nanog* (G.), were assayed by microarray for hepatic gene expression. *Apobec1* is the only candidate in this region that matches our allele effects such that A/J mice appear to express higher levels of *Apobec1*.

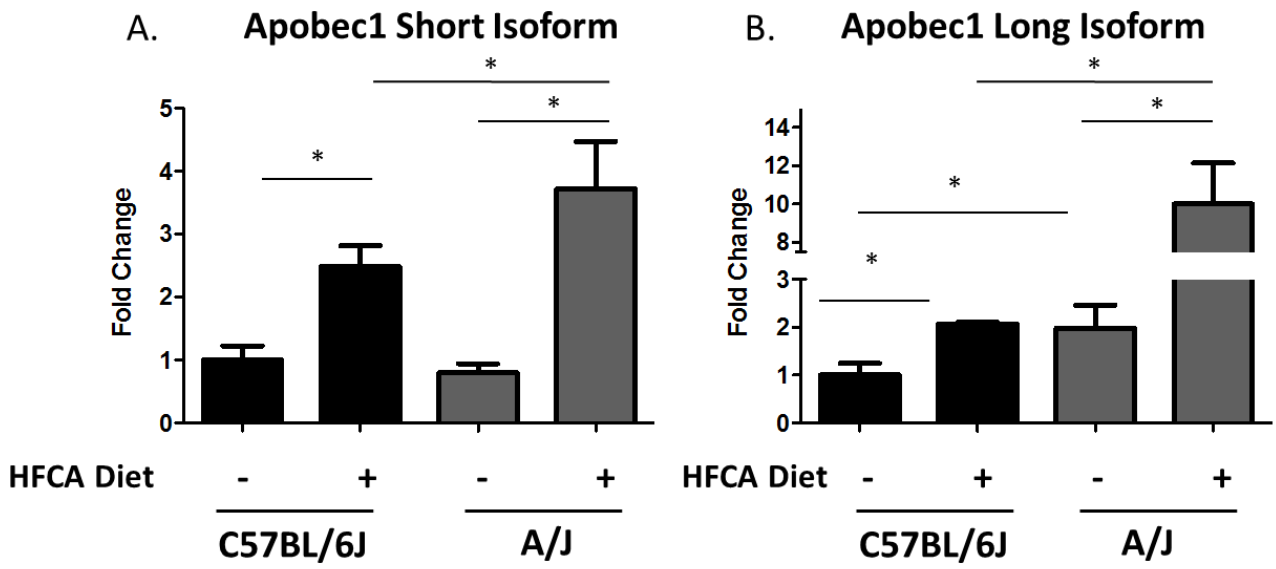


Figure 2.9- A/J preferentially expresses the long isoform of Apobec1 in response to a high-fat, cholic acid diet. *Apobec1* expression levels of the short (A.) and long (B.) isoforms from RNA from liver tissue from A/J and C57/BL6 founder strains. A/J mice on a high-fat, cholic acid diet exhibit increased expression of both the long and short transcripts of *Apobec1* in a diet-dependent manner, * indicates p> 0.05. *Apobec1* levels for each sample were normalized relative to RPS20. Fold changes are reported as the relative expression in A/J versus C57/BL6 samples. Data are presented as mean $\pm$ SD and significance was determined using a Student's t-test.

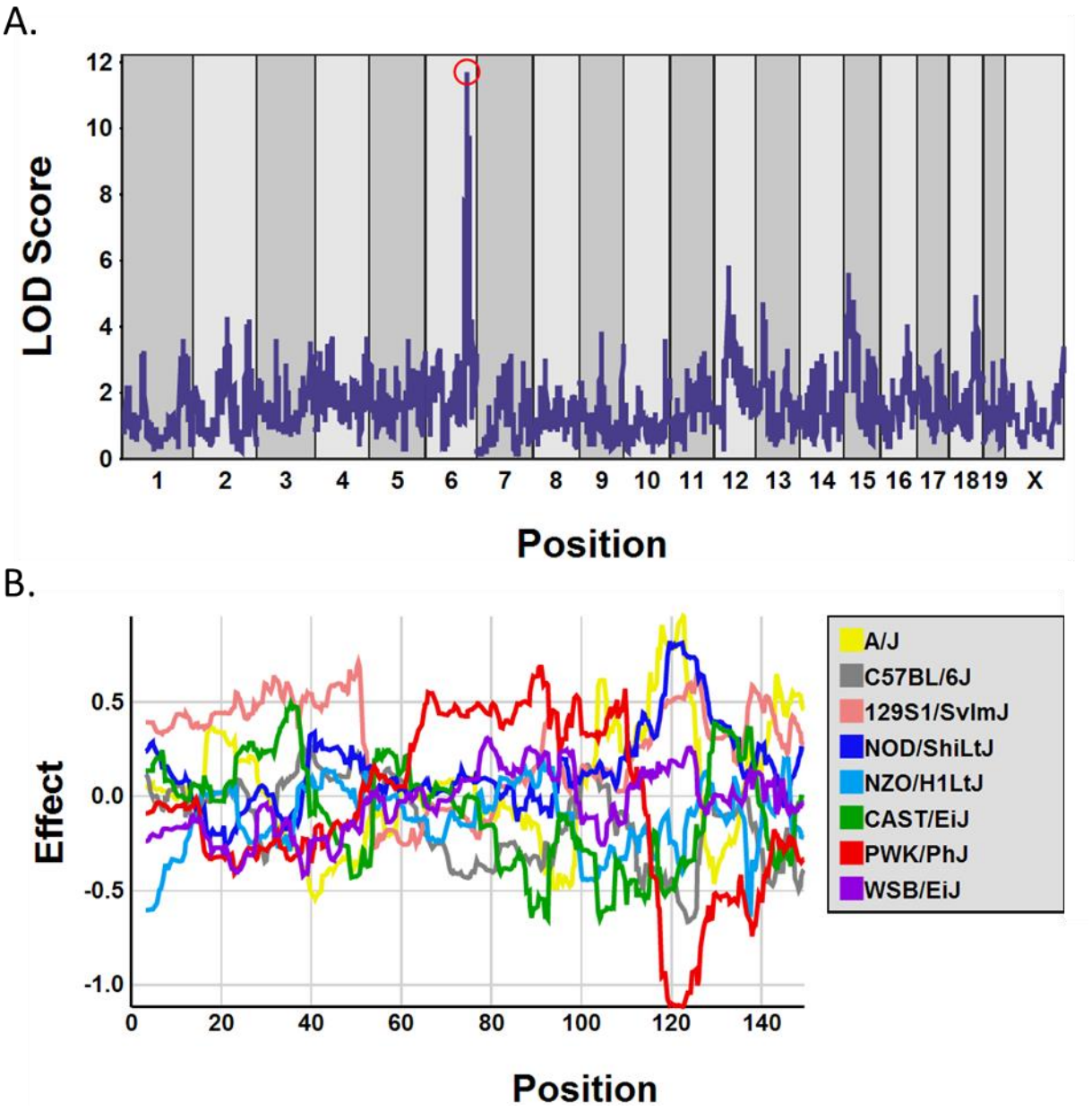


Figure 2.10: Identification of a cis eQTL on Chromosome 6 for *Apobec1* expression. The Jackson Laboratory's Diversity Outbred eQTL viewer data located at http://cgd.jax.org/apps/eqtlviewer-beta/ was queried for eQTL associated with *Apobec1* mRNA expression. A cis eQTL on Chromosome 6 was identified as associated with *Apobec1* expression in the DO mice (LOD= 11.7). The eight coefficients of the QTL model show the effects on the phenotype contributed by each founder haplotype on Chromosome 6 (B.). These data demonstrate that A/J alleles are associated with higher expression of *Apobec1*.

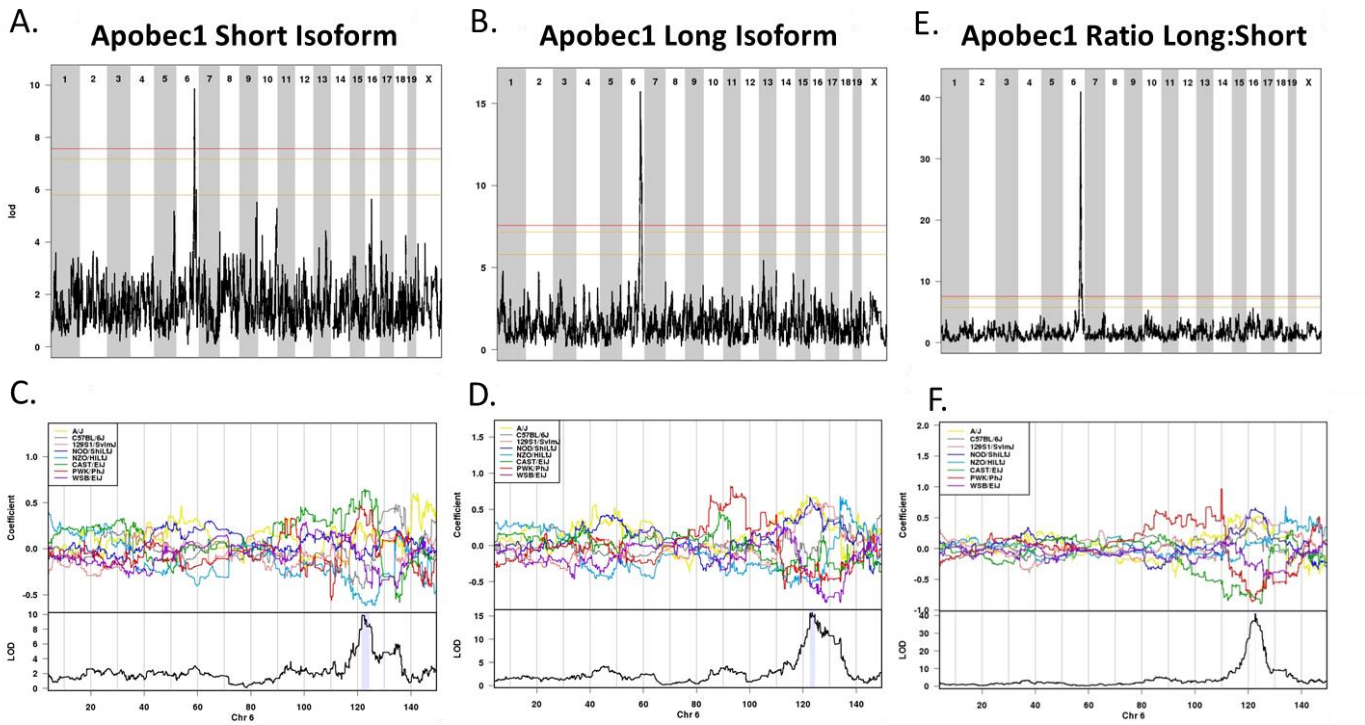


Figure 2.11- *Cis*-eQTL for hepatic *Apobec1* short and long isoforms in the DO mice. Genome-wide QTL scan for eQTL regulating expression of the short (A.) and long (B.) *Apobec1* isoforms in the DO mice. Chromosomes 1 through X are represented numerically on the *x*-axis and the *y*-axis represents the LOD score. The relative width of the space allotted for each chromosome reflects the relative length of each chromosome. Colored lines show permutation-derived significance thresholds at $P = 0.05$ (red), $P = 0.10$ (orange), and $P = 0.63$ (yellow). The eight coefficients of the QTL model show the effects on the phenotype contributed by each founder haplotype on Chromosome 6 for mapping of the short (C.) and long (D.) isoforms of *Apobec1*. Shading identifies the 95% Bayesian credible interval around the peak.

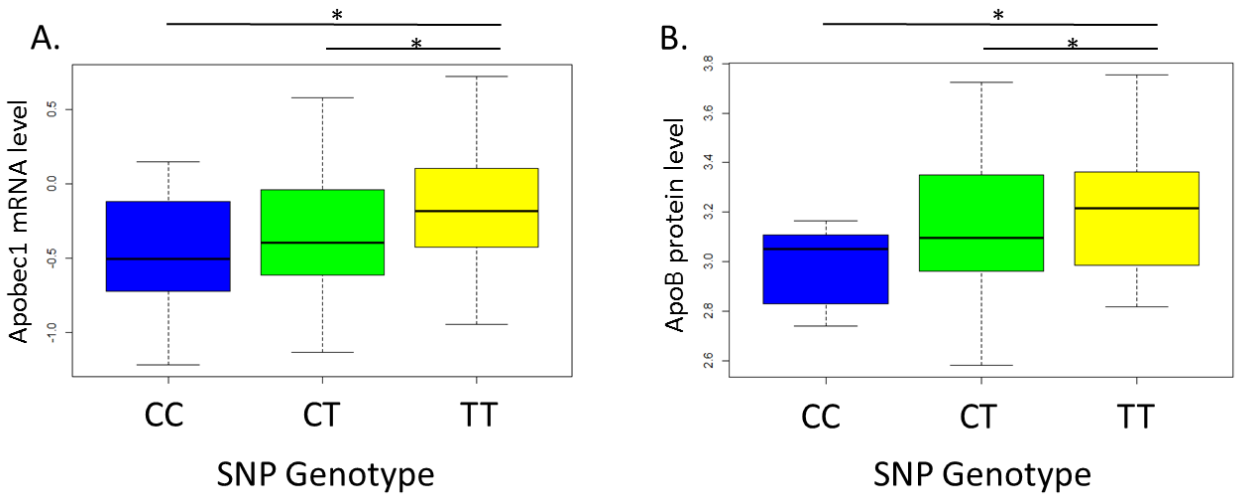


Figure 2.12- ***Apobec1*** **and ApoB levels are dependent on the genotype of UNC11996440, the Chromosome 6 peak SNP associated with atherosclerosis.** Genotyping was performed using the Mouse Universal Genotyping Array (MegaMUGA). Apobec1 mRNA levels were measured by QPCR from liver tissue from the DO mice. Expression levels of the *Apobec1* long isoform differed significantly between genotype classes ($p > 0.001$) (A.). ApoB protein levels were measured in a subset of the mice phenotyped for atherosclerosis (N=80 mice) using a mouse Apolipoprotein B Sandwich-ELISA method in 96-well format. ApoB expression levels differed significantly between the genotype classes (B.).

Product #	**D10001**		**D12083101**		**D12109C**	
%	**gm**	**kcal**	**gm**	**kcal**	**gm**	**kcal**
Protein	20.3	20.8	40.6	40	22.5	20
Carbohydrate	66.0	67.7	40.6	40	45	40
Fat	5.0	11.5	9.1	20	20	40
Total	91.3	100	90.3	100	87.5	100
kcal/gm	3.90		4.07		4.5	
Ingredient	**gm**	**kcal**	**gm**	**kcal**	**gm**	**kcal**
Casein, Lactic	0	0	400	1600	200	800
Casein, 30 Mesh	200	800	0	0	0	0
L-Cystine	0	0	6	24	3	12
DL-Methionine	3	12	0	0	0	0
Corn Starch	150	600	212	848	212	848
Maltodextrin 10	0	0	71	284	71	284
Sucrose	500	2000	113	452	113	452
Cellulose, BW200	50	0	50	0	50	0
Corn Oil	50	450	0	0	0	0
Soybean Oil	0	0	25	225	25	225
Cocoa Butter	0	0	66	594	155	1395
Mineral Mix S10001	35	0	0	0	0	0
Mineral Mix S10021	0	0	10	0	10	0
Dicalcium Phosphate	0	0	13	0	13	0
Calcium Carbonate	0	0	5.5	0	5.5	0
Potassium Citrate	0	0	16.5	0	16.5	0

Vitamin Mix V10001	10	40	10	40	10	40
Choline Bitartrate	2	0	2	0	2	0
Cholesterol	0	0	0	0	11.25	0
Sodium Cholate	0	0	0	0	4.5	0
Red Dye	0	0	0	0	0.05	0
Blue Dye	0	0	0.05	0	0.05	0
Yellow Dye	0	0	0.05	0	0	0
Total	**1000**	**3902**	**1000.1**	**4067**	**901.85**	**4056**

Table 2.1: Compositions of the AIN-76A (D10001), high protein (D12083101), and atherogenic (D12109C) diets used in the study.

The three diets used in this study were manufactured by Research Diets. AIN-76A was fed to the study population from 4-6 weeks of age in order to ensure that there were no spurious effects due to the potential variable composition of standard laboratory chow. Mice were then fed either D12109C or D12083101 for 18 weeks from 6-24 weeks of age. These diets differed by composition, specifically in terms of fat: protein ratio and cholesterol content, D12083101 containing 5% fat and 20.3% protein and D12109C containing 20% fat, 1.25% cholesterol, and 0.5% cholic acid. D12109C is considered atherogenic and was intended to induce the formation of atherosclerosis in the DO mice.

	Baseline (AIN-76A)		High Protein		High-fat, cholic acid	
	n	Mean	n	Mean	n	Mean
Cholesterol (mg/dl)	277	91.7 ± 25.1	128	97.6 ±31.5**	136	199.9 ± 68.6*,**
Triglycerides (mg/dl)	262	59.5 ± 26.5	128	57.7±30.8**	136	32.3 ±12.1*,**
Glucose (mg/dl)	257	155.2 ± 43.8	130	190.5 ±49.9	137	177.9 ± 45.1
Insulin (ng/ml)	235	0.8 ± 0.4	129	1.7±1.1*	133	1.4 ± 0.7*

*Significant difference between baseline diet and post diet group, $p < 0.05$

**Significant difference between post diet groups, $p < 0.05$

Table 2.2: Effects of high protein and high-fat, cholic acid diets on clinical markers of cardiovascular disease in the DO mice.

The values shown are means ± SD. Plasma clinical chemistries were measured at baseline, after two weeks on the synthetic defined diet AIN-76A, and with 4 hours fasting (in 6 week old mice). Mice were transferred to one of two diets, either a high-fat, cholic acid diet or a high protein diet, and maintained on diet for 18 weeks. Plasma clinical chemistries were measured again after diet treatment and with 4 hours fasting (in 24 week old mice). Significant differences between the baseline diet measures and either diet treatment group (*, where $p < 0.05$) or significant differences between the two different diets after 18 weeks of diet treatment (**, where $p < 0.05$) are indicated.

	Phenotype	Chr	Peak LOD	Position (Interval)	Significant (p< 0.05)	Previously Known QTL (Interval)	Known Gene Candidates	Refs.
Baseline	Triglycerides	9	11.3	51.4 (50.2-51.6 Mb)	Y	*Trigq1* (40-70 Mb)	*ApoA5*	Suto, 2003
	Total Cholesterol	13	6.5	30.4 (28.7-43.8 Mb)	N	*Hmgcs1* (20-100 Mb)	*Hmgcr*	Welch, 1996
	Glucose	5	7.0	92.6 (86.3-99.0 Mb)	N	*Bglu13* (84-110 Mb)	*Hnf1a, Pdx1*	Zhang, 2012
	Insulin	19	5.4	58.6 (22.6-60.8 Mb)	N	*Tanidd1* (36-61 Mb)	Not Determined	Kim, 2001
	Triglycerides	12	5.8	97.6 (49.6-102.3 Mb)	N	*Tglq2* (55-89 Mb)	*ApoB*	Srivastava, 2006
	Total Cholesterol	9	7.5	48.3 (47.8-70.0 Mb)	Y	*Cq4* (15-50 Mb), Cq5 (32-66 Mb)	*ApoA4*	Suto, 2003
After diet	Glucose	12	5.7	70.9 (26.7-75.1 Mb)	N	*Bglu15* (10-42 Mb)	*Adam17*	Zhang, 2012
	Insulin	13	7.0	8.7 (5.7-10.4 Mb)	N	NA	NA	NA

Table 2.3: Quantitative trait loci for clinical markers of cardiovascular disease in the DO mice at baseline and after dietary treatment.
Clinical markers of cardiovascular disease were measured in study animals after 4 hours of fasting at both 6 weeks and 24 weeks of age. Baseline measures are from 6-week-old fasted mice fed a synthetic diet. Measures at 24 weeks of age, after diet, include mice on both the high protein diet and the high-fat, cholic acid diet; diet was used as a covariate in the analysis. The statistic reported is log of the odds ratio (LOD).

CHAPTER 3[2]: IDENTIFICATION OF NUMB AS A POTENTIAL GENETIC REGULATOR OF THE ATHEROSCLEROSIS-ASSOCIATED METABOLITE TRIMETHYLAMINE-N-OXIDE IN DIVERSITY OUTBRED MICE

CHAPTER OVERVIEW

Dietary choline and its derivatives have been associated with various aspects of lipid metabolism. Recently, the choline metabolite trimethylamine-N-oxide (TMAO) has been associated with atherosclerosis in both mice and humans. Traditional studies of atherosclerosis in mice use genetic or dietary manipulation to induce atherosclerosis in inbred mouse strains. Inbred mice exhibit strain-specific variation in susceptibility to atherosclerosis and dyslipidemia rendering them useful as models in the dissection of the genetic architecture of these complex diseases. However, traditional quantitative trait locus (QTL) mapping studies using inbred strains often identify large genomic regions, containing many genes, due to limited recombination and/or sample size which hamper candidate gene identification and translation of these results into possible risk factors and therapeutic targets. As an alternative approach, here we use the multi-parental Diversity Outbred (DO) mouse panel for genetic mapping in order to aid in the identification of causal genes and variants associated with TMAO. We mapped QTL for choline metabolites using a linear regression model [101]. Among our highly significant hits, we detected a 4.6 Mb QTL interval on Chromosome 12 with a peak at 86.3 Mb (LOD= 10.0,

[2] This manuscript is in preparation for submission with contribution from and affiliations as follows: Tangi L. Smallwood[1], Kuo-Chen Jung[1], Liyang Zhao[3], Kunjie Hua[1], Daniel Pomp[1,2], Brian J. Bennett[1,2,3] (1. Department of Genetics, University of North Carolina Chapel Hill, North Carolina 27599; 2. Nutrition Research Institute, University of North Carolina Kannapolis, North Carolina 28081; 3. Department of Nutrition, University of North Carolina Chapel Hill, North Carolina 27599

$p<0.05$) for the atherosclerosis-associated metabolite trimethylamine-N-oxide at baseline which was also identified as a suggestive QTL after dietary treatment (LOD= 6.7, $p<0.1$). The Chromosome 12 QTL associated with TMAO contains the positional candidate gene *Numb*, which encodes a clathrin adapter with a role in endocytosis that was recently shown to modulate intestinal cholesterol absorption. We show that TMAO and *Numb* exhibit inverse strain variation across the DO founder strains and that the novel Chromosome 12 TMAO locus co-localizes with a highly significant cis-expression QTL for *Numb*, indicating a potential functional relationship.

INTRODUCTION

Dietary choline and its derivatives have been associated with various aspects of lipid metabolism [132-134]. Several papers published in the last few years have indicated that intestinal microbiota metabolism has an impact on cardiovascular disease risk and the development of atherosclerosis [70, 135-139]. Specifically, Wang et al. published a pivotal paper in 2011 in which they identified choline and two of its metabolites, trimethylamine-N-oxide (TMAO) and betaine, as predictive of cardiovascular disease events in humans. They went on to show that in mice, dietary supplementation with choline or TMAO promoted atherosclerotic lesion development [70]. The results of choline challenge studies in humans in which TMAO levels were found to be elevated after ingestion of hard-boiled eggs corroborated the initial findings that elevated TMAO levels predicted an increased risk of cardiovascular disease events [140]. Additionally, researchers showed that administration of antibiotics suppressed TMAO levels, indicating the vital role of the microbiota as a first line of regulation in monitoring circulating TMAO levels [136]. In addition to regulation by gut microbiota, there is increasing evidence that host genetics plays a role in the regulation of TMAO levels [72, 93, 141].

TMAO is derived from microbiota metabolism of dietary choline and phosphatidylcholine. Choline-containing compounds are metabolized by the gut flora in the

intestines to produce an intermediate compound, TMA, which is then absorbed into the host circulation and neutralized to TMAO primarily by FMO3 in the liver [72, 141]. Aside from the regulation by FMO3, little is known about how the host genetics interact with and regulate circulating TMAO levels.

A recent genome-wide association study using the hybrid mouse diversity panel identified a Chromosome 3 locus associated with TMAO levels and implicated a potential role for the solute carrier family 30 member 7 (SLC30A7) protein, although a functional relationship has yet to be determined. The same study identified two QTL on human chromosomes 1q23.3 and 2p12 with suggestive evidence of association with plasma TMAO levels [93]. *Fmo3* is located on mouse Chromosome 1 and, although it is known to regulate TMAO levels, *Fmo3* has not been identified in genetic mapping studies of plasma TMAO levels, possibly due to the absence of genetic variation around the *Fmo3* locus.

Studies to identify additional genetic regulators of TMAO levels will be important to further elucidate the impact of host genetics on regulation of TMAO. Here, we utilized the newly developed DO mouse population specifically designed for high resolution genetic mapping to identify novel genetic regulators influencing TMAO levels which could present potential therapeutic targets for CVD in humans.

MATERIALS AND METHODS

Animals and Diets

Female Diversity Outbred mice (n = 292; J:DO, JAX stock number 009376) were obtained from the Jackson Laboratory (Bar Harbor, ME) as 146 full sibling pairs at 4 weeks of age and at outbreeding generation 11 (G11) (received September, 2012). The mice were group housed (n = 5 mice per cage) with non-irradiated pine bedding and provided with HEPA-filtered

air and free access to sterile water in a climate-controlled facility under a 12 hour light:dark cycle. Mice were maintained on a defined synthetic diet, AIN-76A, until 6 weeks of age to control for differences due to variable components of standard chow (D10001, Research Diets, New Brunswick, NJ); subsequently, 146 mice were transferred to a synthetic high-fat, cholic acid (HFCA) diet, containing 20% fat, 1.25% cholesterol, and 0.5% cholic acid, to induce atherosclerotic lesions and 146 mice were maintained on a high protein diet containing 5% fat and 20.3% protein which is not atherogenic (D12109C and D12083101, respectively, Research Diets, New Brunswick, NJ). One sibling from each of the 146 sibling pairs was randomly assigned to each one of the diets. The source of fat from the diets varied between the baseline diet (corn oil) fed to the mice from 4 to 6 weeks of age and the dietary treatment groups (soybean oil plus cocoa butter) fed to the mice from 6 to 24 weeks of age. All mice were maintained on their respective diets until 24 weeks of age, for a total of 18 weeks. All procedures were approved by the IACUC at UNC Chapel Hill (IACUC Protocol Number 11-299).

Measurement of Choline Metabolites

Measurement of choline metabolites was performed by the Metabolomics Core Facility (Kannapolis, NC). In brief, plasma was extracted with three volumes of acetonitrile spiked with internal standards TMAO-d9 (DLM-4779-1, Cambridge Isotope Laboratories), creatinine-d3 (D-3689, CDN Isotopes Inc.), choline-d9 (DLM-549-1, Cambridge Isotope Laboratories) and betaine-d9 (616656, Sigma-Aldrich), incubated on ice for 10 minutes, and centrifuged at 15,000 g for 2 minutes. Quantification of TMAO, creatinine, choline and betaine was performed using liquid chromatography-stable isotope dilution-multiple reaction monitoring mass spectrometry (LC-SID-MRM/MS). Chromatographic separations were performed on an Atlantis Silica HILIC 3µm 4.6×150mm column (Waters Corp, Milford, USA) using a Waters ACQUITY UPLC

system. The column was heated to 40°C, and the flow rate was maintained at 1 mL/min. The gradient was 5% A for 0.05 min, to 15% A in 0.35 min, to 20% A in 0.6 min, to 30% A in 1 min, to 45% A in 0.55 min, to 55% A in 0.05 min, at 55% A for 0.9 min, to 5% A in 0.05 min, at 5% A for 1.45 min, where A is 10% acetonitrile/90%water with 10 mM ammonium formate. 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 from its calibration curve using peak area ratio of the metabolite to its isotope. Significant differences between groups were determined using a Student's t-test.

Genotyping

DNA was extracted and purified from tail biopsies taken from 6-week-old mice using Qiagen DNeasy kit according to the manufacturer's instructions. Genotyping was performed using the Mega Mouse Universal Genotyping Array (MegaMUGA) by GeneSeek (Neogen, Lansing, MI) [113]. The call rate exclusion criteria was set at >95% and twelve mice were excluded based on this criteria; the average call rate of the genotyped mice used in the study was 98%. The MegaMUGA array is built on the Illumina Infinium platform and contains 77,808 SNP markers that are distributed throughout the genome at an average spacing of 33 Kb. The MegaMUGA SNPs were subset to include 57,977 informative SNPs that distinguish among the genotypes of the eight founder strains. For the mapping, genomes were reconstructed based on the X and Y allele intensities from the array and founder haplotypes were reconstructed using a hidden Markov model. The founder allele dosages based on the reconstructed haplotypes were then used to perform linkage mapping.

QTL Mapping

QTL mapping was performed using DOQTL Bioconductor release version 1.0.0 for the R environment [101]. DOQTL reconstructs the genomes of the DO mice in terms of 8 state founder allele dosages and performs QTL mapping by regressing the phenotypes on the founder allele dosages of the condensed additive haplotype model with the assumption that the effects of the founder alleles are additive. Phenotypes were natural log-transformed to satisfy the model assumption of a normal distribution. Diet was included as an additive covariate in the mapping model for measurements including phenotypes obtained from 24 week old mice after dietary treatment. Candidate genes were identified by position based on the Wellcome Trust Sanger mouse genomes database, www.sanger.ac.uk, release 1303 based on genome assembly GRCm38 [114]. QTL support intervals were defined by the 95% Bayesian credible interval, calculated by normalizing the area under the QTL curve on a given chromosome [115]. The mapping statistic reported is log of the odds ratio (LOD). The significance thresholds were determined by performing 1000 permutations of genome-wide scans by shuffling phenotypic data in relation to individual genotypes. Significant QTL were determined at a genome-wide p-value of <0.05 and suggestive QTL were determined at a p-value of <0.63. The latter corresponds to one false positive per genome scan [116].

Differential expression of genes in peak regions across the founder strains

Publicly available expression data for candidate genes from liver tissue of female C57BL6/J, A/J, NOD/ShiLtJ, NZO/HiLtJ, WSB/EiJ, CAST/EiJ, PWK/PhJ and 129S1/SvImJ mice (http://cgd.jax.org/gem/strainsurvey26/v1) were obtained from The Jackson Laboratory Gene Expression Strain Survey and analyzed in R to determine differential expression. Mice included in the Jackson Laboratory Gene Expression Survey were maintained for 11 weeks on

standard chow diet (4% fat content). Genes were identified as differentially expressed between the DO founder strains using analysis of variance and significant between strains differences were calculated using Tukey's Post Hoc test. We used a Bonferroni correction to determine statistical significance and correct for multiple comparisons.

Numb mRNA expression in the DO mice

QPCR was performed in triplicate using a High Capacity Reverse Transcriptase Kit (Applied Biosystems, Foster City, CA). Following cDNA conversion, 1 µL of sample cDNA, 2 µL KAPA Sybr Fast qPCR mastermix (KK4610), and 0.5 µM primers was added to each well of a 384-well plate. *Numb* primers were custom designed and ordered from Eurofins MWG Operon (Huntsville, AL). Amplification was performed using the following primer set 5'-caccacagtctcctgtgttacaa-3' (left primer) and 5'-cagtattggctggcttagcagt-3' (right primer). This primer set was designed to recognize the known *Numb* transcripts *Numb-002* and *Numb-003* (ENSMUST00000129335.1 and ENSMUST00000154043.1, respectively). A serial dilution of pooled samples was used to create a standard curve. Sterile water was used as a negative control.

RESULTS

Metabolites are important metabolic intermediates that represent intermediate phenotypes between DNA variation and clinical phenotypes. TMAO is a common metabolite found in mammals that has recently been identified as a clinical predictor of atherosclerosis [134, 140, 142, 143]. TMAO levels are regulated through both the microbiome and through host genetics, though the influence of host genetics on TMAO levels is not yet fully understood [72, 93]. Therefore, we were interested in gaining a better understanding of genetic regulators of choline metabolism by identifying novel genetic regulators of TMAO and choline.

Changes in choline metabolites in response to diet

Our lab has previously shown strain-specific variation in TMAO levels across the eight progenitor strains used to generate the DO mice [144]. Here, we were interested in evaluating the extent of variation of TMAO and its precursor molecule choline in the DO mice. Mice were maintained on a synthetic diet from 4 to 6 weeks of age. From 6 weeks to 24 weeks of age, one mouse from each sib-pair was randomly assigned to either the HFCA diet group or the HP diet group. At baseline, while all mice were still on a controlled, synthetic diet, choline levels differed slightly between the mice assigned to the atherogenic diet versus the high protein diet ($p<0.001$), and this difference was still apparent after diet treatment ($p<0.05$). In response to diet, the mice on the high protein diet exhibited no change in choline levels, while the mice on the atherogenic diet exhibited a significant decrease in choline levels ($p<9 \times 10^{-10}$), **Figure 3.1A**. There was no difference in TMAO levels between the mice from each of the diet groups at baseline. However, after 18 weeks on different diets, we observed a highly significant increase in TMAO levels in response to the atherogenic diet ($p<4 \times 10^{-21}$); while TMAO levels in the mice maintained on the high protein diet did not differ from the baseline levels, **Figure 3.1B**.

Identification QTL on Chromosome 12 associated with choline in 24 week old mice

Choline was measured in the mice at both 6 weeks and 24 weeks of age. While we identified only a suggestive QTL on Chromosome 12 (LOD= 7.3, $p< 0.1$) associated with choline levels in mice at baseline**, Table 3.1**, we identified a significant QTL on distal Chromosome 12 (LOD=7.8, $p< 0.05$) associated with choline in mice after dietary treatment, **Figure 3.2A**. The significant Chromosome 12 QTL has a 4.3 Mb support interval with a peak SNP at 116.8 Mb (113.2-117.5 Mb). By plotting the founder allele coefficients from the additive haplotype model, we can visualize the effect of each founder haplotype on the phenotype. Based

on the founder allele effects, mice with allelic contribution from the C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HiLtJ, and CAST/EiJ founders within the QTL interval have higher choline levels, while mice with allelic contribution from A/J and PWK/PhJ within the QTL interval exhibit lower choline levels, **Figure 3.2B**. This strain distribution pattern suggests that a SNP shared by and unique to A/J and PWK/PhJ may be causal at this locus. There are 68,162 SNPs within the Bayesian credible interval and only 234 have alleles shared by A/J and PWK/PhJ that are different from C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HiLtJ, and CAST/EiJ, consistent with the allele effects we observed suggesting one of these SNPs may be causally associated with regulation of choline. There are 31 positional candidates within this QTL interval, **Figure 3.2C**.

Identification of Chromosome 14 QTL associated with TMAO at baseline

We identified a novel QTL on Chromosome 14 (LOD= 8.2, $p < 0.05$) that was associated with TMAO levels at baseline, **Figure 3.3A**. Based on the founder allele effects we observed, C57BL/6J,129S1/SvImJ, CAST/EiJ, PWK/PhJ alleles in this interval are associated with higher TMAO levels, while A/J, NOD/ShiLtJ, NZO/HiLtJ, and WSB/EiJ alleles in this interval are associated with lower TMAO levels, **Figure 3.3C**. There are 233,673 SNPs within the Bayesian credible interval. 23,108 SNPs are shared by A/J, NOD/ShiLtJ, NZO/HiLtJ, and WSB/EiJ and these are located throughout the interval, from 86,148,966-94,774,798 Mb. Therefore, we are unable to narrow down the genes of interest based on the allele effects within this region. While the Chromosome 14 QTL is 16.3 Mb in size, it is located in a region of relatively few coding genes and there are 49 positional candidates within the interval, **Figure 3.3E**.

Identification of Chromosome 12 QTL associated with TMAO at baseline and after dietary treatment

We identified a novel QTL on Chromosome 12 that was associated with TMAO levels at baseline and after dietary treatment, **Figure 3.3A and 3.4A**, respectively. In 6 week old mice, the Chromosome 12 QTL is significantly associated with TMAO levels (LOD=10.0, $p< 0.05$), **Figure 3.3A**. This QTL has a 4.6 Mb (83.6-88.2 Mb) support interval and contains 116 genes, **Figure 3.3C**. Based on the founder allele effects, we see that allelic contribution from A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HiLtJ, and WSB/EiJ within the interval is associated with the high TMAO levels, while allelic contribution from CAST/EiJ and PWK/PhJ alleles within the interval are associated with low TMAO levels, **Figure 3.3B**. While the allele effects within the QTL interval can often be used to narrow down the list of candidate genes to those genes that have expression matching the allele effects or functional variants exhibiting the same strain distribution pattern, CAST/EiJ and PWK/PhJ are wild-derived strains and they contribute many private alleles that span the entire QTL interval. Indeed, of the 61,363 genotyped SNPs within the 4.6 Mb interval, 15,386 match our allele effects and are shared by and unique to CAST/EiJ and PWK/PhJ. Further characterization of SNPs within the interval indicates that there are 259 3' UTR SNPs, 55 5' UTR SNPs, 78 non-synonymous SNPs, 26 splice variants, 16 insertions, and 35 deletions spanning the QTL interval (between 83,569,930-88,217,420 Mb) that are shared by CAST/EiJ and PWK/PhJ (www.sanger.ac.uk), **Table 3.2**.

We identified the same QTL on Chromosome 12 (LOD= 6.7, $p <0.05$) as associated with TMAO after dietary treatment, though it did not reach genome-wide significance after diet **Figure 3.4A**. The allele effects for the QTL also indicate that allelic contribution from CAST/EiJ and PWK/EiJ within the interval is associated with low circulating TMAO levels, while allelic

contribution from A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HiLtJ, and WSB/EiJ within the interval is associated with the high TMAO levels **Figure 3.4B**. The QTL for TMAO after dietary treatment has a 3.8 Mb (83.1 Mb- 87.1 Mb) support interval with a peak SNP at 85.7 Mb, less than 1 Mb from the peak SNP associated with TMAO in mice at 6 weeks of age, and the interval contains 85 positional candidates, **Figure 3.4C**. These results suggest that a SNP shared by CAST/EiJ and PWK/EiJ is likely causal for the association and that the gene underlying the association is influenced by diet.

Differential expression of TMAO positional candidates

Because the Chromosome 12 QTL associated with TMAO was highly significant and appeared to be affected by diet, we decided to prioritize this QTL for candidate gene testing. In addition to a causal variant resulting in a structural change in a gene associated with TMAO levels, this QTL could be due to a variant influencing expression of the causal gene. We hypothesized that if the QTL we have identified results from altered gene expression of the causal gene, then we may be able to narrow our list of candidate genes by cross referencing the allele effects in our DO study population with expression levels among the founder strains. We compared the hepatic gene expression among the founder strains for the genes in the region for which expression data was available (http://cgd.jax.org/gem/strainsurvey26).

Based on the hepatic expression levels, we identified candidate genes within the QTL that were differentially expressed across the founder strains and then looked for the differentially expressed genes that most closely matched these allele effects patterns observed in our DO study population. Of the genes in this locus, 60 were quantitated by microarray and 37 of these were differentially expressed between all strains. The 8 most differentially expressed genes are: *Gstz1, Entpd5, Ahsa1, Rps6kl1, Ttll5, Esrrb, Numb,* and *Dlst*, **Figure 3.5A-H**. *Gstz1* and *Entpd5*

exhibited lower expression in CAST/EiJ mice, but not in PWK/PhJ, compared to the other strains (3- and 4-fold lower levels in the CAST/EiJ strain compared to all other strains, respectively). Likewise, *Dlst* was more highly expressed in CAST/EiJ, but not PWK/PhJ compared to the other strains (1.5-fold higher in the CAST/EiJ strain compared to all other strains). Of the 8 differentially expressed genes, *Numb* is the only candidate gene that matches the allele effects patterns we see based on the founder coefficients and is expressed at 1.4-fold lower levels in CAST/EiJ and 1.5-fold lower in PWK/PhJ when compared to the average *Numb* expression exhibited by the other 6 founder strains.

Identification of a highly significant *cis*-eQTL for Numb on Chromosome 12

We next measured *Numb* mRNA expression specifically by QPCR. We hypothesized that if Numb is causally associated with TMAO and being regulated at the level of expression, *Numb* expression would also map to Chromosome 12. We measured mRNA expression of the major isoforms of *Numb* from liver tissue from the DO mice and performed QTL mapping using the additive haplotype model. We identified a highly significant *cis*-eQTL on Chromosome 12 with the peak SNP UNC21514106 located at 83.6 Mb and with a maximum LOD score of 36.2, **Figure 3.6A**. When we estimate the effect of each founder at each marker along Chromosome 12, we see that both CAST/EiJ and PWK/PhJ alleles are associated with higher expression of *Numb*, **Figure 3.6B**. Taken together, these results indicate that mice contributing CAST/EiJ and PWK/PhJ alleles within the QTL interval on Chromosome 12 have higher expression of *Numb* and exhibit lower circulating TMAO levels. Consistent with this finding, *Numb* expression and TMAO levels are highly negatively correlated in the DO mice at both 6 weeks (r= -0.65) and 24 weeks (r= -0.40), **Table 3.3**.

We identified 15,386 CAST/EiJ and PWK/PhJ shared SNPs within the QTL on Chromosome 12 associated with TMAO at baseline, including many non-synonymous coding SNPs, splice variants, insertions, deletions, as well as SNPs in noncoding regions which may influence transcription factor binding and regulation of expression, **Table 3.2**. Of the 259 CAST/EiJ and PWK/PhJ shared 3' UTR SNPs, 14 are located in the 3' UTR of *Numb*. Of the 78 CAST/EiJ and PWK/PhJ shared non-synonymous coding SNPs, two are located in the coding region of the *Numb* gene, including a T→ G at 83,797,280 (no rs#) and an A → G at 83,801,288 resulting in a valine to alanine substitution. Of the 16 CAST/EiJ and PWK/PhJ shared insertions and the 35 CAST/EiJ and PWK/PhJ shared deletions; there are two insertions and one deletion in the *Numb* gene. Within 10 kb of *Numb*, there are many additional SNPs shared by CAST/EiJ and PWK/PhJ and it is currently unclear which SNPs may be causally associated with TMAO, **Table 3.4**.

In conclusion, here we have identified here two novel QTL associated with TMAO in the DO mice. The Chromosome 12 QTL colocalizes with a highly significant *cis*-eQTL for the gene *Numb*. *Numb* encodes a clathrin adaptor with a reported role in NPC1L1-dependent endocytosis [145]. Consistent with the known function of *Numb* in mediating intestinal cholesterol absorption, we observe that *Numb* expression is highly correlated with cholesterol in our cohort of DO mice (r= 0.81), **Table 3.3**. We report here that *Numb* is highly negatively correlated with TMAO in these mice, consistent with the founder allele effects we observed. These results and the colocalization of QTL for plasma TMAO and *Numb* mRNA expression, suggest that *Numb* may play a role in the regulation of TMAO by host genetics. Additional studies investigating the association between *Numb* and TMAO and the functional relevance of this association should be pursued.

DISCUSSION

Metabolites are produced in the cell as a result of various enzymatic reactions and in part reflect the metabolic state of the cell and in some instances are used as biomarkers for human disease [146, 147]. We were particularly interested in choline and its metabolites because, in collaborative studies, we identified choline and its metabolite TMAO as predictive of cardiovascular disease events in humans [72, 134]. TMAO has been shown to promote atherosclerotic lesion development in mice and this metabolite exhibits complex regulation by both the gut microbiota and host genetics [72, 93, 134]. To date, there has been one QTL identified on mouse chromosome 1 as significantly associated with plasma TMAO and two suggestive QTL identified in humans [93]. Therefore, we utilized the newly developed DO mouse resource to identify novel genetic regulators of choline and TMAO.

Chromosome 12 QTL associated with choline in 24 week old mice

We identified a significant QTL on Chromosome 12 (LOD=7.8, $p < 0.05$) associated with choline in 24 week old DO mice. Based on the founder allele effects, DO mice with allelic contribution from the C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HiLtJ, and CAST/EiJ founders within the QTL interval have higher choline levels, while mice with allelic contribution from A/J and PWK/PhJ within the QTL interval exhibit lower choline levels. Therefore, we predict that the causal variant will have an allele shared by A/J and PWK/PhJ. We identified 234 of the 68,162 SNPs within the QTL interval as exhibiting this strain distribution pattern. Most of the 234 SNPs with alleles shared by A/J and PWK/PhJ are intergenic or intronic and they are positioned throughout the interval, from 113,202,771-117,469,183 Mb. Any of these could influence the expression or splicing of nearby genes. There are 10 non-synonymous coding SNPs with alleles shared by A/J and PWK/PhJ in the QTL interval and these are in the following

genes: *AC079181.3, AC160473.1, Gm16747, and Ighv6-6*. These genes have all been identified as encoding immunoglobulin heavy chain variable regions that contribute to mounting an immune response against pathogens. *Ighv* family genes are known to undergo mutation readily to adapt to antigens presented by foreign bacteria and viruses [148]. Indeed, within the QTL interval on Chromosome 12 there are a total of 1330 non-synonymous SNPs and only 56 are in non-*Ighv* family genes and none of those are shared by A/J and PWK/PhJ, suggesting that the variant underlying the QTL is likely one of the 224 intergenic or intronic SNPs affecting expression or isoform usage of one of the positional candidates.

Chromosome 14 QTL associated with baseline TMAO

We identified a QTL on Chromosome 14 (LOD= 8.2, $p < 0.05$) that spans a 16.3 Mb and contains only 49 positional candidate genes. Based on the founder allele effects we observed, C57BL/6J, 129S1/SvImJ, CAST/EiJ, PWK/PhJ alleles in this interval are associated with higher TMAO levels, while A/J, NOD/ShiLtJ, NZO/HiLtJ, and WSB/EiJ alleles in this interval are associated with lower TMAO levels. Based on analysis of the 233,674 SNPs in this region, we identified 23,108 genotyped SNPs shared between A/J, NOD/ShiLtJ, NZO/HiLtJ, and WSB/EiJ. This includes 7 non-synonymous variants matching these allele effects in the genes *Diap3* and *Tdrd3* and one non-synonymous variant specific to A/J, NOD/ShiLtJ, and WSB/EiJ in the gene *Pcdh9*.

Diap3 encodes the Diaphanous Homolog 3 which is involved in actin remodeling. *Tdrd3* encodes the Tudor Domain Containing 3 protein which has been shown to be a scaffolding protein and also acts as a transcriptional activator in the nucleus. *Pcdh9* is a protocadherin involved in cell-cell adhesion. None of these genes has been previously associated with TMAO or atherosclerosis.

Within the interval, there is one gene that has been associated with atherosclerosis, *Klf5*. *Klf5* encodes Krüppel-like factor 5 which is a zinc-finger transcription factor that plays a role in cell proliferation in the intestinal epithelium and mediates tissue remodeling in atherosclerosis [149-151]. In humans, SNPs in *Klf5* have been associated with risk for hypertension [149]. This evidence suggests that *Klf5* may be an attractive candidate for modulating the relationship between TMAO and atherosclerosis. There are 380 documented SNPs in *Klf5* (www.sanger.ac.uk). However, none of these exhibits the same strain distribution pattern we identified based on the allele effects from the additive haplotype model. Therefore, while biologically *Klf5* represents the most interesting candidate underlying the Chromosome 14 QTL, we do not identify SNPs matching our allele effects in this gene and we are unable to narrow down potential candidates among the 49 positional candidate genes based on our allele effects at this QTL.

Chromosome 12 QTL associated with TMAO at baseline and after dietary treatment

We identified a QTL on Chromosome 12 significantly associated with TMAO at baseline and suggestively associated with TMAO after dietary treatment. We observed that DO mice with CAST/EiJ and PWK/PhJ alleles within the Chromosome 12 QTL had lower TMAO levels than mice with A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HiLtJ, and WSB/EiJ alleles at the locus. We analyzed this region for CAST/EiJ and PWK/PhJ private SNPs and missense variants, CAST/EiJ and PWK/PhJ are two of the three wild-derived strains used to create the DO and carry many private SNPs compared to the other inbred strains. Therefore, we identified numerous SNPs shared by these strains as well as non-synonymous mutations in 38 of the 116 genes. It is possible that a SNP shared by CAST/EiJ and PWK/PhJ results in a functional variant

in one of the genes within this locus and the 38 genes with missense variants shared by CAST/EiJ and PWK/PhJ cannot be ruled out as potential candidates.

We also looked at expression of the genes in this region. We compared the hepatic gene expression among the founder strains for the genes in the region using the publicly available Jackson Strain Expression Survey (http://cgd.jax.org/gem/strainsurvey26). We identified the 8 most differentially expressed genes as *Gstz1, Entpd5, Ahsa1, Rps6kl1, Ttll5, Esrrb, Numb,* and *Dlst*. Of the 8 differentially expressed genes, *Numb* is the only candidate gene that matches the allele effects patterns we observed in the DO mice. We measured *Numb* mRNA expression from liver tissue from the DO mice and mapped expression of this gene to a highly significant *cis*-eQTL on Chromosome 12. Interestingly, these allele effects mirror the allele effects observed for TMAO, suggesting that DO mice with CAST/EiJ and PWK/PhJ alleles at the Chromosome 12 QTL specifically exhibit high levels of *Numb* and low levels of TMAO. Consistent with this finding, *Numb* expression is highly negatively correlated with plasma TMAO both at baseline (r= -0.65) and after dietary treatment (r= -0.40), **Table 3.3**.

While we cannot exclude the possibility that other genes in the region are causal, we have identified *Numb* as a potential candidate gene that may play a role in regulating TMAO levels. Classically, *Numb* was identified in Drosophila for its role in cell fate determination and it interacts with and directly binds Notch [152]. More recently, a pivotal role for Numb in the intestinal absorption of cholesterol was identified via its interaction with Niemann-Pick c1-like 1 protein (NPC1L1) [153]. Additionally, recent studies have reported that dietary supplementation with TMAO causes decreased mRNA expression of *Npc1L1* in the mouse intestine, indicating a potential negative feedback loop in Npc1L1-mediated absorption, which is mediated exclusively

by the adaptor protein Numb [142]. Follow up studies should be conducted to determine whether there is a causal association between *Numb* and TMAO.

In conclusion, we identify here for the first time two novel QTL associated with TMAO in mice. The significant QTL on Chromosome 12 was present both before and after dietary treatment. Based on the allele effects we observed and differential expression across the founder strains, we identified *Numb* as a high priority candidate gene underlying the association. We measured expression of *Numb* in the DO mice and identified a highly significant *cis*-eQTL regulating expression of this gene. The allele effects of the *cis*-eQTL mirrored the allele effects we observed for plasma TMAO, suggesting that CAST/EiJ and PWK/PhJ alleles within the Chromosome 12 QTL interval contribute to decreased TMAO levels and increased *Numb* expression. Consistent with this, TMAO and *Numb* are highly negatively correlated in the DO mice. Based on this evidence, we hypothesize that the candidate gene *Numb* may be regulating TMAO in these mice and may represent a novel pathway by which the atherosclerosis-associated metabolite is being regulated by host genetics. While *Numb* has not yet been functionally linked to atherosclerosis, we believe that this may represent a new pathway to the development and progression of atherosclerotic lesions.

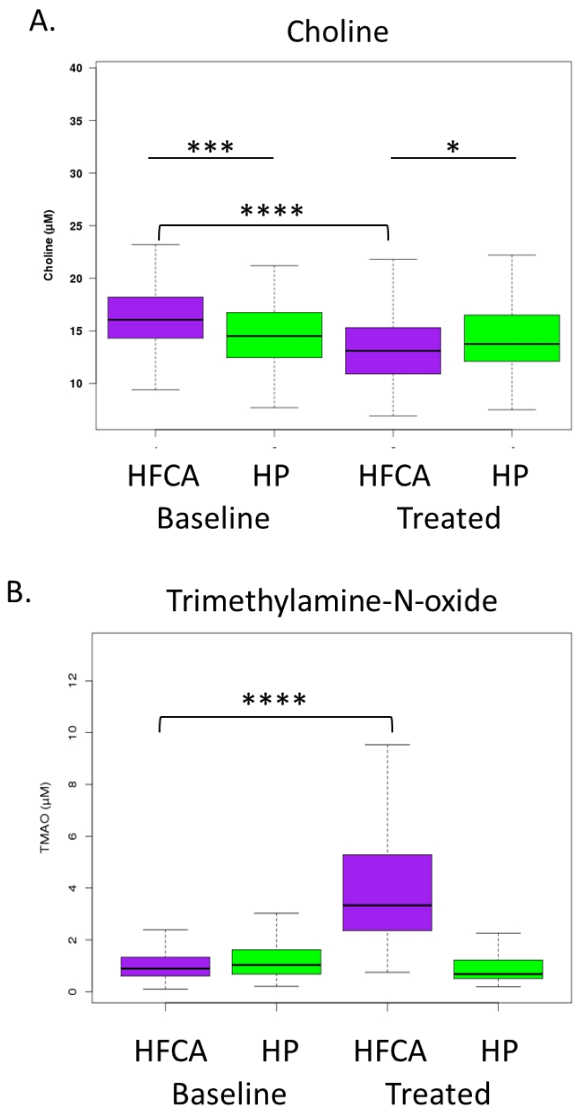


Figure 3.1: Effects of high fat, cholic acid and high protein diets on choline and TMAO levels in the DO mice. Mice were maintained on a synthetic diet for two weeks, fasted for four hours, and then phenotyped for plasma clinical chemistries at 6 weeks of age (Baseline). Following two weeks of synthetic diet, mice were transferred to either a high protein diet (HP) or an atherogenic diet (HFCA). Plasma was taken from 24-week-old mice after 18 weeks on their respective diets, and with four hours fasting, and then phenotyped for plasma clinical chemistries after diet treatment (Treated) (*p< 0.05, **p< 0.01, ***p< 0.001, ****p< 0.0001).

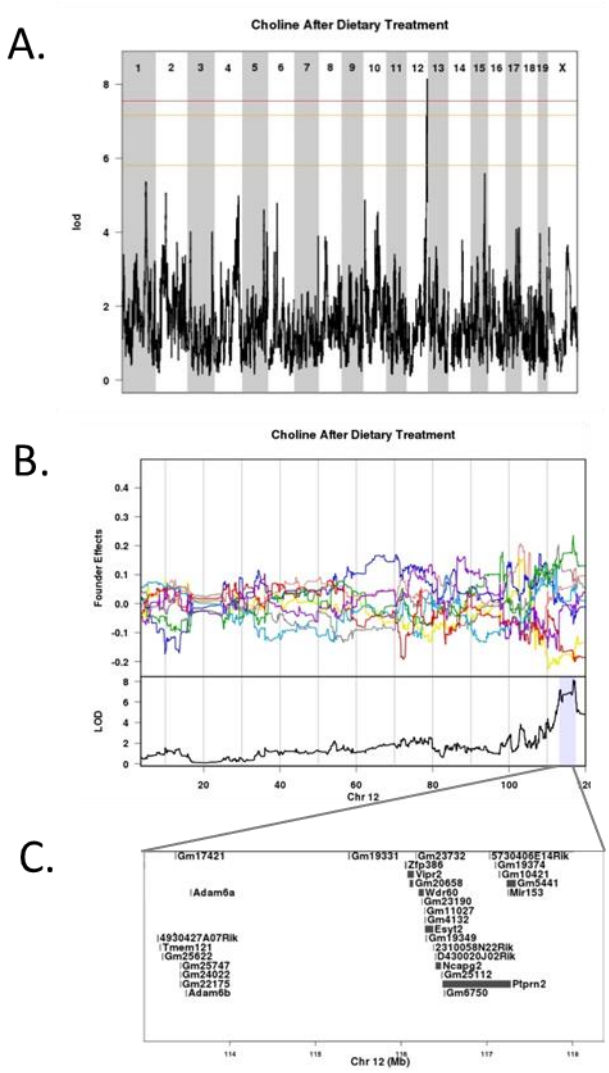


Figure 3.2: QTL mapping of choline in the DO mice after dietary treatment. Genome-wide QTL scan for loci affecting choline levels after 18 weeks of dietary treatment (A). Chromosomes 1 through X are represented numerically on the *x*-axis and the *y*-axis represents the LOD score. The relative width of the space allotted for each chromosome reflects the relative length of each chromosome. Plasma was taken from 24-week-old mice after 18 weeks of dietary treatment. Colored lines show permutation-derived significance thresholds (N=1000) at $P = 0.05$ (LOD=7.57, shown in red), $P = 0.10$ (LOD=7.17, shown in orange), and $P = 0.63$ (LOD=5.79, shown in yellow). The eight coefficients of the QTL model show the effect of each founder haplotype on the phenotype. Shading identifies the 95% Bayesian credible interval around the peak (B.). There are 31 potential candidate genes within the Chromosome 12 locus associated with choline after dietary treatment (C.).

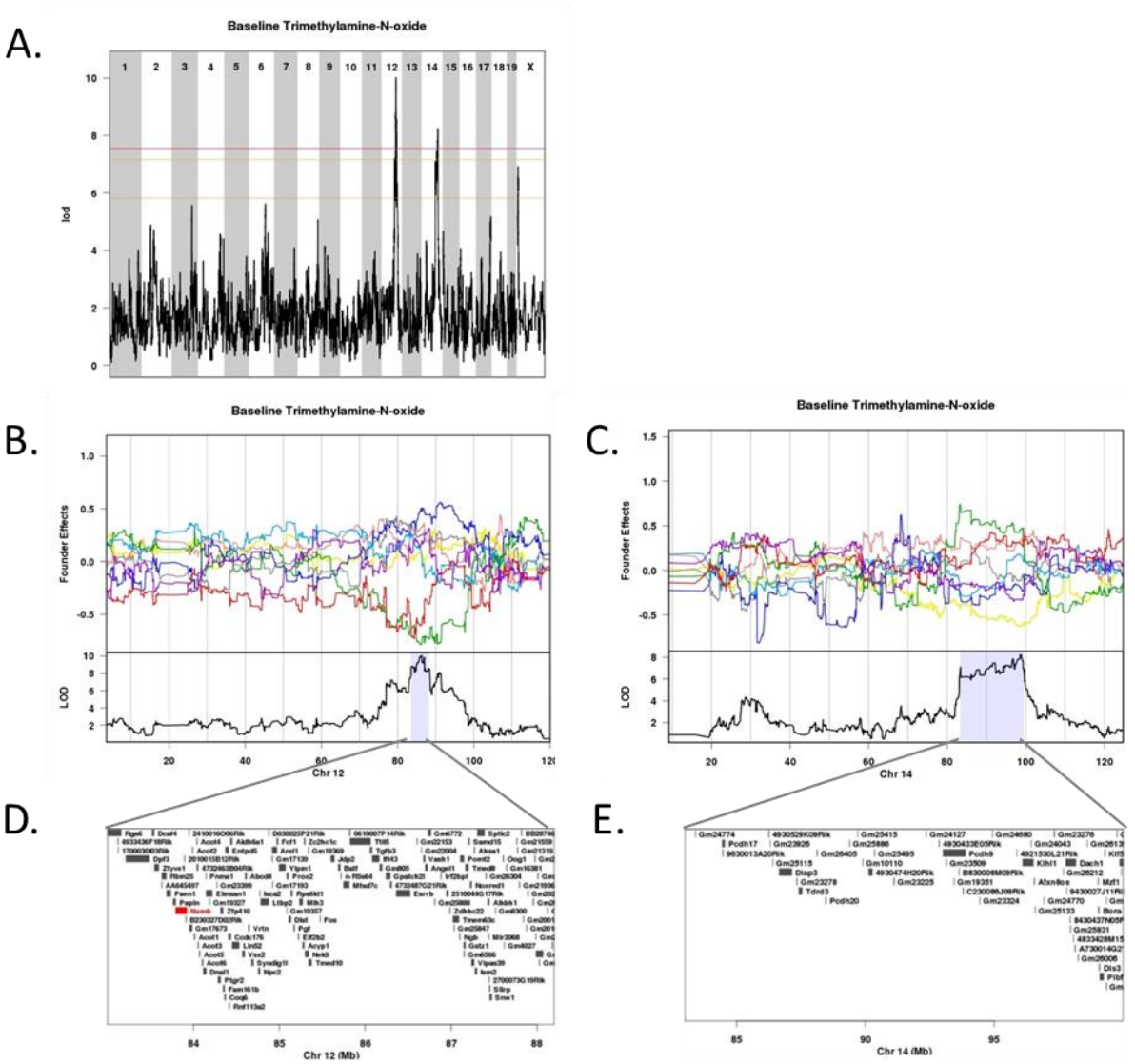


Figure 3.3: QTL mapping of TMAO in the DO mice at baseline. Genome-wide QTL scan for loci affecting choline in 6 week old mice. Chromosomes 1 through X are represented numerically on the *x*-axis and the *y*-axis represents the LOD score. The relative width of the space allotted for each Chromosome reflects the relative length of each Chromosome. Mice were maintained on a synthetic diet for 2 weeks and then phenotyped for plasma clinical chemistries at 6 weeks of age. Colored lines show permutation-derived significance thresholds (N=1000) at P = 0.05 (LOD=7.5, shown in red), P = 0.10 (LOD=7.1, shown in orange), and P = 0.63 (LOD=5.8, shown in yellow) (A.). The eight coefficients of the QTL model show the effect of each founder allele on the phenotype along Chromosome 12 (B.) and Chromosome 14 (C.). Shading identifies the 95% Bayesian credible interval around the peak. There are 116 potential candidate genes within the Chromosome 12 locus. The candidate gene *Numb* is highlighted in red (D.). There are 49 potential candidate genes within the Chromosome 14 locus (E.).

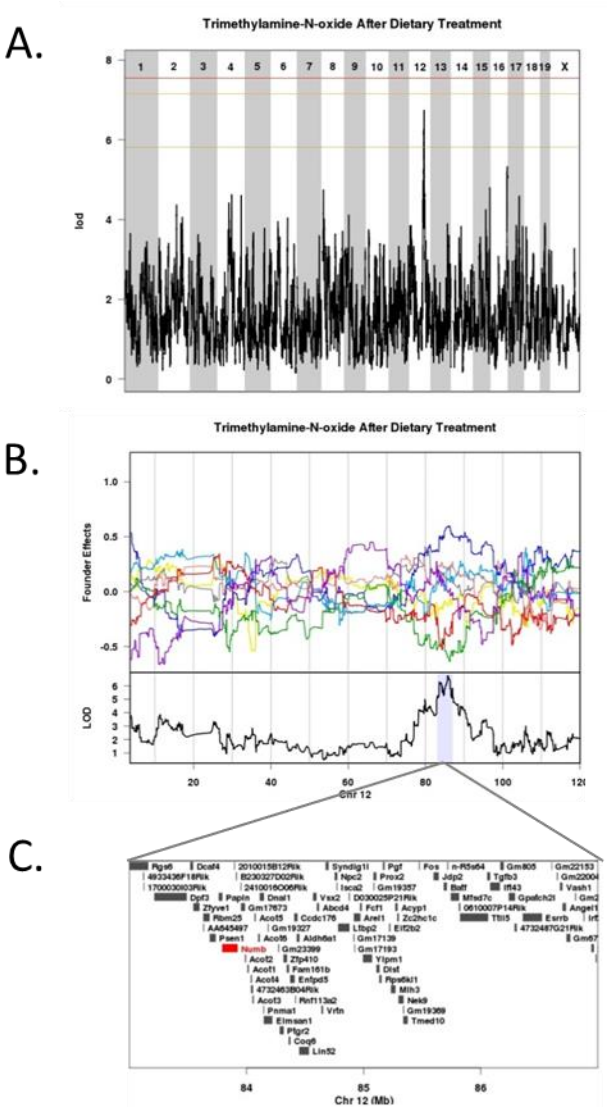


Figure 3.4: QTL mapping of TMAO in the DO after dietary treatment mice. Genome-wide QTL scan for loci affecting TMAO levels after 18 weeks of dietary treatment (A). Chromosomes 1 through X are represented numerically on the *x*-axis and the *y*-axis represents the LOD score. The relative width of the space allotted for each chromosome reflects the relative length of each chromosome. Plasma was taken from 24-week-old mice after 18 weeks of dietary treatment. Colored lines show permutation-derived significance thresholds (N=1000) at $P = 0.05$ (LOD=7.57, shown in red), $P = 0.10$ (LOD=7.17, shown in orange), and $P = 0.63$ (LOD=5.79, shown in yellow). The eight coefficients of the QTL model show the effect of each founder haplotype on the phenotype. Shading identifies the 95% Bayesian credible interval around the peak (B.). There are 85 potential candidate genes within the Chromosome 12 locus associated with TMAO after dietary treatment. The candidate gene *Numb* is highlighted in red (C.).

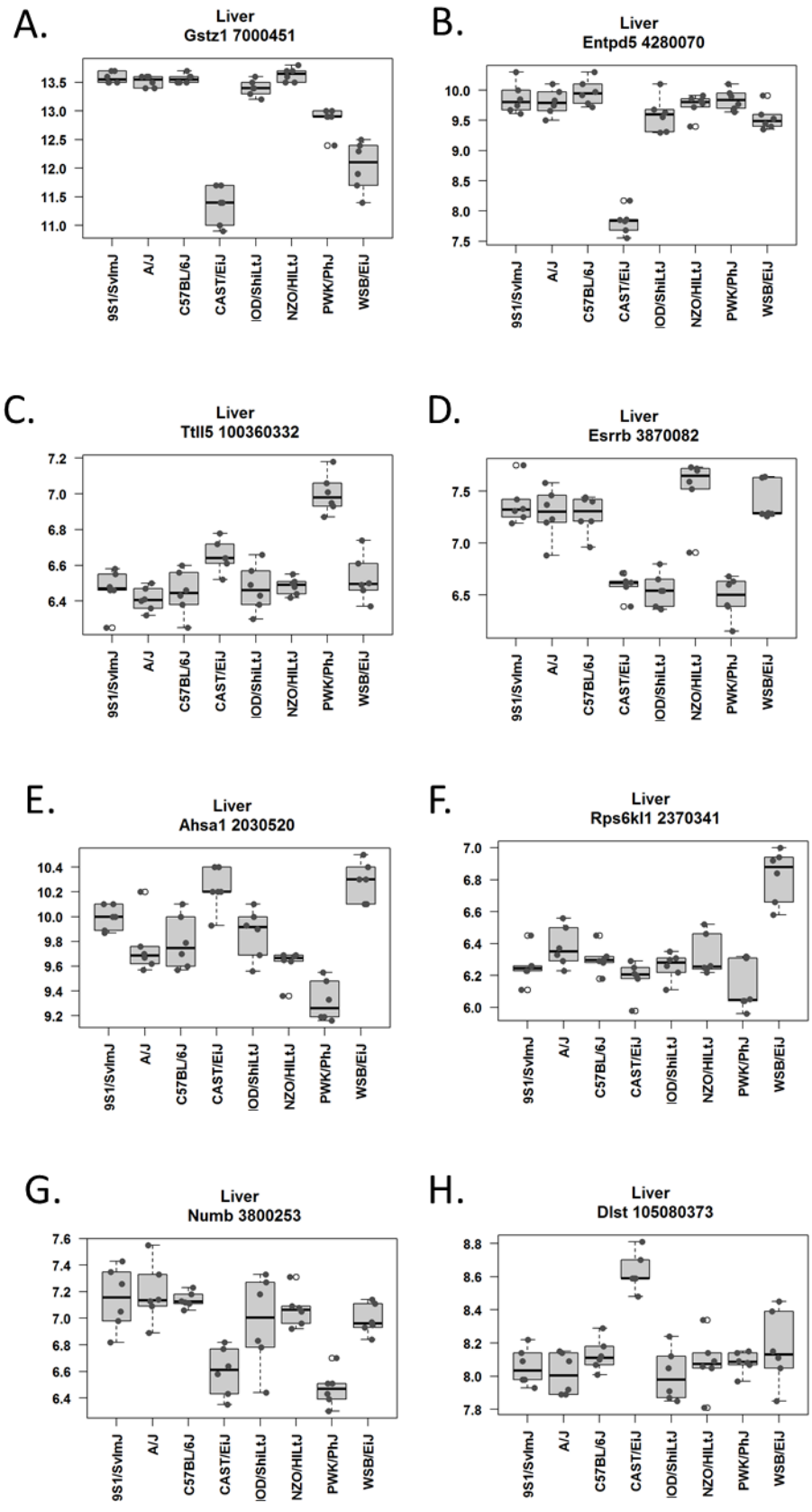

A.
Liver
Gstz1 7000451
B.
Liver
Entpd5 4280070
C.
Liver
Ttll5 100360332
D.
Liver
Esrrb 3870082
E.
Liver
Ahsa1 2030520
F.
Liver
Rps6kl1 2370341
G.
Liver
Numb 3800253
H.
Liver
Dlst 105080373
9S1/SvImJ
A/J
C57BL/6J
CAST/EiJ
IOD/ShiLtJ
NZO/HlLtJ
PWK/PhJ
WSB/EiJ

Figure 3.5- Differentially expressed positional candidates within the Chromosome 12 QTL associated with TMAO. Hepatic gene expression across the DO founder strain mice. We identified the 8 most differentially expressed genes as *Gstz1*, *Entpd5*, *Ahsa1*, *Rps6kl1*, *Ttll5*, *Esrrb*, *Numb*, and *Dlst* (A.-H.). *Numb* most closely matches the founder allele effects identified by QTL mapping.

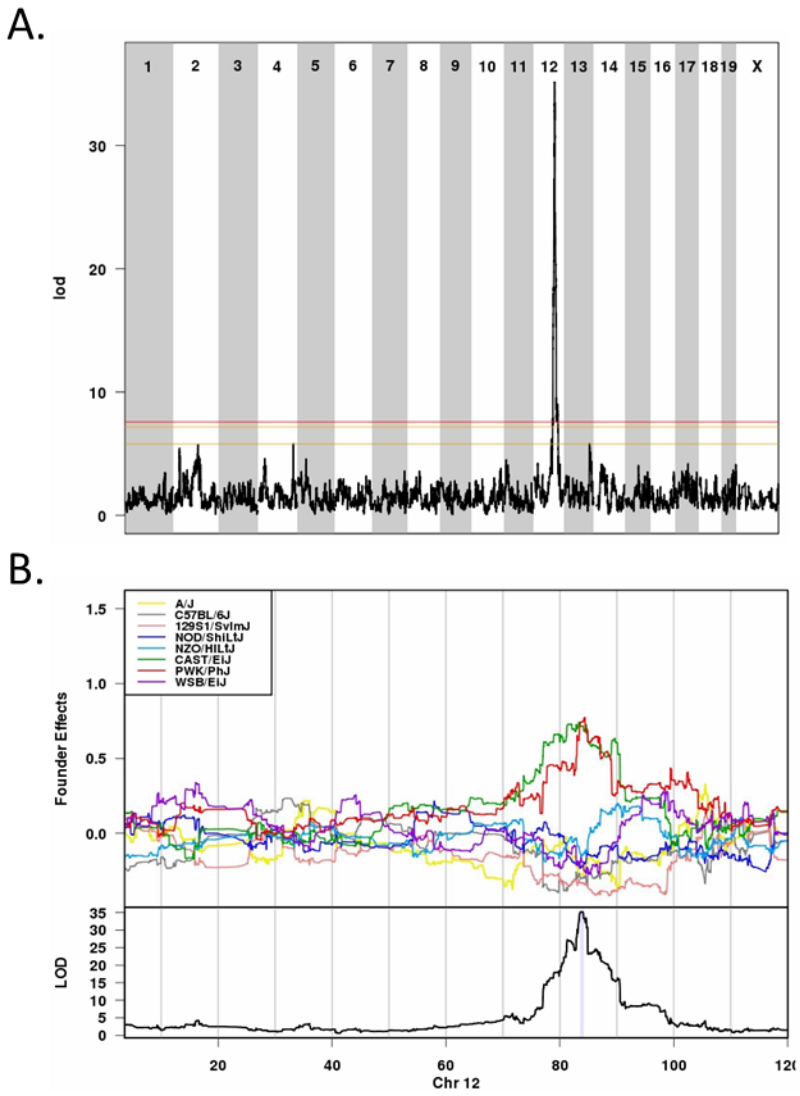


Figure 3.6: Identification of a *cis*-eQTL for *Numb* in the DO mice. Genome-wide QTL scan for eQTL regulating expression of *Numb* in the DO mice. Chromosomes 1 through X are represented numerically on the *x*-axis and the *y*-axis represents the LOD score. The relative width of the space allotted for each chromosome reflects the relative length of each chromosome. Colored lines show permutation-derived significance thresholds at $P = 0.05$ (red), $P = 0.10$ (orange), and $P = 0.63$ (yellow). The eight coefficients of the QTL model show the effects on the phenotype contributed by each founder haplotype on Chromosome 12. Shading identifies the 95% Bayesian credible interval around the peak (B.). Hepatic expression of *Numb* across the founder strains and the effect of the HFCA diet on *Numb* expression in the progenitor strain mice. The HFCA diet induces *Numb* expression in a strain-specific manner, specifically *Numb* expression was significantly increased in response to diet in 129/SvImJ, CAST/EiJ, and PWK/EiJ strains.

Table 3.1: Summary of QTL identified as associated with choline and TMAO in the DO.

Phenotype		Chr	LOD	Peak SNP	Peak Position Mb (Bayesian interval)	Positional Candidates
TMAO	Baseline	12	10	UNC21547552	86.3 (83.6-88.2)	116
		14	8.2	UNC24586024	98.6 (83.3-99.6)	49
	After diet	12	6.7	UNC21539522	85.7 (83.1-87.1)	85
Choline	Baseline	12	7.3	UNC21862809	109 (99.4-110.2)	279
	After diet	12	8.1	UNC21982912	116.8 (113.2-117.5)	31

Table 3.2: SNP analysis of Chromosome 12 QTL associated with TMAO at baseline.

SNP Type	Total SNPs	CAST, PWK shared SNPs
3' UTR variant	1301	259
5' UTR variant	229	55
Missense variant	302	78
Insertion	114	16
Deletion	158	35
Splicing variant (includes acceptor, donor, and region)	96	26
Stop (includes gained, lost, and retained)	3	0

Table 3.3: SNP analysis of known variants within 10 kb of the candidate gene *Numb* on Chromosome 12.

SNP types	Total number (Number specific to CAST/EiJ and PWK/PhJ)
3' UTR variant	17(14)
5' UTR variant	1(0)
Downstream gene variant	291 (204)
Intergenic variant	28(12)
Intron variant	1313(779)
Missense variant	4(2)
Insertion	8 (2)
Deletion	1 (1)
NMD transcript variant	1222 (750)
Nc transcript variant	514(233)
Non coding exon variant	1(0)
Regulatory region variant(includes ablation, amplification, and variant)	0(0)
Splicing variant (includes acceptor, donor, and region)	0(0)
Stop (includes gained, lost, and retained)	0(0)
TF binding site (includes variant, ablation, and amplification)	0(0)
Transcript ablation or amplification	0(0)
Upstream gene variant	138(108)

Table 3.4: Spearman's correlations between *Numb* expression and cardiovascular traits in the DO mice.

			Baseline		After Dietary treatment		
		Numb	**TMAO**	**Cholesterol**	**TMAO**	**Cholesterol**	**Atherosclerosis**
Baseline	***Numb***	-	-	-	-	-	-
	TMAO	-0.654	-	-	-	-	-
	Cholesterol	-0.127	0.374	-	-	-	-
After Dietary Treatment	**TMAO**	-0.398	0.173	0.311	-	-	-
	Cholesterol	0.809	-0.3	0.065	-0.3	-	-
	Atherosclerosis	-0.245	0.003	0.061	0.143	0.068	-

CHAPTER 4: QUANTITATIVE TRAIT LOCUS ANALYSIS OF METABOLIC PHENOTYPES IN THE DO MICE

CHAPTER OVERVIEW

We performed linkage mapping in DO mice that were genotyped at using the high-density MegaMUGA genotyping array. Linkage mapping was performed using the 8 founder state haplotype contributions. Mice were fed either a HFCA diet or a high protein diet for 18 weeks and diet was included as an additive covariate in the mapping model for all traits which included phenotypes collected from mice at 24 weeks after dietary treatment. Here, I report and summarize the results of mapping 8 quantitative traits in this cohort of DO mice. We identified 21 suggestive loci and three significant loci ($p<0.05$) including novel loci for measures of blood urea nitrogen (BUN) and insulin resistance. The QTL on Chromosome 1 associated with BUN at baseline encompasses a 900 kb region containing only 5 genes, including the candidate gene, *Esrrg*, which has been shown to be involved in the organization of renal papillae. We were able to identify a single SNP that matches out allele effects patterns that is located in an intron of *Esrrg*. The QTL on Chromosome 8 associated with HOMA-IR in 24 week old mice is 1.8 Mb in size and encompasses 42 positional candidate genes, including the high priority candidate *Fgfr1*. These results confirm the power of the DO to identify novel loci and fine map QTL to sub-Mb genomic regions containing far fewer positional candidates for follow up study, thus expediting efforts to identify causal genes and variants underlying the significant associations.

INTRODUCTION

QTL mapping is a statistical approach used to associate genomic intervals with phenotypic traits. QTL mapping has been widely and successfully used to identify loci associated with human disease. QTL mapping studies using inbred mice have aided in the discovery of genes underlying many human disease associated QTL [154]. However, the use of inbred mice for QTL mapping is limited by the lack of allelic diversity present in inbred strains and when QTL intervals are identified they are often too large to quickly identify the underlying causal gene responsible for the genetic association, sometimes requiring many years of follow up studies before the causal gene is identified [155]. Recently, efforts have been made to develop heterogeneous and outbred populations of mice designed to be informative for QTL mapping studies. Mouse populations such as the HS and the DO are derived from multiple inbred founder strains, providing the introduction of allelic variants, and are maintained by outbreeding, resulting in the accumulation of recombination events at each successive generation of outbreeding [79, 81, 97, 98].

The DO mice are a heterogeneous population derived from the following eight inbred founder strains: A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HiLtJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ. The DO mice are maintained by randomized outbreeding as 175 outbred lines at the Jackson Laboratory in Bar Harbor, Maine and at the time of this study were at outbreeding generation 11 (received September, 2012). The allelic diversity present in the DO population represents 90% of the known allelic diversity present in inbred strains. The DO contain approximately 45 million segregating SNPs, including 37.8 million SNPs and 6.9 million insertions, deletions, and structural variants [101]. Additionally, the genetic variation captured by the DO mice is uniformly distributed throughout the mouse genome [98]. Because the DO mice

are highly heterozygous outbred animals that were designed to capture the majority of known allelic variants across the eight inbred founder strains, which include three wild-derived strains, these mice have proven to be highly informative for QTL mapping studies [97, 98, 100-103, 156, 157].

While multi-parent mapping populations like the DO present a more powerful resource for QTL mapping, they often also present new challenges in data analysis. The mosaic genomes of the DO mice are informative for mapping due to the accumulation of meiotic recombinations, but the identification of small haplotype blocks requires high-density genotyping of individual mice in order to accurately reconstruct haplotypes. For our study, the mice were genotyped using the high-density MegaMUGA array which assays 77,808 SNPs spanning the mouse genome which an average spacing of 33 kb (GeneSeek, 2013). For haplotype reconstructions and QTL mapping, we used DOQTL, which is a specialized software package that uses an HMM for haplotype reconstructions and uses the information from the HMM to input the founder allele dosages for QTL mapping [101].

The earliest DO mapping studies had much success using DOQTL to identify QTL associated with phenotypes including complex traits related to cardiovascular disease and behavioral phenotypes [97-99]. Here, I report the results of QTL mapping of 8 additional quantitative traits in this cohort of DO mice not previously reported in Chapters 2 and 3. Among our significant findings, we identify novel loci for blood urea nitrogen and HOMA-IR, a measure of insulin resistance, **Table 4.1**. We identify a single SNP in an intron of *Esrrg* underlying the Chromosome 1 QTL associated with BUN. These results highlight the usefulness of the DO as a resource for QTL mapping and the identification of causal variants underlying QTL.

MATERIALS AND METHODS

Animals and Diets

Female Diversity Outbred mice (n = 292; J:DO, JAX stock number 009376) were obtained from the Jackson Laboratory (Bar Harbor, ME) as 146 full sibling pairs at 4 weeks of age and at outbreeding generation 11 (G11) (received September, 2012). The mice were group housed (n = 5 mice per cage) with non-irradiated pine bedding and provided with HEPA-filtered air and free access to sterile water in a climate-controlled facility under a 12 hour light:dark cycle. Mice were maintained on a defined synthetic diet, AIN-76A, until 6 weeks of age to control for differences due to variable components of standard chow (D10001, Research Diets, New Brunswick, NJ); subsequently, 146 mice were transferred to a synthetic high-fat, cholic acid (HFCA) diet, containing 20% fat, 1.25% cholesterol, and 0.5% cholic acid, to induce atherosclerotic lesions and 146 mice were maintained on a high protein diet containing 5% fat and 20.3% protein which is not atherogenic (D12109C and D12083101, respectively, Research Diets, New Brunswick, NJ). One sibling from each of the 146 sibling pairs was randomly assigned to each one of the diets, as previously described in Chapter 2 of this thesis. The diet composition, including source of fat from the diets is described in Table 2.3 of this thesis. All mice were maintained on their respective diets until 24 weeks of age, for a total of 18 weeks. All procedures were approved by the IACUC at UNC Chapel Hill (IACUC Protocol Number 11-299).

Plasma clinical chemistries

At 6 weeks of age, the mice were anesthetized using isoflurane and blood was collected after 4 hours of fasting into EDTA-containing microtubes from the retro-orbital sinus. This was repeated at 24 weeks of age, after 18 weeks of dietary treatment. Blood samples were centrifuged

for 10 minutes at 10000 rpm at 4°C and stored at -80°C. The plasma levels of glucose, ALT, AST, and BUN were quantitated using a Biolis 24i Analyzer (Carolina Liquid Chemistries, Winston-Salem, NC). HDL was precipitated using the Heparin-$MgCl^2$ precipitation method. Samples were run in triplicate in a 96-well plate and optical densities were measured at 515 nm using a microplate reader and analyzed with Gen5 Data Analysis Software (Bio-Tek, Winooski, VT). LDL was calculated as the difference of HDL cholesterol subtracted from total cholesterol. Insulin was determined using the Alpco Mouse Ultrasensitive Insulin ELISA assay (Alpco Inc, Salem, NH); samples and controls were run in duplicate and optical densities were measured at 450 nm using a microplate reader and analyzed with Gen5 Data Analysis Software (Bio-Tek, Winooski, VT). The homeostatic model assessment of insulin resistance (HOMA-IR) was calculated using the equation $[(G_0 \times I_0)/405)]$, where G_0 and I_0 refer to 4 hour fasting plasma glucose and insulin [158]. Data are presented as mean $\pm$ SD and significance was determined using a Student's t-test.

Measurement of Choline Metabolites

Measurement of choline metabolites was performed by the Metabolomics Core Facility (Kannapolis, NC). In brief, plasma was extracted with three volumes of acetonitrile spiked with internal standards TMAO-d9 (DLM-4779-1, Cambridge Isotope Laboratories), creatinine-d3 (D-3689, CDN Isotopes Inc.), choline-d9 (DLM-549-1, Cambridge Isotope Laboratories) and betaine-d9 (616656, Sigma-Aldrich), incubated on ice for 10 minutes, and centrifuged at 15,000 g for 2 minutes. Quantification of TMAO, creatinine, choline and betaine was performed using liquid chromatography-stable isotope dilution-multiple reaction monitoring mass spectrometry (LC-SID-MRM/MS). Chromatographic separations were performed on an Atlantis Silica HILIC 3µm 4.6×150mm column (Waters Corp, Milford, USA) using a Waters ACQUITY UPLC

system. The column was heated to 40°C, and the flow rate was maintained at 1 mL/min. The gradient was 5% A for 0.05 min, to 15% A in 0.35 min, to 20% A in 0.6 min, to 30% A in 1 min, to 45% A in 0.55 min, to 55% A in 0.05 min, at 55% A for 0.9 min, to 5% A in 0.05 min, at 5% A for 1.45 min, where A is 10% acetonitrile/90%water with 10 mM ammonium formate. 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 from its calibration curve using peak area ratio of the metabolite to its isotope.

Genotyping

DNA was extracted and purified from tail biopsies taken from 6-week-old mice using Qiagen DNeasy kit according to the manufacturer's instructions. Genotyping was performed using the Mega Mouse Universal Genotyping Array (MegaMUGA) by GeneSeek (Neogen, Lansing, MI) [113]. The call rate exclusion criteria was set at >95% and twelve mice were excluded based on this criteria; the average call rate of the genotyped mice used in the study was 98%. The MegaMUGA array is built on the Illumina Infinium platform and contains 77,808 SNP markers that are distributed throughout the genome at an average spacing of 33 Kb. The MegaMUGA SNPs were subset to include 57,977 informative SNPs that distinguish among the genotypes of the eight founder strains. For the mapping, genomes were reconstructed based on the X and Y allele intensities from the array and founder haplotypes were reconstructed using a hidden Markov model. The founder allele dosages based on the reconstructed haplotypes were then used to perform linkage mapping.

QTL Mapping

QTL mapping was performed using the R package DOQTL version 1.0.0 (DOQTL Bioconductor release, 2014). DOQTL reconstructs the genome in terms of founder haplotypes and performs QTL mapping by regressing the phenotypes on the founder haplotypes with an adjustment for kinship between the mice. Phenotypes were natural log-transformed to satisfy the model assumption of a normal distribution. Diet was included as an additive covariate in the mapping model for measurements including phenotypes obtained from 24 week old mice after dietary treatment. The mapping statistic reported is log of the odds ratio (LOD). QTL support intervals were defined by the 95% Bayesian credible interval, calculated by normalizing the area under the QTL curve on a given chromosome [115]. The significance thresholds were determined by performing 1000 permutations of genome-wide scans by shuffling phenotypic data in relation to individual genotypes. Significant QTL were determined at a genome-wide p-value of <0.05 and suggestive QTL were determined at a p-value of <0.63. The latter corresponds to one false positive per genome scan [116]. Candidate genes were identified by position based on the Wellcome Trust Sanger mouse genomes database, www.sanger.ac.uk, release 1303 based on genome assembly GRCm38 [114]. Association mapping was performed for association of the phenotype with known SNPs within the identified QTL by imputing the founder SNPs onto the DO genomes.

RESULTS

Effects of HFCA Diet on Diversity Outbred Mice

The overall aim of this study was to identify loci associated with atherosclerotic lesion development and cardiovascular risk factors in the DO mice. In order to induce lesions in these mice, we fed mice a synthetic HFCA diet designed to induce lesion formation but reduce adverse pathological effects on liver and kidney function [41]. In order to assess liver and kidney

function, the mice were phenotyped for ALT, AST, and BUN. Prior to dietary treatment, there were no significant differences in ALT, AST, or BUN in the mice at baseline, **Figure 4.1**. After dietary treatment, ALT increased significantly in response to both the high protein diet ($p < 5x10^{-10}$) and the HFCA diet ($p < 2x10^{-34}$), **Figure 4.1A**. AST increased significantly in response to both the high protein diet ($p < 1x10^{-15}$) and the HFCA diet ($p < 6x10^{-13}$), **Figure 4.1B**. BUN increased significantly in response to both the high protein diet ($p < 2x10^{-19}$) and the HFCA diet ($p < 4x10^{-4}$), **Figure 4.1C**. The increase in ALT corresponds to a 3-fold increase in ALT levels in response to the HP diet and a 7-fold increase in response to the HFCA diet. Previous studies have indicated a 10-fold increase in plasma ALT as indicative of fatty liver disease [159]. While we did observe hepatic steatosis in some of the mice, the overall impact of the diet did not increase mortality. Specifically, 1% of mice died prior to the end of the study (N=5) and of those only one mouse was receiving the HFCA diet.

Identification of Chromosome 1 QTL associated with BUN in 6 week old mice

We detected two loci associated with BUN in 6 week old DO mice. A significant QTL interval on Chromosome 1 (LOD= 7.6, $p < 0.05$) and a suggestive QTL interval on Chromosome 16 (LOD= 7.2, $p < 0.1$), **Figure 4.2A**. The Chromosome 1 QTL has a support interval of only 900 kb with a peak SNP at 187.5 Mb (187.0- 187.8 Mb). The allele effects of the significant QTL revealed that allelic contribution from the DO founder strains A/J, C57BL/6J, and NOD/ShiLtJ is associated with lower BUN levels, while allelic contribution from the DO founder strains 129S1/SvImJ, NZO/HiLtJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ are associated higher BUN levels, **Figure 4.2B**.

This interval contains only 6 positional candidates, including the known genes *Spata17*, *Gpatch2*, and *Esrrg*, as well as the predicted genes *Gm15509*, *Gm21710*, and *9330162B11Rik*,

Figure 4.2C. Of the 19,530 genotyped SNPs within the interval, only one SNP, *rs32769253*, is shared by and unique to the A/J, C57BL/6J, and NOD/ShiLtJ founders as compared to the other five founder strains. According to the Sanger SNP database, this SNP represents a C/T polymorphism located in an intron of the *Esrrg* gene at 187,919,386 Mb (GRC Build 38). The strains A/J, C57BL/6J, and NOD/ShiLtJ share a T at this position, while the strains 129S1/SvImJ, NZO/HiLtJ, CAST/EiJ, PWK/PhJ/ and WSB/EiJ exhibit a shared T->C variant. Interestingly, *Esrrg* has been shown to regulate the morphology of renal papilla [160]. The intronic SNP we have identified here may affect enhancer binding and expression of *Esrrg*, although we cannot determine the functional relevance of this SNP without follow up studies. Therefore, follow up studies on the impact of this variant on *Esrrg* expression and more generally on the role of *Esrrg* in regulating BUN levels are warranted.

Identification of QTL associated with HOMA-IR in 24 week old mice

We were unable to identify loci significantly associated with independent measures of either insulin or glucose. These results were previously discussed in chapter 2 of this thesis, **Table 2.3**. Therefore, we decided to use measurements of fasting glucose and fasting insulin to calculate HOMA-IR in the DO mice. HOMA-IR is an indicator of insulin resistance that has been widely used clinically since it was described in the early 80s [161]. Previous work out of our lab identified significant strain-specific variation in HOMA-IR across the 8 founder strains used to create the DO, suggesting that this variation should be present in the DO mice [144]. Indeed, we observe HOMA-IR ranging from 0.04-1.4 in 6 week old mice fed a synthetic diet. Prior to transferring mice to either the HFCA or the high protein diet, there were no significant differences in HOMA-IR. After dietary treatment, HOMA-IR ranged from 0.03-3 in 24 week old mice and there were significant differences between the diet groups ($p < 0.01$).

We identified a significant QTL on Chromosome 8 (LOD= 7.7, p <0.05) associated with HOMA-IR in 24 week old DO mice and a suggestive QTL on Chromosome 6 (LOD= 6.6, p <0.63), **Figure 4.3A**. The Chromosome 8 QTL encompasses a 1.8 Mb interval (24.3-26.1 Mb). The effects of the founder alleles within this interval do not appear to fall out into two distinct categories. We identify C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HiLtJ, and PWK/PhJ alleles within this interval as associated with high HOMA-IR levels; A/J and CAST/EiJ associated with intermediate HOMA-IR levels; and WSB/EiJ alleles are associated with low HOMA-IR levels, **Figure 4.3B**. This data indicates that there may be more than one SNP type within the interval that is associated with HOMA-IR. There are 42 positional candidates within this QTL interval, **Figure 4.3C**. There are 24,366 genotyped SNPs within the Bayesian credible interval including SNPs that affect splicing, SNPs in 3' UTRs, SNPs in 5' UTRs, non-synonymous coding SNPs, multiple insertions and deletions, and one SNP shared by C57BL/6J and PWK/PhJ that results in the loss of the termination codon resulting in a long isoform of the gene *Adam3*, **Table 4.3**.

While we cannot currently exclude the other 41 genes in this region as candidates, we are particularly interested in the gene that encodes the fibroblast growth factor receptor 1, *Fgfr1*, as a candidate gene for regulating HOMA-IR. Fgfr1 is a transmembrane receptor which is activated by Fgf21 and Fgf21 has been shown to improve insulin sensitivity and reduce circulating glucose levels [162-164]. Additionally, the use of Fgfr1 agonists has recently been proposed for the treatment of metabolic syndrome and non-alcoholic fatty liver disease [164]. Therefore, *Fgfr1* is a high priority candidate for this QTL.

There are 1101 variants within 10 kb of the *Fgfr1* gene that may regulate expression or function of this gene, including 25 in the 3' UTR, 10 in the 5' UTR, 5 splice variants, and one

WSB/EiJ-specific non-synonymous coding variant, **Table 4.4**. Based on our allele effects, we identified WSB/EiJ alleles within the Chromosome 8 QTL interval as associated with low HOMA-IR compared to the other founder strains. While this QTL could result from a change in expression of the causal gene, the presence of a WSB-EiJ-specific variant in *Fgfr1* is intriguing. The non-synonymous SNP, *rs47309288* at position 25,555,535 Mb, represents a T→C transition specific to WSB/EiJ within the coding sequence of *Fgfr1* resulting in a valine to alanine amino acid substitution that affects five of the eight protein-coding transcripts of *Fgfr1*. The AA substitution occurs at position 40 in the protein sequence resulting from the *Fgfr1-001, Fgfr1-002, Fgfr1-012,* and *Fgfr1-013* transcripts and at position 53 of the protein sequence resulting from the *Fgfr1-202* transcript. The functional significance of this AA change and more generally of *Fgfr1* in regulating insulin resistance should be further explored.

Chromosome 9 QTL associated with change in total cholesterol exhibits complex genetic regulation by multiple SNP types.

We previously reported a 22.2 Mb QTL interval on Chromosome 9 (LOD = 7.54, p <0.05) associated with cholesterol after dietary treatment, **Figure 2.4A and Table 2.2**. This region of Chromosome 9 has been identified as associated with triglycerides and cholesterol in several previous studies and the regulation of lipid metabolism by genes in this region is thought to be complex. The *ApoA5-ApoA4-ApoC3-ApoA1* gene cluster is located at 46.2 Mb, just proximal to the 22.2 Mb QTL interval associated with cholesterol after dietary treatment. Here, we used total cholesterol levels in 6 and 24 week old mice to determine the change in total cholesterol from 6 to 24 weeks and then mapped this quantitative trait. We identified the same Chromosome 9 QTL as significantly associated with the change in total plasma cholesterol from

6 to 24 weeks (LOD= 7.8, p <0.05) and the 25 Mb Bayesian defined interval includes the *ApoA5-ApoA4-ApoC3-ApoA1* gene cluster, **Figure 4.5A**.

The founder allele effects in this region appear complex, likely due to multiple variants with different strain distribution patterns influencing the phenotype. Within the interval, CAST/EiJ private alleles between 45.3- 60 Mb are associated with an increase in cholesterol over time, while CAST/EiJ and 129S1/SvImJ shared alleles between 60-70.3 Mb are associated with an increase in cholesterol over time, **Figure 4.5B**. These results suggest that there may be at least two types of SNPs influencing cholesterol metabolism within this interval. In order to test this, we performed association mapping in the DO mice by imputing the Sanger allele calls onto the DO genomes. We then fit a linear regression model, similar to the additive haplotype model used to perform linkage mapping, but with 2-state SNP parameters in the model and estimated the additive effect of allelic substitution [101]. There are 91,134 CAST/EiJ-specific alleles in the QTL interval and 18,391 of these are highly associated with change in total cholesterol (LOD > 4 for the additive SNP model). There are 4,314 CAST/EiJ and 129S1/SvImJ shared alleles within the interval and 22 of these are highly associated with change in total cholesterol (LOD > 4 for the additive SNP model). As expected based on our allele effects, the CAST/EiJ alleles associated with the phenotype are located between 47,929,312- 51,417,095 Mb and the CAST/EiJ and 129S1/SvImJ alleles associated with the phenotype are located between 69,353,741- 69,530,721 Mb, **Figure 4.5C**.

In order to determine if these are indeed independent loci, we regressed out the effect of the CAST/EiJ allele by adding the peak SNP, *UNC16319105*, as a covariate in the QTL mapping model. Indeed, when we regress out the peak SNP, the proximal QTL region no longer reaches significance and the LOD score for the peak SNP of the distal QTL region associated with the

phenotype increases, **Figure 4.5D**. The effect of the CAST/EiJ alleles between 45.3- 60 Mb is diminished but the Bayesian interval is unchanged, **Figure 4.5E**. When we performed association mapping, we no longer see the CAST/EiJ alleles associated with the phenotype between 47,929,312- 51,417,095 Mb, while the same 22 CAST/EiJ and 129S1/SvImJ shared alleles within the interval are still highly associated with the phenotype (LOD > 4 for the additive SNP model), **Figure 4.5F**.

Identification of suggestive QTL associated with various metabolic phenotypes

We also identified 21 QTL that reached the suggestive level of significance and these are summarized in **Table 4.2**. More than one suggestive QTL was identified for measures of baseline ALT, AST, and creatinine, as well as glucose: insulin ratio after dietary treatment and creatinine after dietary treatment. Although these results represent suggestive QTL that did not reach significance in our study, previous studies using the DO mice have identified and validated suggestive QTLs of less than 10 Mb. For example, Recla et al. identified a 3.8 Mb QTL on Chromosome 8 (LOD= 6, p <0.63) containing 44 positional candidates and identified *Hydin* as a high priority candidate gene associated with acute pain thermal sensitivity [100]. Of our 21 suggestive QTL, 9 suggestive QTL have a Bayesian credible support interval of less than 10 Mb and 7 suggestive QTL contain less than 100 positional candidate genes. Further analysis of these intervals may aid us in prioritizing genes underlying these loci. Phenotyping additional DO mice for traits exhibiting such suggestive QTL that are trending toward genome-wide significance is one approach to assess the validity of suggestive QTL prior to conducting follow up studies.

DISCUSSION

In order to identify and take advantage of genetic differences in susceptibility to atherosclerosis, cardiovascular disease risk is often determined to identify risk factors known to

be associated with this condition. In our study, we characterized atherosclerosis and several cardiovascular risk factors in DO mice fed either an atherogenic or a high protein diet. In Chapter 2 of this thesis, I have reported QTLs identified as associated with atherosclerosis, glucose, insulin, triglycerides, and total cholesterol in the DO mice. In Chapter 3 of this thesis, I reported QTLs associated with the metabolites TMAO and choline. Here, I report the results of QTL mapping of 8 additional quantitative traits in this cohort of DO mice. We identified three significant QTL ($p< 0.05$) including novel loci for blood urea nitrogen and insulin resistance, as well as 21 suggestive loci. We discuss each of these results in turn.

Identification of Chromosome 1 QTL associated with BUN in 6 week old mice

The Chromosome 1 QTL associated with BUN at baseline is a prime example of the power of the DO to detect QTL to sub-Mb support intervals and expedite the identification of the underlying causal variants responsible for the observed association. We were interested in detecting BUN levels as a measure of kidney function in these mice to assess the effect of diet on overall health. In our original study design, we did not use these mice to model human kidney disease; therefore no validation studies of the 6 candidate genes within this interval have been pursued to date. However, one of these candidate genes, *Esrrg*, has been shown to regulate the morphology of renal papilla. *Esrrg* encodes the estrogen-receptor related gamma protein, which is a nuclear receptor that contains DNA-binding domains and is a transcriptional activator. *Esrrg* $^{-/-}$ knockout mice exhibit deformed renal papillae and exhibit complete postnatal lethality [160]. Biologically, the association of *Esrrg* with plasma BUN is intriguing, as renal papillae are composed of the collecting ducts of the nephrons where BUN is reabsorbed. Therefore, this gene may affect the way the nephrons are organized or possibly the permeability of the ducts.

Our allele effects indicated that a variant shared by the A/J, C57BL/6J, and NOD/ShiLtJ founder strains may be underlying this QTL. We identified only one SNP, *rs32769253*, matching the allele effects pattern that we observed, such that the strains A/J, C57BL/6J, and NOD/ShiLtJ share a T at this position, while the strains 129S1/SvImJ, NZO/HiLtJ, CAST/EiJ, PWK/PhJ/ and WSB/EiJ exhibit a shared T->C variant. This SNP is located in an intron of *Esrrg*, between exons 2 and 3 of the isoforms *Esrrg-003* and *Esrrg-004*. The variant consequence is identified within the Sanger database as intronic, based on the Ensembl Variant Effect Predictor. However, it is just upstream of the coding sequences of the *Esrrg-001* and *Esrrg-002* isoforms, indicating it could play a functional role in processing of these transcripts. While there are 46 QTL previously identified that encompass this region of mouse Chromosome 1, none have been identified as associated with BUN levels or kidney function (rgd.mcw.edu). Therefore, the significant QTL associated with baseline BUN represents a novel locus that has not been previously characterized. The identification of such a refined interval associated with this phenotype demonstrates the power of the DO mice to expedite candidate gene testing and the determination of the functional significance between *Esrrg* and BUN levels warrants further study.

Chromosome 8 locus associated with HOMAIR in 24 week old mice

We identified a significant QTL on Chromosome 8 (LOD= 7.7, $p< 0.05$) associated with HOMA-IR in 24 week old mice. One positional candidate gene within the QTL interval is *Fgfr1*. Fgfr1 is a member of the fibroblast growth factor receptor family that is activated by fibroblast growth factors. While there are four Fgf receptors (Fgfr1-4), there are 22 Fgfs which are involved in a many diverse cellular functions [165]. The activity of Fgfs is regulated in part by ligand binding affinity and tissue distribution of Fgf receptors [166]. There are eight isoforms of Fgfr1 in mice and alternative mRNA splicing leads to complex tissue-specific regulation of

Fgfr1 function [167]. In adipocytes, Fgfr1 forms a complex with ß-Klotho and is activated by Fgf19 and Fgf21 [162, 163]. Fgf21 is expressed in adipose and liver tissue and it acts on adipose and muscle tissue primarily to release stored energy during times of stress, such as fasting. Fgf21 stimulates glucose uptake and it has been shown that pharmacological administration of Fgf21 lowers glucose levels. Additionally, the use of Fgfr1 agonists has recently been proposed for the treatment of metabolic syndrome and non-alcoholic fatty liver disease [164]. Therefore, Fgfr1 plays a functional role in the regulation of glucose uptake at least in part via its activation by Fgf21 and exhibits complex regulation by multiple Fgfs. Although we have not ruled out the 41 other positional candidates, we have identified *Fgfr1* as a high priority candidate gene based on its known biological function in the maintenance of glucose homeostasis.

Chromosome 9 QTL associated with change in total cholesterol exhibits complex genetic regulation by multiple SNP types.

Previous reports have identified at least 6 QTL on mouse Chromosome 9 associated with total cholesterol, including the overlapping QTL *Cq4* and *Cq5*, and this same region of Chromosome 9 has been associated with triglycerides as well [123]. In our study alone, we identified 4 coincident QTL on Chromosome 9 associated with measures of triglycerides and cholesterol with overlapping intervals and peak SNPs less than 3 Mb apart. Here, we reported the QTL associated with the change in total plasma cholesterol from 6 to 24 weeks. This QTL has a 25 Mb support interval with a peak SNP at 48.3 Mb (45.3-70.3 Mb) and it includes the *ApoA5-ApoA4-ApoC3-ApoA1* gene cluster located at 46.2 Mb.

Based on the observed founder allele effects, we identified complex genetic regulation with contribution from multiple variants with different strain distribution patterns within this interval. Using an additive SNP model, we identified two distinct regions within the QTL

interval that are associated with variants exhibiting different strain distribution patterns, with at least one CAST/EiJ-specific allele contributing to the proximal QTL region and at least one CAST/EiJ and 129S1/SvImJ shared allele contributing to the distal QTL region. When we regress out the effect of the CAST/EiJ-specific peak SNP, the effect of the CAST/EiJ alleles at the proximal peak is diminished; while the CAST/EiJ and 129S1/SvImJ shared alleles at the distal peak are still highly associated with the phenotype. Our results showed that there are two variants within this interval.

Taken together, it appears that the QTL we identified on Chromosome 9 represents at least two separate and distinct QTLs that have different causal variants underlying the associations. Additionally, when we use the 2-LOD support interval to identify the boundaries of the QTL region, these peaks are distinct. The proximal Chromosome 9 QTL has a 2.4 Mb support interval (47.9-50.3 Mb) with the peak SNP located at 48.3 Mb (LOD= 7.8, $p< 0.05$), while the distal Chromosome 9 QTL has a 8.5 Mb support interval (64.7-73.2 Mb) with the peak SNP located at 70.2 Mb (LOD= 6.2, $p< 0.63$). Our results corroborate the previous findings that there are multiple underlying variants within this region of Chromosome 9 that are regulating total cholesterol.

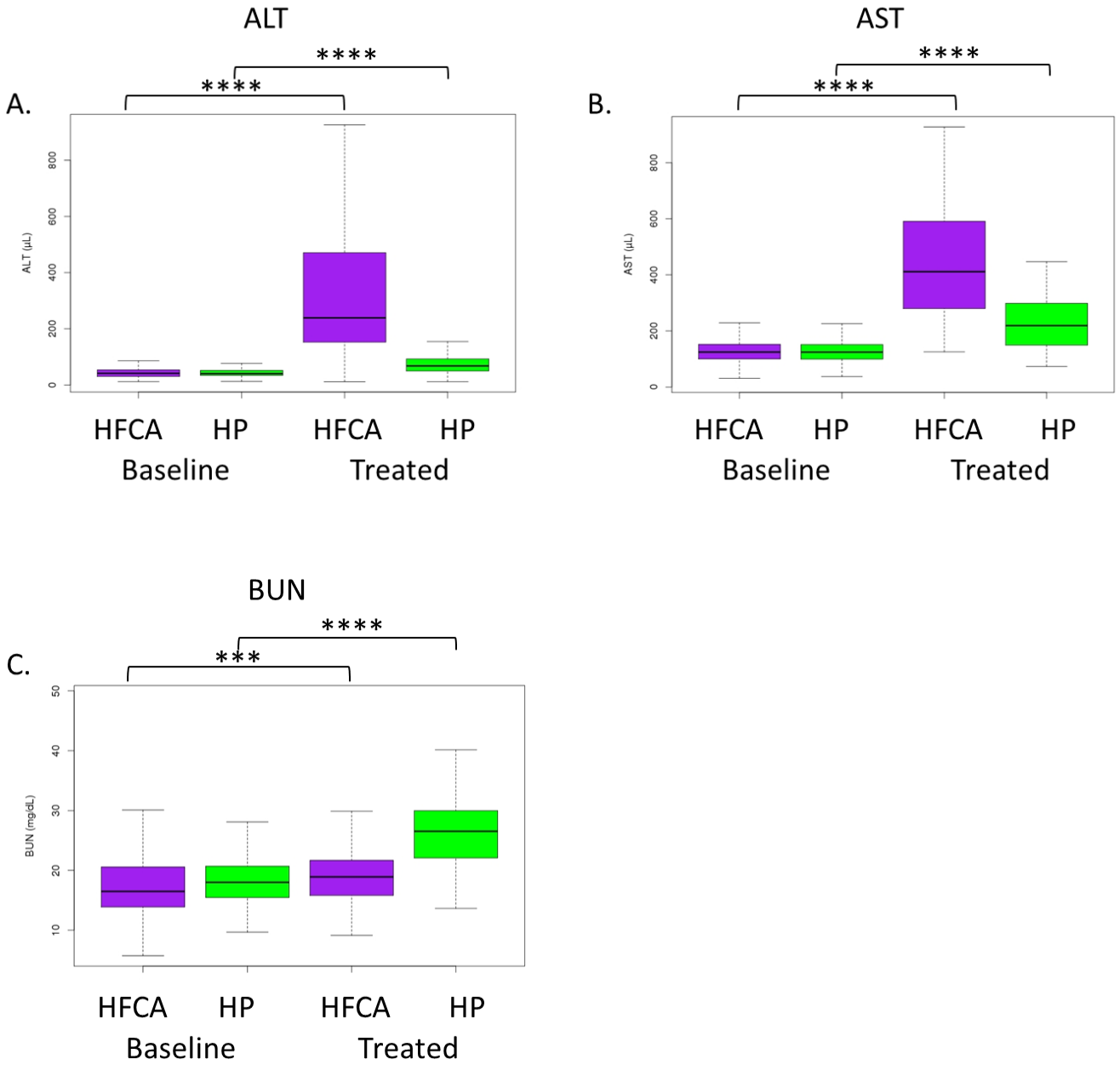


Figure 4.1: Effects of HFCA diet and high protein diet on ALT, AST, and BUN levels in the DO mice. Mice were maintained on a synthetic diet for two weeks, fasted for four hours, and then phenotyped for plasma clinical chemistries at 6 weeks of age (Baseline). Following two weeks of synthetic diet, mice were transferred to either a high protein diet (HP) or an atherogenic diet (HFCA). Plasma was taken from 24-week-old mice after 18 weeks on their respective diets, and with four hours fasting, and then phenotyped for plasma clinical chemistries after diet treatment (Treated). After dietary treatment, ALT increased significantly in response to both the high protein diet ($p < 5x10^{-10}$) and the HFCA diet ($p < 2x10^{-34}$). AST increased significantly in response to both the high protein diet ($p < 1x10^{-15}$) and the HFCA diet ($p < 6x10^{-13}$). BUN

increased significantly in response to both the high protein diet ($p < 2x10^{-19}$) and the HFCA diet ($p < 4x10^{-4}$). The increase in ALT corresponds to a 3-fold increase in ALT levels in response to the HP diet and a 7-fold increase in response to the HFCA diet (***$p < 0.001$, ****$p < 0.0001$).

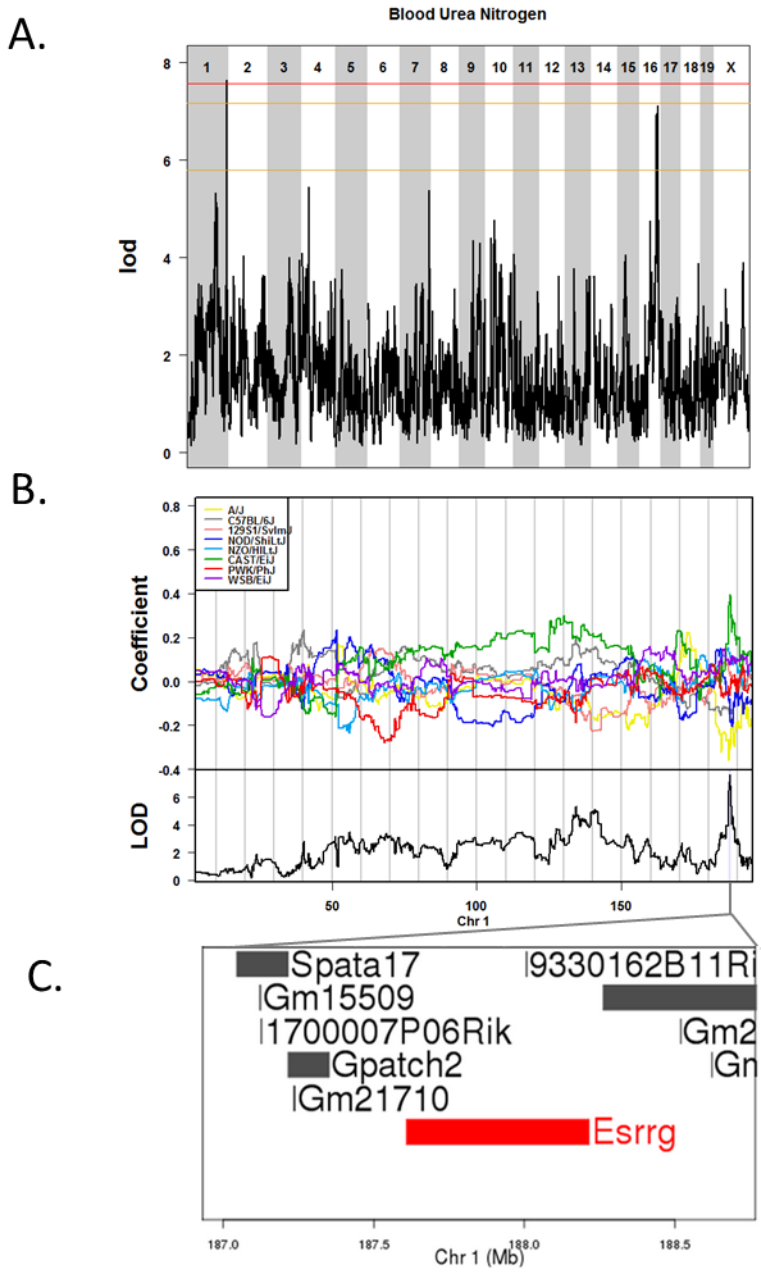


Figure 4.2: QTL mapping of blood urea nitrogen levels in DO mice at baseline. Genome-wide QTL scan for loci affecting plasma levels of blood urea nitrogen in 6 week old mice (A.). Chromosomes 1 through X are represented numerically on the *x*-axis and the *y*-axis represents the LOD score. The relative width of the space allotted for each Chromosome reflects the relative length of each Chromosome. Mice were maintained on a synthetic diet for 2 weeks and then phenotyped for plasma clinical chemistries at 6 weeks of age. Colored lines show permutation-derived significance thresholds (N=1000) at $P = 0.05$ (LOD=7.5, shown in red), $P =$

0.10 (LOD=7.1, shown in orange), and *P* = 0.63 (LOD=5.8, shown in yellow). The eight coefficients of the QTL model show the effect of each founder allele on the phenotype along Chromosome 1. Shading identifies the 95% Bayesian credible interval around the peak (B.). The 900 kb QTL interval on Chromosome 1 contains 6 positional candidates, including the known genes *Spata17*, *Gpatch2*, and *Esrrg*, as well as the predicted genes *Gm15509*, *Gm21710*, and *9330162B11Rik*. *Esrrg* is a high priority candidate for this QTL.

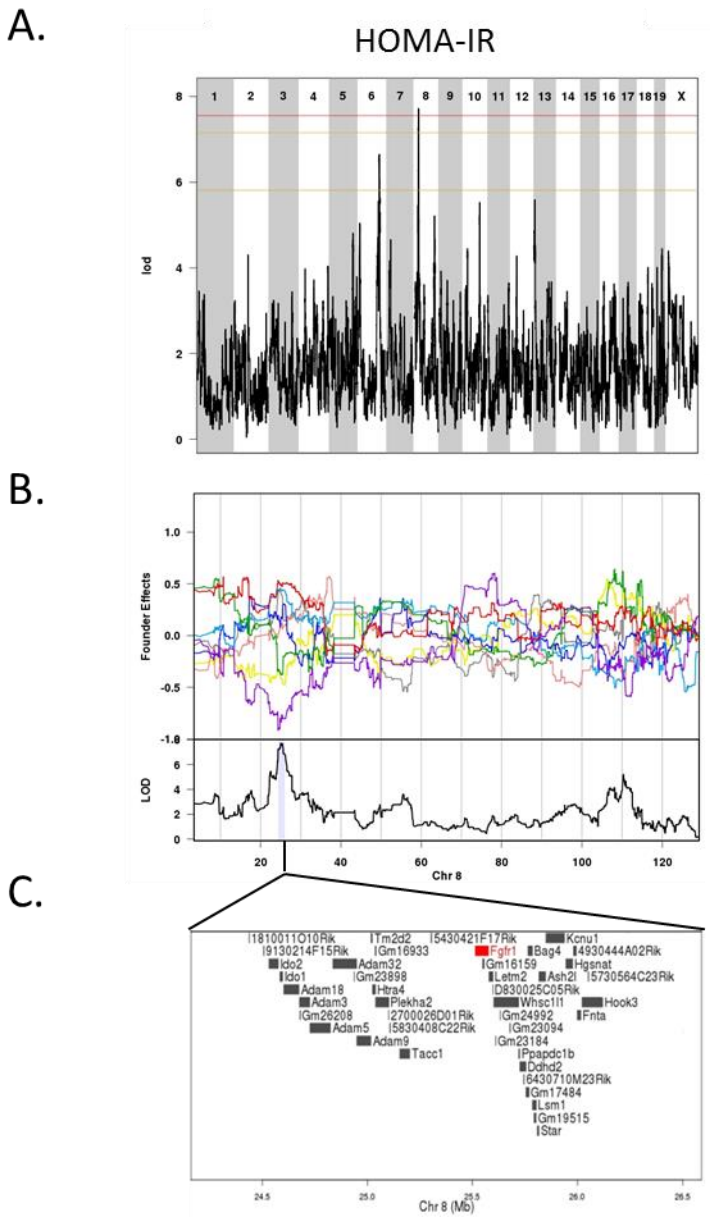
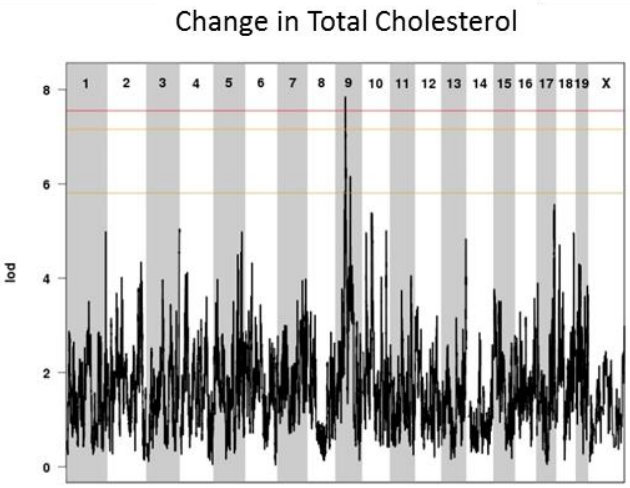


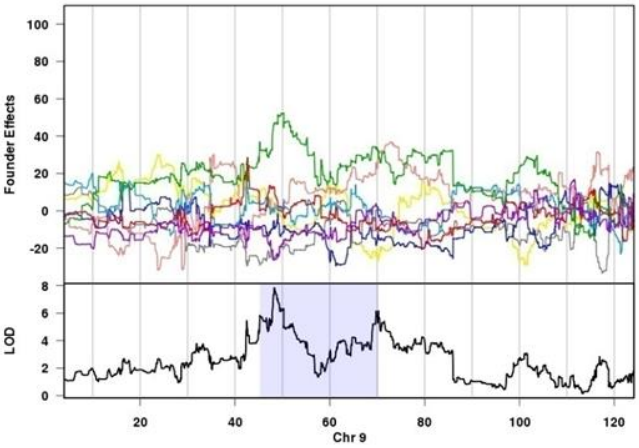
Figure 4.3: Mapping of HOMA-IR in 24 week old mice. Genome-wide QTL scan for loci affecting calculated HOMA-IR in 24 week old mice (A.). Chromosomes 1 through X are represented numerically on the *x*-axis and the *y*-axis represents the LOD score. The relative width of the space allotted for each chromosome reflects the relative length of each chromosome. Plasma was taken from 24-week-old mice after 18 weeks of dietary treatment. Diet was included as an additive covariate in the mapping model. Colored lines show permutation-derived significance thresholds (N=1000) at $P = 0.05$ (LOD=7.5, shown in red), $P = 0.10$ (LOD=7.1, shown in orange), and $P = 0.63$ (LOD=5.8, shown in yellow). The eight coefficients of the QTL

model show the effect of each founder haplotype on the phenotype long Chromosome 8. Shading identifies the 95% Bayesian credible interval around the peak (B.). There are 42 candidate genes within the 1.8 Mb QTL interval on Chromosome 8, including *Fgfr1* (C.).

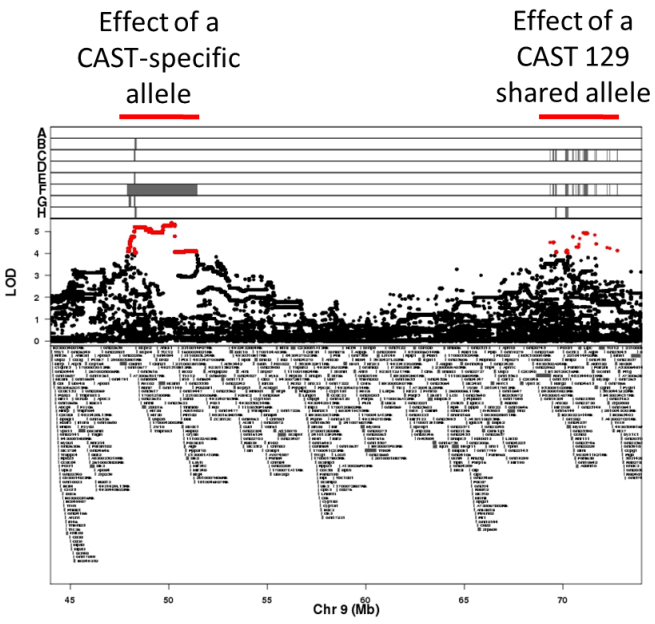
A.


Change in Total Cholesterol
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 X
lod
0
2
4
6
8

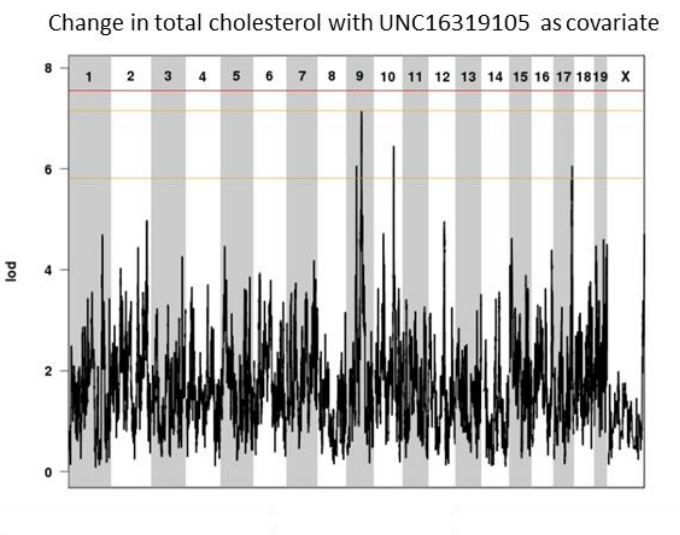
B.


Founder Effects
100
80
60
40
20
0
-20
LOD
8
6
4
2
0
20
40
60
80
100
120
Chr 9

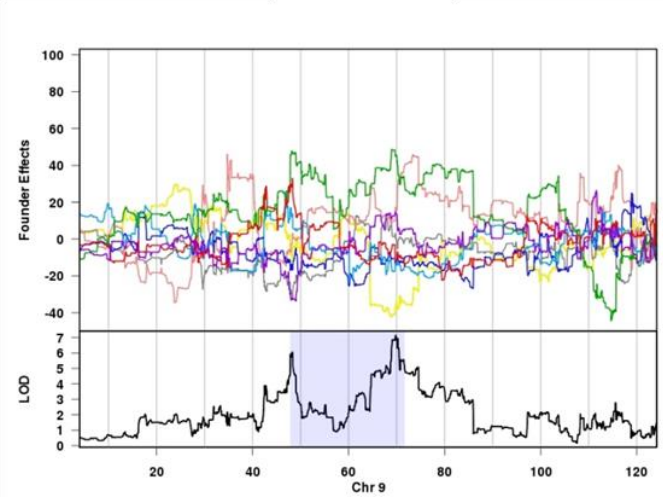
C.


Effect of a CAST-specific allele
Effect of a CAST 129 shared allele
A
B
C
D
E
F
G
H
LOD
0
1
2
3
4
5
45
50
55
60
65
70
Chr 9 (Mb)

D.


Change in total cholesterol with UNC16319105 as covariate
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 X
lod
0
2
4
6
8

E.


Founder Effects
LOD
20 40 60 80 100 120
Chr 9

F.

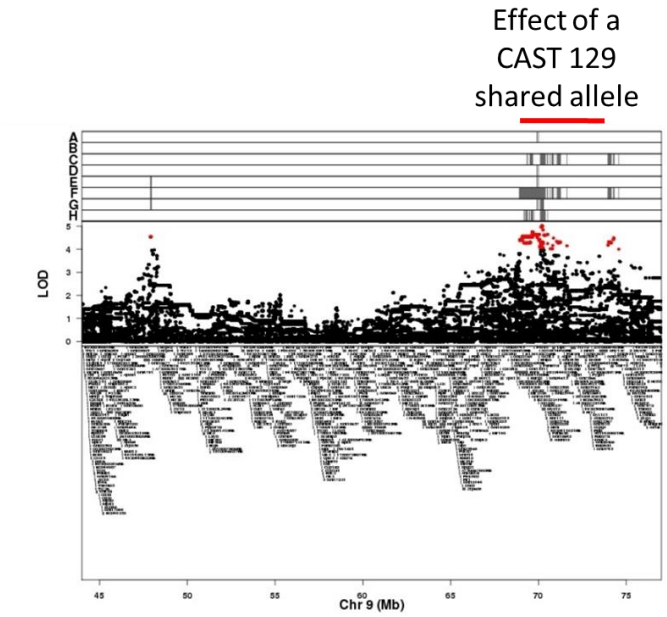

Effect of a CAST 129 shared allele
A
B
C
D
E
F
G
H
LOD
45 50 55 60 65 70 75
Chr 9 (Mb)

Figure 4.4: Chromosome 9 QTL associated with change in total plasma cholesterol exhibits complex genetic regulation by multiple SNP types. Genome-wide QTL scan for loci affecting change in cholesterol from 6 to 24 weeks of age (A.). Chromosomes 1 through X are represented numerically on the *x*-axis and the *y*-axis represents the LOD score. The relative width of the space allotted for each Chromosome reflects the relative length of each Chromosome. Mice were maintained on a synthetic diet for 2 weeks and then phenotyped for plasma clinical chemistries at 6 weeks of age. Plasma was taken from 24 week old mice after 18 weeks either HFCA or high protein diet. Diet was included as an additive covariate in the mapping model. Colored lines show permutation-derived significance thresholds (N=1000) at $P = 0.05$ (LOD=7.5, shown in red), $P = 0.10$ (LOD=7.1, shown in orange), and $P = 0.63$ (LOD=5.8, shown in yellow). The eight coefficients of the QTL model show the effect of each founder allele on the phenotype along Chromosome 9. Shading identifies the 95% Bayesian credible interval around the peak. (B.). Using the additive SNP model, we identified SNPs associated with the phenotype. Highly associated SNPs are shown in red (LOD> 4). The strains that share haplotypes for associated SNPs are shown as grey bars (A=A/J, B=C57BL/6J, C=129S1/SvImJ, D=NOD/ShiLtJ, E=NZO/HiLtJ, F=CAST/EiJ, G=PWK/PhJ, H=WSB/EiJ). We identified 2 regions within the QTL interval that appear to be influenced by multiple variants with different strain distribution patterns, with CAST/EiJ alleles between 47,929,312- 51,417,095 Mb associated with the phenotype and CAST/EiJ and 129S1/SvImJ shared alleles between 69,353,741- 69,530,721 Mb associated with the phenotype (C.). Results of the genome-wide QTL scan for loci affecting change in total cholesterol from 6 to 24 weeks of age with the genotype of the peak SNP included as a covariate (D.). Allele effects along Chromosome 9 with the CAST-EiJ-specific peak SNP UNC16319105 included as a covariate in the model (E.). Using the additive SNP model, we identified SNPs associated with the phenotype. Highly associated SNPs are shown in red (LOD> 4). The strains that share haplotypes for associated SNPs are shown as grey bars (A=A/J, B=C57BL/6J, C=129S1/SvImJ, D=NOD/ShiLtJ, E=NZO/HiLtJ, F=CAST/EiJ, G=PWK/PhJ, H=WSB/EiJ). The effect of the CAST/EiJ alleles at the proximal peak is diminished; while the CAST/EiJ and 129S1/SvImJ shared alleles at the distal peak are still highly associated with the phenotype (F.).

Table 4.1: QTL associated with BUN, HOMA-IR, and change in total cholesterol ($p < 0.05$).

Phenotype	Chr	LOD	Peak SNP	Peak Position Mb (Bayesian interval)	Positional Candidates
Baseline BUN	1	7.6	UNC2390037	187.5 (187.0-187.9)	6
HOMA-IR after diet	8	7.7	UNC14334951	25.0 (24.3-26.1)	42
Change in TC	9	7.8	UNC16319105	48.3 (45.3-70.3)	429

Table 4.2: Summary of 21 suggestive QTL identified for ALT, AST, BUN, betaine, creatinine, glucose insulin ratio, and HOMA-IR at baseline and after dietary treatment.

Phenotype		Chr	LOD	Peak SNP	Peak Position Mb (Bayesian interval)	Positional Candidates
ALT	Baseline	18	7.1	UNC29404880	61.9 (61.7-70.4)	108
	Baseline	19	6.1	UNC30116610	25.6 (16.5-40.7)	275
	Baseline	7	6	UNC13607268	110.7 (73.3-113.5)	756
	After diet	6	6.8	UNC10897933	37.3 (21.5-136.7)	1584
AST	Baseline	18	6.2	UNC29520496	69.2 (11.9-74.1)	703
	Baseline	14	6	UNC24141257	65.9 (64.6-66.3)	33
	Baseline	1	6	UNC784253	62.0 (58.3-95.3)	519
	Baseline	19	5.9	UNC30308285	40.2 (23.3-40.5)	209
	After diet	4	5.8	UNC7811724	98.5 (94.0-156.2)	1370
BUN	Baseline	16	7.2	UNC27277467	84.1 (77.1-85.9)	57
	After diet	7	5.9	UNC12706670	43.4 (19.5-44.3)	683
Glucose Insulin Ratio	Baseline	11	6.3	UNC20019015	83.0 (82.1-89.5)	208
	After diet	2	6.4	UNC2483107	4.4 (3.2-9.6)	75
	After diet	8	5.8	UNC15550042	114.9 (16.8-115.1)	1303
HOMA-IR	After diet	6	6.6	JAX00145678	115.4 (13.2-117.8)	1228
Betaine	Baseline	6	7	UNC11841141	110.5 (107.4-111.8)	18
	After diet	8	6	UNC14504469	36.1 (23.4-37.7)	146
Creatinine	Baseline	16	6.7	UNC27278724	84.2 (79.2-87.5)	37
	Baseline	12	6.6	UNC21480034	80.6 (79.8-82.5)	58
	After diet	5	7.3	UNC10197506	132.2 (127.0-146.7)	394
	After diet	4	6.2	UNC7752549	93.4 (88.3-95.1)	84

Table 4.3: SNP analysis of the Chromosome 8 QTL associated with HOMA-IR after dietary treatment.

SNP Type	Total SNPs
3' UTR variant	616
5' UTR variant	55
Missense variant	100
Insertion	50
Deletion	73
Splicing variant (includes acceptor, donor, and region)	52
Stop (includes gained, lost, and retained)	1

Table 4.4: Variants within 10 kb of the candidate gene *Fgfr1*.

SNP Type	*Fgfr1* SNPs
3' UTR variant	25
5' UTR variant	10
Missense variant	1
Insertion	2
Deletion	2
Splicing variant (includes acceptor, donor, and region)	5
Stop (includes gained, lost, and retained)	0

CHAPTER 5: SUMMARY, DISCUSSION AND CONCLUSIONS

SUMMARY AND SIGNIFICANCE

This dissertation focuses on improving our current understanding of the genetic processes, and potential therapeutically relevant targets, underlying atherosclerosis, an increasingly prevalent human health issue in today's society. We have taken advantage of state of the art mouse resources and bioinformatics tools to refine known QTL to sub-Mb intervals, identify novel QTL, and even identify single causal variants underlying novel QTL. The Diversity Outbred Mouse population used in this study is a multi-parent outbred mouse resource designed to be highly informative for QTL mapping due to the inclusion of allelic variants from eight different inbred strains of mice and due to the high density of recombination events that accumulate through successive generations of outbreeding [97, 98].

Our analyses provide information about the effects on the phenotype of each of the founder haplotypes at the loci we identified, allowing us to focus attention on particular variants with specific strain distribution patterns. The Ensembl variant consequence predictor was used to inform what effects the particular variants within a region may have on the phenotype. Using this bioinformatics-based approach, we were able to identify QTL related to phenotypic variation of clinically relevant complex traits associated with cardiovascular disease.

The studies presented in this dissertation represent several novel findings that contribute to our understanding or the genetic architecture of atherosclerosis and complex traits related to CVD susceptibility. This is the first study to date to characterize atherosclerosis in the DO mice;

therefore the future use of the DO to study atherosclerotic lesions will be informed by the findings we present here. Using diet to induce lesion formation, we were able to identify *Apobec1* as a candidate gene contributing to variation in lesion size. We also present here the identification of novel QTL regulating the newly discovered atherosclerosis-associated metabolite TMAO and the identification of *Numb* as a high priority positional candidate associated with TMAO levels in the DO mice. Additionally, we report the identification of a single SNP in *Esrrg* that is associated with variation in BUN levels in these mice.

The identification of a single variant in a single mouse study is a level of resolution that had not been previously possible using traditional QTL mapping approaches. Here we have refined larger QTL identified in previous studies to regions containing fewer candidate genes, thereby expediting the identification of biologically relevant genes that are responsible for these associations. Additionally, we have identified novel QTL which will lead to the identification of new genes associated with these phenotypes, leading to a better understanding of the biological mechanisms of these diseases. Taken together, these findings provide a basis upon which to direct future studies to identify the causal genes or variants underlying the QTL not yet resolved at the SNP level. Here, I highlight key findings from each chapter and suggest follow up studies which will build upon this work.

CONCLUSIONS AND DISCUSSION

Identification of *Apobec1* as a Candidate Gene Regulating Atherosclerosis

Atherosclerosis is a chronic inflammatory condition that leads progressively over time to adverse CVD events such as heart attack or stroke [1]. CVD risk is governed by a variety of factors including a combination of both environmental and genetic risk factors. Certain individuals will be susceptible to CVD despite concerned efforts to avoid environmental risk,

while other individuals are less susceptible to CVD despite lifestyle choices that would seem to put them at higher risk for CVDs. These differences in susceptibility to disease are based on underlying genetic differences and we must make efforts to understand the genetics underlying these complex diseases [5].

Mouse studies using inbred strains have taken advantage of the ability to tightly regulate the environment in which the mice are maintained in order to elucidate many genetic factors influencing susceptibility to CVD. Studies in mice have shown that different mouse strains exhibit differential susceptibility to CVD based on underlying genetic differences [40]. It is these genetic differences that can be exploited to better understand the complex mechanisms underlying such diseases and aid in the development of better therapeutic options.

While the use of inbred mouse strains has led to the identification of over 30 atherosclerosis-associated QTL, identification of genes associated with atherosclerosis in inbred crosses is limited by the high levels of homozygosity of inbred animals, such that causal variants not segregating in the study population will not ever be associated with the phenotype. Therefore, researchers have begun to use multi-parent mouse populations in order to introduce heterozygosity and genetic variation, including allelic variants from multiple parental strains, for refined QTL mapping. These features of multi-parent mouse populations, such as the DO, when combined with genotyping using high density arrays, such as the MegaMUGA used in the current study, to detect small recombination blocks allows for fine mapping of QTL intervals associated with complex diseases.

The current study represents the first to quantify atherosclerosis in this newly developed DO mouse population. The DO mice used in this study were received in September of 2012 and were at Generation 11 of outbreeding. The animals were genotyped on the high density

MegaMUGA array, with an average marker spacing of 33 kb, and the DO genomes were reconstructed based on the allele intensities from the MegaMUGA array. Using this approach, we are able to capture the accumulation of recombination events in these animals at a rate that is close the theoretically expected number of events for G11 animals (390 recombinations/G11 animal) [101].

We fed G11 DO mice either a HFCA diet to induce atherosclerosis or a high protein, low fat diet not expected to be atherogenic. Based on the heterogeneous nature of the DO genomes, which are a mosaic of the eight founder strains, we hypothesized that DO mice fed the HFCA atherogenic diet would exhibit variation in susceptibility to diet-induced lesions. We measured lesions throughout the aortic sinuses of these mice. Indeed, we observe that none of the mice fed the high protein, low-fat diet developed lesions (N= 140 mice), while 76% of the mice fed the HFCA diet exhibited aortic lesions (N= 99 of 131 mice). We observed a large amount of variation in susceptibility to lesion formation among the DO animals. The lesions we observed ranged in both size and number, with multiple lesions observed in 65% of the mice (N=85 of 131 mice).

We performed QTL mapping using an additive haplotype model to identify loci associated with variation in atherosclerotic lesion size. We identified one highly significant QTL on Chromosome 6 (LOD= 10.7, $p < 0.05$) associated with variation in lesion size in these mice. This region of Chromosome 6 has been previously identified as associated with atherosclerosis [78]. The previously identified QTL, *Ath37,* has a 12 Mb (116.6- 128.5 Mb) support interval which contains 256 genes, including *Apobec1*. Apobec1 is a cytidine deaminase that was first identified for its role in editing ApoB100 to the shorter isoform ApoB48 [168]. Apobec1 acts as a dimer in complex with ACF to deaminate the cytosine at position 6666 of ApoB mRNA to

uracil [169]. This posttranscriptional modification results in an inframe stop codon at residue 2153 in the mature protein. In humans, tissue-specific expression of *Apobec1* is responsible for the predominance of ApoB48 in the intestine, while ApoB100 is predominantly found in the liver. ApoB100 is the primary apolipoprotein of VLDL, IDL, and LDL and ApoB48 is the primary apolipoprotein of chylomicrons. In mice, *Apobec1* edits ApoB mRNA in both the intestine and the liver resulting in lower levels of ApoB100-containing particles which is thought to contribute to the resistance to atherosclerosis in mice [170].

Despite its apparent role in lipid metabolism, *Apobec1* was not originally implicated as a candidate gene underlying the *Ath37* locus. Three genes, *Adipor2, CD163*, and *A2m*, were described as positional candidates for *Ath37* based on the known biological roles for these genes [78]. *Adipor2* was suspected because it encodes the receptor that binds the ligand adiponectin, which has been shown to be atheroprotective in mice [171, 172]. *CD163* is a macrophage marker that is predictive of CVD in humans [173]. *A2m* is a protease inhibitor that may play a role in reverse cholesterol transport in humans [174]. While these genes may also play a role in atherosclerosis, based on the locations of *Adipor2* (119,353,150- 119,417,704 Mb), *CD163* (124,304,651- 124,330,527 Mb), and *A2m* (121,636,173- 121,679,237 Mb) relative to the QTL we identified here (122,650,900- 122,735,399 Mb), these are not positional candidates for this QTL.

Our atherosclerosis-associated QTL on Chromosome 6 contained only six positional candidate genes: *Apobec1, Gdf3, Gm26168, Dppa3, Nanog*, and *Slc2a3*. *Gm26168* is a predicted gene with no known function; therefore we are not able to predict whether or not this gene may be involved with atherosclerosis without performing additional studies. *Dppa3* and *Nanog* both encode transcription factors that regulate stem cell pluripotency and have not previously been

implicated in atherosclerotic processes [175, 176]. *Slc2a3* encodes the solute carrier family 2 member 3, which functions in glucose transport across cell membranes. While increased expression of *Slc2a3* has been identified in atherosclerotic arteries obtained from human patients by one independent study [177], in mice *Slc2a3* is nominally expressed in heart tissue and *Slc2a1*, *Slc2a4*, and *Slc2a8* are the predominant isoforms in the murine heart [178]. *Gdf3* encodes the growth differentiation factor 3 and was initially a high priority candidate based on the previously characterized role of GDFs in cardiovascular disease [179]. Specifically, elevated serum levels of GDF15 have been associated with increased CVD risk in women [180].

The additive haplotype model used to identify QTL in the DO mice estimates the effects of the founder alleles along each chromosome. The regression coefficients represent the additive genetic effects of each founder haplotype. By plotting the founder allele coefficients, we are able to visualize the effect that alleles contributed by each founder have within the Bayesian defined interval. Within the Chromosome 6 QTL interval, we observed that A/J alleles were specifically associated with larger lesion size compared to alleles from C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HiLtJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ. When we queried the Jackson Strain Expression Survey for differential expression of the genes in this locus, *Apobec1* was the only differentially expressed gene matching the founder allele effects we observed such that A/J specifically expressed higher levels of *Apobec1* mRNA. Therefore, based on the known functions of the positional candidates and the expression patterns we observed in the founder strain mice compared to the allele effects patterns we observed in this study population, we predicted that *Apobec1* is the causal gene contributing to variation in lesion size.

Interestingly, strain-specific differences in expression of *Apobec1* isoforms between the inbred mouse strains A/J and C57BL/6J were recently identified, such that C57BL/6J mice

preferentially express an alternatively spliced truncated *Apobec1* mRNA while A/J mice preferentially express a long *Apobec1* transcript [119]. We exploited these findings to investigate whether there was a difference in the genetic regulation of each of these isoforms across the eight DO founder strains, which includes A/J and C57BL/6J. We identified *cis*-eQTL for both *Apobec1* isoforms and we found that A/J alleles of *Apobec1* were associated with increased expression specifically of the long isoform, but not the short isoform. These results indicate a potential functional relationship between isoform usage and mRNA editing efficiency of *Apobec1* and atherosclerosis.

Identification of *Numb* as a potential genetic regulator of TMAO

Recently, several studies have identified the dietary-derived metabolite TMAO as associated with complex diseases in humans, including atherosclerosis and chronic kidney disease [134, 140, 181]. TMAO is derived from dietary choline and circulating TMAO levels have been identified as predictive of cardiovascular disease risk in humans [70, 132]. Additionally, studies of transgenic mice with an atherosclerosis-sensitizing mutation (C57BL/6J *ApoE*$^{-/-}$) fed diets supplemented with choline or TMAO for 16 weeks showed that significant increases in circulating TMAO were directly correlated with lesion size independent of measures of plasma cholesterol, triglycerides, and glucose [70].

It is known that TMA produced by the breakdown of choline-containing compounds by certain gut bacteria is transported to the liver where it is oxidized by flavin monooxygenase enzymes, primarily by FMO3, resulting in the production of TMAO [141]. The pivotal role of FMO3 in regulating TMAO levels has been well characterized in mice. Specifically, it has been shown that transgenic overexpression of hepatic *Fmo3* results in significantly higher TMAO and lower TMA levels, while antisense-mediated *Fmo3* knockdown results in significantly lower

TMAO and higher TMA levels [72]. Therefore, it would make sense that variants influencing *Fmo3* expression or function would be identified in studies of host genetic regulation of variation in TMAO.

Hartiala et al. did identify a variant in the human *Fmo3* gene that was significantly associated with *Fmo3* mRNA expression; however, this SNP did not demonstrate evidence of association with either plasma TMAO or CVD risk [93]. The same group did not identify the *Fmo3* locus as associated with TMAO levels in a GWAS using HDMP mice. This could be due to factors such as posttranscriptional or posttranslational modifications of FMO3 playing a role in its regulation of TMAO, as this would not be detected at the sequence level. Alternatively, this could be due to a lack of variation at the *Fmo3* locus across the inbred mouse strains used to construct the HDMP. According to the Sanger SNP database, most of the documented variation across the *Fmo3* gene region is contributed by the four wild-derived strains SPRET/EiJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ and one inbred mouse strain DBA/2J. DBA/2J was one of the inbred strains included in the HDMP [82]. Therefore, we were not surprised that we did not detect regulation of TMAO by variants at the *Fmo3* locus in the current study and it appears that the regulation of TMAO by *Fmo3* is complex and may be mediated by posttranscriptional or posttranslational modifications.

In the present study, we identified two novel loci influencing variation in TMAO levels in the DO mice, a QTL on Chromosome 14 (LOD= 8.2, $p < 0.05$) and a QTL on Chromosome 12 (LOD=10.0, $p < 0.05$). The Chromosome 14 QTL is 16.3 Mb in size and it is located in a region of relatively few coding genes. There are 49 positional candidates within the interval. Based on the known biological roles of the positional candidate genes, we did prioritize the Krüppel-like factor 5, *Klf5*, as an attractive candidate for modulating the relationship between TMAO and

atherosclerosis. *Klf5* is a zinc-finger transcription factor that plays a role in cell proliferation in the intestinal epithelium and mediates tissue remodeling in atherosclerosis [149-151]. While biologically *Klf5* represents the most interesting candidate underlying the Chromosome 14 QTL, we are unable to narrow down potential candidates among the 49 positional candidate genes based on our allele effects at this QTL and additional studies will be required to prioritize candidate genes.

The novel QTL on Chromosome 12 has a 4.6 Mb (83.6-88.2 Mb) support interval and contains 116 genes. In this case, CAST/EiJ and PWK/PhJ allelic contribution at the Chromosome 12 locus was specifically associated with lower levels of TMAO, while allelic contribution from A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HiLtJ, and WSB/EiJ within the interval was associated with the higher TMAO levels. While the allele effects within this QTL interval can be used to narrow down the list of candidate genes to those genes that have regulatory or functional variants exhibiting the same strain distribution pattern, we identified 15,386 of the 61,363 genotyped SNPs spanning the QTL interval as shared by and unique to CAST/EiJ and PWK/PhJ. It is possible that a SNP shared by CAST/EiJ and PWK/PhJ results in a functional variant in one of the genes within this locus. There are 78 non-synonymous coding SNPs in 38 genes matching our allele effects that cannot be ruled out as potential candidates.

We also identified 259 3' UTR SNPs, 55 5' UTR SNPs, and 26 splice variants spanning the QTL interval that are shared by CAST/EiJ and PWK/PhJ and it is possible that one of these influences expression of the causal gene. Therefore, we compared hepatic gene expression among the founder strains for the genes in this region using the Jackson Strain Expression Survey. Based on this analysis, we identified *Numb* as the candidate gene that most closely matches the allele effects patterns we observed in the DO mice.

Numb is an interesting candidate because it has recently been shown to be required for cholesterol absorption mediated by NPC1L1 endocytosis in humans and mice and studies in mice have shown that dietary supplementation with TMAO reduces *Npc1L1* mRNA expression at least in the intestines, indicating a potential functional relationship [142, 153].

Numb is expressed in multiple tissues and it may affect processing and trafficking of TMAO independent of intestinal absorption. Indeed, our data focused on *Numb* expression from livers of DO and founder strain mice. Here, we measured *Numb* mRNA expression from liver tissue from the DO mice and mapped expression of this gene to a highly significant *cis*-eQTL on Chromosome 12. The colocalization of QTLs for plasma TMAO and hepatic *Numb* expression suggests that differences in *Numb* expression could be contributing to the variation in TMAO observed in the DO mice. This evidence supports a possible role for hepatic *Numb* in the regulation of TMAO levels.

Identification of a single variant in *Esrrg* associated with variation in blood urea nitrogen

In Chapter 2, we described QTL mapping of loci associated with atherosclerotic lesion development and cardiovascular risk factors in the DO mice. Because atherosclerosis does not occur spontaneously in mice, we did not expect DO mice to be susceptible to atherosclerosis. Therefore, we fed the mice a synthetic HFCA diet designed to induce lesion formation but reduce adverse pathological effects on liver and kidney function [41].

In order to assess the impact of this diet on liver and kidney function, we measured plasma levels of ALT, AST, and BUN. While we were primarily interested in cardiovascular phenotypes, we took advantage of these data to identify QTL associated with ALT, AST, and BUN. While we did not identify any significant loci associated with ALT and AST, we identified a highly significant QTL on Chromosome 1 (LOD= 7.6, $p < 0.05$) associated with BUN in 6

week old DO mice. BUN is a measure of the amount of nitrogen in the blood that comes from urea which is made when proteins are broken down in the body. In the presence of kidney dysfunction, the kidneys may not filter urea efficiently and this results in elevated BUN levels [182]. Based on the MGI Genes and Markers Query Form (http://www.informatics.jax.org/) there are no QTL in this genomic interval previously reported as associated with BUN, therefore, this represents a novel QTL.

Interestingly, the allele effects at this locus indicate that a single variant contributes to variation in BUN levels. Based on the allele effects, allelic contribution from the founder strains 129S1/SvImJ, NZO/HiLtJ, CAST/EiJ, PWK/PhJ/ and WSB/EiJ was associated with higher BUN levels, while allelic contribution from the A/J, C57BL/6J, and NOD/ShiLtJ founder strains was associated with lower BUN levels. We identified only one SNP, *rs32769253*, matching this allele effects pattern, such that the strains A/J, C57BL/6J, and NOD/ShiLtJ share a T at this position, while the strains 129S1/SvImJ, NZO/HiLtJ, CAST/EiJ, PWK/PhJ/ and WSB/EiJ exhibit a shared T->C variant.

The variant consequence is identified within the Sanger database as intronic, based on the Ensembl Variant Effect Predictor [183]. However, it should be noted that variants may have different effects in different transcripts. *Esrrg* has four known protein-coding transcripts. The SNP we have identified here is located in an intron of *Esrrg*, between exons 2 and 3 of the isoforms *Esrrg-003* and *Esrrg-004*. However, the SNP is upstream of the TSS of the *Esrrg-001* and *Esrrg-002* isoforms, indicating it may affect expression of these isoforms differentially.

Recently, *Esrrg* has been shown to play a role in kidney development. *Esrrg* encodes the estrogen-receptor related gamma protein, which is a nuclear receptor that contains DNA-binding domains and is a transcriptional activator. *Esrrg* is strongly expressed in the developing kidneys

and *Esrrg*$^{-/-}$ knockout mice exhibit deformed renal papillae and a reduction in the total number of nephrons [160, 184]. Biologically, the association of *Esrrg* with plasma BUN is intriguing, as renal papillae are composed of the collecting ducts of the nephrons where BUN is reabsorbed [185].

Based on the location and predicted effect of the single variant at the Chromosome 1 QTL that matches our allele effects pattern and the known biological role of *Esrrg* in kidney development, we predict that this SNP influences *Esrrg* isoform usage and is the causal factor contributing to variation in BUN at this locus. The intronic SNP we have identified here may affect enhancer binding and expression of *Esrrg* isoforms, although we cannot determine the functional relevance of this SNP without follow up studies. Interestingly, intronic SNPs in *Esrrg* have been associated with human disease previously, indicating that isoform-specific expression of *Esrrg* plays a large role in mediating the functions of this gene [186, 187]. In order to understand the role of the single SNP that we identified here, we must perform additional follow up studies.

FUTURE DIRECTIONS

The data presented in this dissertation highlight the impact that the current era of mouse genomics has on advancing our understanding of complex human diseases. Historically, QTL mapping studies have sought to identify the causal genes and variants contributing to variation in quantitative traits related to disease states in humans. While traditional QTL experiments have successfully identified loci associated with complex traits, identification of causal genes and variants underlying these associations was often hampered by the size of the QTL intervals, requiring years of follow up study to determine functional associations.

The DO mice are a newly developed multi-parent mouse population that was specifically designed to be highly informative for QTL mapping. The DO mouse resource was first introduced in 2009 and the earliest studies of mapping quantitative traits in the DO were published in 2012. Since then, six additional studies have published results of QTL mapping in the DO mice, including the study presented in Chapter 2 of this dissertation, as summarized in **Table 1.2**.

While highly informative, QTL mapping in multi-parent mouse populations such as the DO mice presents certain analytical challenges. Because the genome of each animal is a unique mosaic of the eight inbred founder strains, mice must be genotyped using high-density genotyping arrays and the haplotypes must be reconstructed to determine the contribution of founder alleles along the chromosomes. Genotyping mice using higher density arrays as well as continued improvements to haplotype reconstruction algorithms will likely continue to improve the mapping resolution made possible by using the DO mice. Additionally, the accumulation of meiotic recombination events will result in smaller and smaller haplotype blocks at each successive outbreeding generation of DO mice. Therefore the future of QTL studies in mice are likely to rely heavily on multi-parent outbred populations such as the DO and on the availability of analytical tools to properly perform the analysis.

Using the DO mice in the current study, we have shown that it is possible to directly identify single variants associated with a phenotype, as exemplified by the identification of a single intronic SNP in *Esrrg* that is associated with variation in BUN levels. This level of resolution is something that was not previously possible in single experiments using inbred mouse strains. This is significant in that it represents a unique opportunity to expedite the discovery of causal genes and variants associated with complex traits. While we have shown

here that fine mapping of refined QTL intervals is possible using a conservative population of ~300 DO animals, follow up studies must still be performed to validate the findings and definitely identify the biological relevance of the genes and variants identified by QTL mapping. Below, I will discuss the proposed experiments that should be conducted in order to follow up on and validate the major findings from the current study.

Identification and functional significance of the A/J allele of *Apobec1* in development of atherosclerotic lesions

In the current study, we identified a single significant QTL on Chromosome 6 (LOD= 10.7, $p < 0.05$) associated with atherosclerosis. Interestingly, this region of Chromosome 6 is within the previously identified 12 Mb *Ath37* locus; however, we have refined this region to ~100 kb containing only 6 candidate genes. Based on the known functions of these genes, we identified *Apobec1* as a high priority candidate for regulating lesion size and we identified a highly significant *cis*-eQTL for *Apobec1*, suggesting that differential expression may contribute to differences in lesion size.

We found that A/J alleles at the Chromosome 6 locus were associated with larger lesion size, suggesting that an A/J-specific variant may contribute to larger and more extensive atherosclerotic lesions in these mice. Interestingly, a recent study found A/J mice preferentially express a specific *Apobec1* isoform that alters the editing efficiency of *Apobec1*, such that inbred A/J mice exhibit higher rates of editing compared to C57BL/6J mice which preferentially express a truncated isoform of *Apobec1* [119]. In light of the results of the present study, this indicates a potential functional relationship between isoform usage and mRNA editing efficiency of *Apobec1* and atherosclerosis.

The previous study that identified differences in *Apobec1* isoform usage between A/J and C57Bl/6J animals was performed using bone marrow-derived macrophages from the mice. Similarly, we could look for differences in *Apobec1* isoform levels and differential editing of *Apobec1* targets by isolating macrophages from animals that have A/J allelic contribution at the Chromosome 6 locus compared to those that have contribution from the other founders. While we did not collect these cells from the DO mice used in the current study, the Collaborative Cross mice represent a unique and valuable resource for validating our findings. The DO and CC mice are complementary resources and while each DO mouse has a unique genome, the CC mice are maintained as reproducible RILs. Therefore, we can use the CC mice to identify differences in *Apobec1* isoform usage and editing efficiency by obtaining CC lines that have contribution from either A/J or one of the other founder strains at the Chromosome 6 locus we have identified using the DO mice.

There are currently 48 CC RILs available, including 12 Tier 1 lines and 36 Tier 2 lines. Tier 1 CC lines have contribution from all 8 founders, while Tier 2 lines have allelic contribution from 6 to 8 of the founders used to create the DO mice. Currently, both Tier 1 and Tier 2 CC lines are at least 90% homozygous. Of the 48 available CC lines, there are 8 that are homozygous for allelic contribution from the A/J founder strain at *Apobec1*, including CC008, CC011, CC026, CC030, CC040, CC043, CC045, and CC063. There are 3 CC lines that are heterozygous for allelic contribution from the A/J founder at *Apobec1*, including CC017, CC058, and CC075. Of the remaining 37 CC lines, 16 are homozygous C57BL/6J at *Apobec1*. Based on the results of the current study, we would predict that those CC lines with contribution from A/J at *Apobec1* would exhibit larger lesions in response to 18 weeks of the atherogenic diet. Based on the previously identified differences in isoform usage, we would predict that these lines also

differ in isoform usage and editing efficiency based on allelic contribution at *Apobec1*. We plan to test these hypotheses by measuring lesion size in response to the HFCA diet in CC mice and by isolating macrophages and measuring *Apobec1* expression and editing efficiency in these mice.

In addition to its known role in editing APOB, recent studies have identified at least 36 additional mRNA targets of APOBEC1 C to U editing [122, 188]. In the present study, we found that APOB levels differed significantly in mice with different genotypes of the peak SNP at the Chromosome 6 QTL, **Figure 2.12**, suggesting that the A/J allele of *Apobec1* may contribute to regulation of atherosclerosis via editing of APOB. However, it is possible that the A/J allele of *Apobec1* influences lesion size through editing of targets other than APOB and this remains to be tested. Initial studies to pursue the biological mechanisms underlying the association between the A/J allele of *Apobec1* and increased lesion size will be aimed at identifying whether the isoform-specific variation influences APOB editing specifically. Western immunoblotting of hepatic APOB will be performed in order to quantify the relative amounts of APOB100 and APOB48. If the A/J isoform of *Apobec1* is better at editing its primary target APOB, then we would expect to see an increase in APOB48 protein from livers of A/J mice compared to C57BL/6J mice and to the other 6 founder strains.

If the results of these proposed studies suggest that A/J mice indeed harbor a causal variant in *Apobec1* that increases atherosclerosis susceptibility, we would next identify the causal variant by sequencing *Apobec1* in the DO and CC mice that exhibit differing founder contribution at the Chromosome 6 locus in order to identify sequence differences specific to A/J. If we are able to identify a causal variant, then we will pursue *in vitro* studies to determine the functional role of the causal variant in *Apobec1*. There are cell lines that actively synthesize

cholesterol and make serum proteins (for example FL83B) that could be used to determine the effect of a causal variant of *Apobec1* on either editing of APOB or other target genes *in vitro*.

Based on the results of these studies, we could predict the effects of overexpressing or knocking out the A/J allele on *Apobec1* function. We could then test this by overexpressing the A/J variant in either C57BL/6J mice or *Apobec1*$^{-/-}$ mice. The results of these studies would definitely identify the functional role of the A/J isoform of *Apobec1* and would provide a basis for understanding the biological mechanism by which this variant may contribute to atherosclerosis susceptibility.

Further characterization of the Chromosome 12 QTL associated with TMAO and validation of *Numb* as a functional candidate regulating TMAO levels

In the present study, we performed QTL mapping using an additive haplotype model to identify loci associated with variation in TMAO levels. We identified two novel loci associated with this metabolite in the DO mice. The QTL interval on Chromosome 14 contains 49 positional candidate genes, but the allele effects in this region appear complex and do not fall out into two distinct categories. These results indicate that there may be more than one causal variant influencing TMAO levels at this locus. While no follow up studies have been pursued to identify the causal gene underlying this QTL, we do plan to perform microarray analysis of gene expression from liver tissue from the DO mice used in the current study. We predict that if this association is the result of differences in expression of the causal gene, then this should be correlated with TMAO levels in this population of mice. Performing QTL mapping using the additive haplotype model for expression of the candidate genes to identify the presence of *cis*-eQTL would also provide evidence as to which positional candidates should be pursued in follow up studies.

The Chromosome 12 QTL interval contains a total of 116 genes and the allele effects indicate that CAST/EiJ and PWK/PhJ contribution at the locus are specifically associated with lower TMAO levels, while contribution at the locus from the 6 other founder strains is associated with higher TMAO levels. Based on this pattern of allele effects, we queried the publicly available Jackson Laboratory Strain Survey for hepatic expression of candidate genes matching this allele effects pattern. We identified *Numb* as the candidate gene that most closely matches the allele effects patterns we observed in the DO mice. We measured *Numb* mRNA levels by QPCR and identified a highly significant *cis*-eQTL regulating *Numb* expression, indicating that this gene represents a high priority candidate underlying the Chromosome 12 QTL.

Interestingly, the allele effects for the *Numb cis*-eQTL mirrored those of the Chromosome 12 QTL associated with TMAO levels, such that CAST/EiJ and PWK/PhJ allelic contribution at the locus was associated with higher *Numb* expression and lower TMAO levels. Consistent with these results, *Numb* mRNA and TMAO levels were highly negatively correlated in the DO mice at both 6 weeks (r= -0.65) and 24 weeks (r= -0.40).

Functionally, *Numb* is an interesting gene candidate for regulation of TMAO levels because it plays a known role NPC1L1-mediated intestinal cholesterol absorption in both humans and mice. *Numb* is known to be expressed in multiple tissues and it may affect processing and trafficking of TMAO independent of intestinal absorption. Evidence has been published since the studies presented here were conducted that show that *Numb* facilitates NPC1L1-mediated endocytosis in both the mouse and human intestine and liver [145, 189]. A functional relationship has been suggested in mouse studies in which dietary supplementation with TMAO was shown to reduce *Npc1L1* mRNA expression in the intestines, however no

studies to date have identified a direction functional role for NPC1L1 or NUMB in regulating absorption of TMA or TMAO in the liver or intestine.

Based on our results, we observed that *Numb* expression is regulated by a colocalizing *cis*-eQTL at the Chromosome 12 locus and we identified a negative correlation between *Numb* expression and TMAO levels in the DO mice. Based on these results, we hypothesize that *Numb* may play a functional role in absorption of TMA and/or TMAO. If *Numb* expression does contribute to variation in TMAO levels as we have observed in the DO mice, we can validate this association using the CC lines. We plan to measure *Numb* expression and TMAO levels in 6 week old CC mice with founder contribution at the Chromosome 12 locus from either CAST/EiJ, PWK/PhJ, or the other 6 founder strains. Of the 48 available CC lines, there are 8 lines available that are homozygous for PWK/PhJ allelic contribution at the Chromosome 12 QTL and 4 that are homozygous for CAST/EiJ allelic contribution at the Chromosome 12 QTL; while the other 36 available CC lines have founder contribution from the other 6 founder strains at the locus. We hypothesize that CC lines with CAST/EiJ and PWK/PhJ allelic contribution at the Chromosome 12 locus will exhibit lower TMAO levels and should also express higher levels of *Numb*.

Once we have identified CC lines that exhibit high levels of TMAO and express low levels of *Numb*, we can directly test the effects of *Numb* expression on TMAO levels in these mice. We plan to transiently overexpress *Numb* in the livers of these mice by tail vein injection of an adenovirus construct (ADV-266207, Vector Biolabs) and determine the effects of overexpression of *Numb* in the liver on TMAO levels in these mice. The liver is the site of oxidation of TMA to TMAO and it makes sense that an endocytic protein may play a functional role in trafficking TMA to the site of oxidation and that variants in *Numb* may influence the efficiency of this process.

Another approach that will be pursued in order to better understand the potential functional role that *Numb* may play in regulating TMAO levels is a bioinformatics based approach aimed at identifying gene networks that are associated with the phenotype and identifying biologically relevant associated genes within significant networks. This approach, known as weighted gene co-expression network analysis or WGCNA, has been successfully used to identify novel pathways contributing to complex phenotypes in previous studies [84, 190]. Studies are currently underway to measure gene expression and differential transcript abundance of genes across the eight founder strain mice used to create the DO by RNA-Sequencing in our lab. This expression data will then be used to identify networks, or groups of genes, that are highly correlated with TMAO levels in the founder strain mice. Genes identified by WGCNA can then be further investigated using *in vitro* approaches and the effects can be evaluated to determine the potential therapeutic applications of such novel genetic associations.

Based on the identification of *Numb* as a high priority candidate gene, we expect *Numb* to be contained in a module that is highly associated with TMAO levels. The identification of other genes in the same highly correlated network as *Numb* will inform us as to the potential biological pathway by which *Numb* may act to influence regulation of TMAO. For example, if we identify NPC1L1 in the same module, this further supports a role for NUMB-dependent NPC1L1-mediated absorption of TMA and TMAO. Alternatively, we may identify a novel pathway by which *Numb* expression is regulated and associated with TMAO. The results of these studies should provide direction for experimental validation of candidate genes associated with TMAO levels.

In addition to the identification of genetic regulators of TMAO, we must also consider the evidence that TMAO is regulated by modulation of the gut microbiome. It is known that

TMA produced by the breakdown of choline-containing compounds by certain gut bacteria is transported to the liver where it is oxidized by primarily by FMO3, resulting in the production of TMAO [141]. A role for the gut microbiome in regulating TMAO is therefore apparent and studies in humans have shown that antibiotic suppression of the gut flora prevents spikes in circulating TMAO levels after dietary choline challenge [142]. Most recently, a study by Romano et al. identified specific bacterial strains that are able to produce TMA from choline *in vitro* [191]. The identification of genes regulating both complex traits and gut microbial composition has been demonstrated previously [192]. Significantly, if we are able to identify genes as associated with TMAO that are regulating taxa or individual microbial species capable of producing TMA from choline, these results may provide novel therapeutic interventions for the treatment of multiple diseases that have been associated with elevated TMAO, including atherosclerosis and chronic kidney disease.

To this end, the cohort of DO mice used in the current study is part of a meta-analysis that will examine the microbiome of this study population. Specifically, fecal samples were collected from the DO mice used in the current study at three time points and these will be used to conduct a full metagenomics analysis. I plan to perform QTL mapping of microbial abundance measured from these samples in order to identify genes that regulate the gut microbiome. It will be interesting to see if abundance of the intestinal microbial species identified as capable of producing TMA from choline maps to the same loci on Chromosome 12 or Chromosome 14 that we identified here as associated with TMAO levels, as this would suggest that the causal genes underlying these loci may regulate TMAO levels via modulation of the gut microbiome.

Identification of the role of the *Esrrg* variant *rs32769253* in the regulation of BUN and kidney function

In the present study, we fed mice either a HFCA diet designed to induce atherosclerotic lesions or a low fat, high protein control diet that was not expected to be atherogenic. In order to assess the impact of this diet on kidney function, we measured plasma levels of BUN which when elevated can indicate that the kidneys may not be functioning properly and may not be filtering urea efficiently. We mapped variation in this phenotype using the additive haplotype model and identified a highly significant QTL on Chromosome 1 associated with BUN in 6 week old DO mice. Interestingly, only one SNP, *rs32769253*, matched the allele effects patterns that we identified at this locus. According to the Sanger SNP database, this SNP represents a C/T polymorphism located in an intron of the *Esrrg* gene at 187,919,386 Mb (GRC Build 38). The strains A/J, C57BL/6J, and NOD/ShiLtJ share a T at this position, while the strains 129S1/SvImJ, NZO/HiLtJ, CAST/EiJ, PWK/PhJ/ and WSB/EiJ exhibit a shared T->C variant.

This SNP we have identified in the current study as associated with variation in BUN levels is identified as intronic based on its location in an intron of *Esrrg*, between exons 2 and 3 of the isoforms *Esrrg-003* and *Esrrg-004*. However, this SNP is located upstream of the TSS of two other known isoforms of *Esrrg*, *Esrrg-001* and *Esrrg-002*, indicating it may affect expression of these isoforms differentially. Therefore, we hypothesize that this variant may result in differential expression of certain isoforms of *Esrrg*. By designing primers specific for each isoform, we can measure expression of *Esrrg* isoforms in the DO mice. If this variant affects expression of one or both isoforms, we would expect to see a *cis*-eQTL at the locus with allele effects matching the strain distribution pattern exhibited by this SNP.

Additionally, we should determine whether the expression of *Esrrg* varies across the founder strains and if variation in *Esrrg* expression is directly or inversely associated with BUN levels. We hypothesis that *Esrrg* expression of one or more isoforms will be higher or lower in A/J, C57BL/6J, and NOD/ShiLtJ compared to 129S1/SvImJ, NZO/HiLtJ, CAST/EiJ, PWK/PhJ and WSB/EiJ. We should perform QPCR and Western immunoblotting to quantify mRNA and protein expression of *Esrrg* across the founder strains. These studies will be important to inform the directionality of the effect of *Esrrg* expression on variation in BUN levels. We could then confirm a direct role for this SNP in regulating the expression of *Esrrg*, by employing gene-targeting technologies and site-directed mutagenesis [193, 194]. Specifically, by introducing a T → C in strains A/J, C57BL/6J, and NOD/ShiLtJ or a C → T in strains 129S1/SvImJ, NZO/HiLtJ, CAST/EiJ, PWK/PhJ/ and WSB/EiJ at this position, we can directly determine the effect of this SNP on *Esrrg* expression and BUN levels. If this variant is causal, we should see the low BUN expresser strains exhibit increased BUN expression and the high BUN expresser strains exhibit decrease BUN expression.

Interestingly, *Esrrg* has been shown to regulate the morphology of renal papilla and it is therefore biologically plausible that a variant in *Esrrg* that affects kidney development might also influence kidney function [160]. In the current study, we were primarily interested in cardiovascular phenotypes and we therefore did not collect data on renal function in these mice.

As a follow up to the findings here, studies are currently underway to measure urinary sodium excretion volume as a measure of renal function in an independent cohort of DO mice. The mice will be injected with sterile saline and then placed in metabolic cages for collection of urine samples and recording of sample volume at 2, 4, and 8 hour time points. The mice will be fasted for 4 hours and then plasma samples collected and used to measure BUN levels followed

by harvesting of kidney tissue to measure renal *Esrrg* expression. In this way, we can relate renal function, BUN levels, and expression of the candidate gene *Esrrg* in these mice. We predict that DO mice with reduced renal function will have elevated BUN and that these mice should exhibit allelic contribution at the Chromosome 1 locus from the founder strains 129S1/SvImJ, NZO/HiLtJ, CAST/EiJ, PWK/PhJ/ and WSB/EiJ, based on the allele effects we observed in the current DO study.

REFERENCES

1. Pagidipati, N.J. and T.A. Gaziano, *Estimating deaths from cardiovascular disease: a review of global methodologies of mortality measurement.* Circulation, 2013. **127**(6): p. 749-56.

2. Minino, A.M., et al., *Deaths: final data for 2008.* Natl Vital Stat Rep, 2011. **59**(10): p. 1-126.

3. Minino, A.M., *Death in the United States, 2011.* NCHS Data Brief, 2013(115): p. 1-8.

4. Turk-Adawi, K.I. and S.L. Grace, *Narrative Review Comparing the Benefits of and Participation in Cardiac Rehabilitation in High-, Middle- and Low-Income Countries.* Heart Lung Circ, 2014.

5. Breslow, J.L., et al., *Genetic susceptibility to atherosclerosis.* Circulation, 1989. **80**(3): p. 724-8.

6. Strong, J.P., et al., *Prevalence and extent of atherosclerosis in adolescents and young adults: implications for prevention from the Pathobiological Determinants of Atherosclerosis in Youth Study.* Jama, 1999. **281**(8): p. 727-35.

7. Berenson, G.S., S.R. Srinivasan, and T.A. Nicklas, *Atherosclerosis: a nutritional disease of childhood.* Am J Cardiol, 1998. **82**(10b): p. 22t-29t.

8. Krishnan, P., et al., *Cardiovascular risk profile of asymptomatic healthy young adults with increased carotid artery intima-media thickness: the Bogalusa Heart Study.* J La State Med Soc, 2003. **155**(3): p. 165-9.

9. Tracy, R.E., et al., *Histologic features of atherosclerosis and hypertension from autopsies of young individuals in a defined geographic population: the Bogalusa Heart Study.* Atherosclerosis, 1995. **116**(2): p. 163-79.

10. Harker, L.A., S.M. Schwartz, and R. Ross, *Endothelium and arteriosclerosis.* Clin Haematol, 1981. **10**(2): p. 283-96.

11. Schwartz, S.M., *Role of endothelial integrity in atherosclerosis.* Artery, 1980. **8**(4): p. 305-14.

12. Schwartz, C.J., et al., *Monocyte-macrophage participation in atherogenesis: inflammatory components of pathogenesis.* Semin Thromb Hemost, 1986. **12**(2): p. 79-86.

13. Lusis, A.J., *Atherosclerosis.* Nature, 2000. **407**(6801): p. 233-41.

14. Jensen, J. and D.H. Blankenhorn, *The inheritance of familial hypercholesterolemia.* Am J Med, 1972. **52**(4): p. 499-516.

15. Buja, L.M., P.T. Kovanen, and D.W. Bilheimer, *Cellular pathology of homozygous familial hypercholesterolemia.* Am J Pathol, 1979. **97**(2): p. 327-57.

16. Goldstein, J.L., et al., *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, 1973. **52**(7): p. 1544-68.

17. Yang, W.S., et al., *A mutation in the promoter of the lipoprotein lipase (LPL) gene in a patient with familial combined hyperlipidemia and low LPL activity.* Proc Natl Acad Sci U S A, 1995. **92**(10): p. 4462-6.

18. Masucci-Magoulas, L., et al., *A mouse model with features of familial combined hyperlipidemia.* Science, 1997. **275**(5298): p. 391-4.

19. Grundy, S.M., et al., *Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III Guidelines.* J Am Coll Cardiol, 2004. **44**(3): p. 720-32.

20. Gotto, A.M., Jr. and J.E. Moon, *Management of cardiovascular risk: the importance of meeting lipid targets.* Am J Cardiol, 2012. **110**(1 Suppl): p. 3a-14a.

21. Butcher, M. and E. Galkina, *Current views on the functions of interleukin-17A-producing cells in atherosclerosis.* Thrombosis and Haemostasis, 2011. **106**(5): p. 787-795.

22. Butcher, M.J., et al., *The IL-17A/IL-17RA Axis Plays a Proatherogenic Role via the Regulation of Aortic Myeloid Cell Recruitment.* Circulation Research, 2012. **110**(5): p. 675-687.

23. Smith, E., et al., *Blockade of Interleukin-17A Results in Reduced Atherosclerosis in Apolipoprotein E-Deficient Mice.* Circulation, 2010. **121**(15): p. 1746-1755.

24. Wong, B.W., et al., *The biological role of inflammation in atherosclerosis.* Can J Cardiol, 2012. **28**(6): p. 631-41.

25. Nilsson, J., M. Wigren, and P.K. Shah, *Vaccines against atherosclerosis.* Expert Rev Vaccines, 2013. **12**(3): p. 311-21.

26. Lusis, A.J., *Genetics of atherosclerosis.* Trends Genet, 2012. **28**(6): p. 267-75.

27. Collins, F.S., M. Morgan, and A. Patrinos, *The Human Genome Project: lessons from large-scale biology.* Science, 2003. **300**(5617): p. 286-90.

28. Makela, K.M., et al., *Genome-wide association study pinpoints a new functional apolipoprotein B variant influencing oxidized low-density lipoprotein levels but not cardiovascular events: AtheroRemo Consortium.* Circ Cardiovasc Genet, 2013. **6**(1): p. 73-81.

29. Radovica, I., et al., *Association between CETP, MLXIPL, and TOMM40 polymorphisms and serum lipid levels in a Latvian population.* Meta Gene, 2014. **2**: p. 565-78.

30. Lu, Y., et al., *Exploring genetic determinants of plasma total cholesterol levels and their predictive value in a longitudinal study.* Atherosclerosis, 2010. **213**(1): p. 200-5.

31. Yang, R., et al., *A genome-wide linkage scan identifies multiple quantitative trait loci for HDL-cholesterol levels in families with premature CAD and MI.* J Lipid Res, 2010. **51**(6): p. 1442-51.

32. Flint, J. and R. Mott, *Finding the molecular basis of quantitative traits: successes and pitfalls.* Nat Rev Genet, 2001. **2**(6): p. 437-45.

33. Morrisett, J.D., et al., *Genetic susceptibility and resistance to diet-induced atherosclerosis and hyperlipoproteinemia.* Arteriosclerosis, 1982. **2**(4): p. 312-24.

34. Roberts, A. and J.S. Thompson, *Inbred mice and their hypbrids as an animal model for atherosclerosis research.* Adv Exp Med Biol, 1976. **67**(00): p. 313-327.

35. LeBoeuf, R.C., et al., *Genetic control of lipid transport in mice. I. Structural properties and polymorphisms of plasma lipoproteins.* J Biol Chem, 1983. **258**(8): p. 5063-70.

36. Lusis, A.J., et al., *Genetic control of lipid transport in mice. II. Genes controlling structure of high density lipoproteins.* J Biol Chem, 1983. **258**(8): p. 5071-8.

37. Bruell, J.H., A.F. Daroczy, and H.K. Hellerstein, *Strain and sex differences in serum cholesterol levels of mice.* Science, 1962. **135**(3508): p. 1071-2.

38. Weibust, R.S., *Inheritance of plasma cholesterol levels in mice.* Genetics, 1973. **73**(2): p. 303-12.

39. Paigen, B., M.B. Havens, and A. Morrow, *Effect of 3-methylcholanthrene on the development of aortic lesions in mice.* Cancer Res, 1985. **45**(8): p. 3850-5.

40. Paigen, B., et al., *Variation in susceptibility to atherosclerosis among inbred strains of mice.* Atherosclerosis, 1985. **57**(1): p. 65-73.

41. Nishina, P.M., J. Verstuyft, and B. Paigen, *Synthetic low and high fat diets for the study of atherosclerosis in the mouse.* J Lipid Res, 1990. **31**(5): p. 859-69.

42. Piedrahita, J.A., et al., *Generation of mice carrying a mutant apolipoprotein E gene inactivated by gene targeting in embryonic stem cells.* Proc Natl Acad Sci U S A, 1992. **89**(10): p. 4471-5.

43. Weisgraber, K.H., et al., *Apolipoprotein E2(Arg158----Cys) frequency in a hyperlipidemic French-Canadian population of apolipoprotein E2/2 subjects. Determination by synthetic oligonucleotide probes.* Arteriosclerosis, 1989. **9**(1): p. 50-7.

44. Eto, M., K. Watanabe, and I. Makino, *Increased frequencies of apolipoprotein epsilon 2 and epsilon 4 alleles in patients with ischemic heart disease.* Clin Genet, 1989. **36**(3): p. 183-8.

45. Nakashima, Y., et al., *ApoE-deficient mice develop lesions of all phases of atherosclerosis throughout the arterial tree.* Arterioscler Thromb, 1994. **14**(1): p. 133-40.

46. Reddick, R.L., S.H. Zhang, and N. Maeda, *Atherosclerosis in mice lacking apo E. Evaluation of lesional development and progression.* Arterioscler Thromb, 1994. **14**(1): p. 141-7.

47. Zhang, S.H., et al., *Spontaneous hypercholesterolemia and arterial lesions in mice lacking apolipoprotein E.* Science, 1992. **258**(5081): p. 468-71.

48. Zhang, S.H., et al., *Diet-induced atherosclerosis in mice heterozygous and homozygous for apolipoprotein E gene disruption.* J Clin Invest, 1994. **94**(3): p. 937-45.

49. Ishibashi, S., et al., *Hypercholesterolemia in low density lipoprotein receptor knockout mice and its reversal by adenovirus-mediated gene delivery.* J Clin Invest, 1993. **92**(2): p. 883-93.

50. Benlian, P., et al., *A LDL receptor gene homozygous mutation: PCR amplification, direct genomic sequencing, associated haplotype, rapid screening for frequency.* Ann Genet, 1990. **33**(2): p. 65-9.

51. Schuster, H., et al., *Familial defective apolipoprotein B-100. Comparison with familial hypercholesterolemia in 18 cases detected in Munich.* Arteriosclerosis, 1990. **10**(4): p. 577-81.

52. Santos, R.D. and R.C. Maranhao, *What is new in familial hypercholesterolemia?* Curr Opin Lipidol, 2014. **25**(3): p. 183-8.

53. Ishibashi, S., et al., *Massive xanthomatosis and atherosclerosis in cholesterol-fed low density lipoprotein receptor-negative mice.* J Clin Invest, 1994. **93**(5): p. 1885-93.

54. Ishibashi, S., et al., *The two-receptor model of lipoprotein clearance: tests of the hypothesis in "knockout" mice lacking the low density lipoprotein receptor, apolipoprotein E, or both proteins.* Proc Natl Acad Sci U S A, 1994. **91**(10): p. 4431-5.

55. Powell-Braxton, L., et al., *A mouse model of human familial hypercholesterolemia: markedly elevated low density lipoprotein cholesterol levels and severe atherosclerosis on a low-fat chow diet.* Nat Med, 1998. **4**(8): p. 934-8.

56. Purcell-Huynh, D.A., et al., *Transgenic mice expressing high levels of human apolipoprotein B develop severe atherosclerotic lesions in response to a high-fat diet.* J Clin Invest, 1995. **95**(5): p. 2246-57.

57. Sanan, D.A., et al., *Low density lipoprotein receptor-negative mice expressing human apolipoprotein B-100 develop complex atherosclerotic lesions on a chow diet: no accentuation by apolipoprotein(a).* Proc Natl Acad Sci U S A, 1998. **95**(8): p. 4544-9.

58. Rao, D.C., *Genetic dissection of complex traits: an overview.* Adv Genet, 2001. **42**: p. 13-34.

59. Belmont, J.W. and S.M. Leal, *Complex phenotypes and complex genetics: an introduction to genetic studies of complex traits.* Curr Atheroscler Rep, 2005. **7**(3): p. 180-7.

60. Lander, E.S. and N.J. Schork, *Genetic dissection of complex traits.* Science, 1994. **265**(5181): p. 2037-48.

61. Smith, J., *Quantitative trait locus mapping for atherosclerosis susceptibility.* Curr Opin Lipidol, 2003. **14**(5): p. 499-504.

62. Paigen, B., et al., *Ath-1, a gene determining atherosclerosis susceptibility and high density lipoprotein levels in mice.* Proc Natl Acad Sci U S A, 1987. **84**(11): p. 3763-7.

63. Pollard, D.A., *Design and construction of recombinant inbred lines.* Methods Mol Biol, 2012. **871**: p. 31-9.

64. Paigen, B., et al., *Ath-2, a second gene determining atherosclerosis susceptibility and high density lipoprotein levels in mice.* Genetics, 1989. **122**(1): p. 163-8.

65. Paigen, B., et al., *Genetic analysis of strains C57BL/6J and BALB/cJ for Ath-1, a gene determining atherosclerosis susceptibility in mice.* Biochem Genet, 1987. **25**(11-12): p. 881-92.

66. Stewart-Phillips, J.L., J. Lough, and E. Skamene, *ATH-3, a new gene for atherosclerosis in the mouse.* Clin Invest Med, 1989. **12**(2): p. 121-6.

67. Mehrabian, M., et al., *Genetic Locus in Mice That Blocks Development of Atherosclerosis Despite Extreme Hyperlipidemia.* Circulation Research, 2001. **89**(2): p. 125-130.

68. Chen, Y., et al., *Genetic and Genomic Insights into the Molecular Basis of Atherosclerosis.* Cell Metabolism, 2007. **6**(3): p. 164-179.

69. Wang, X., et al., *Identifying novel genes for atherosclerosis through mouse-human comparative genetics.* Am J Hum Genet, 2005. **77**(1): p. 1-15.

70. Wang, Z., et al., *Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease.* Nature, 2011. **472**(7341): p. 57-63.

71. Zschocke, J., et al., *Mild trimethylaminuria caused by common variants in FM03 gene.* The Lancet, 1999. **354**(9181): p. 834-835.

72. Bennett, Brian J., et al., *Trimethylamine-N-Oxide, a Metabolite Associated with Atherosclerosis, Exhibits Complex Genetic and Dietary Regulation.* Cell Metabolism, 2013. **17**(1): p. 49-60.

73. Wang, X., et al., *Positional identification of TNFSF4, encoding OX40 ligand, as a gene that influences atherosclerosis susceptibility.* Nat Genet, 2005. **37**(4): p. 365-72.

74. Gotsman, I., A.H. Sharpe, and A.H. Lichtman, *T-cell costimulation and coinhibition in atherosclerosis.* Circ Res, 2008. **103**(11): p. 1220-31.

75. Huang, Q., et al., *Absence of association between atherosclerotic cerebral infarction and TNFSF4/TNFRSF4 single nucleotide polymorphisms rs1234313, rs1234314 and rs17568 in a Chinese population.* J Int Med Res, 2014. **42**(2): p. 436-43.

76. Mehrabian, M., et al., *Identification of 5-lipoxygenase as a major gene contributing to atherosclerosis susceptibility in mice.* Circ Res, 2002. **91**(2): p. 120-6.

77. Funk, C.D. and X.S. Chen, *5-Lipoxygenase and leukotrienes. Transgenic mouse and nuclear targeting studies.* Am J Respir Crit Care Med, 2000. **161**(2 Pt 2): p. S120-4.

78. Ghazalpour, A., et al., *Complex inheritance of the 5-lipoxygenase locus influencing atherosclerosis in mice.* Genetics, 2006. **173**(2): p. 943-51.

79. Mott, R. and J. Flint, *Simultaneous detection and fine mapping of quantitative trait loci in mice using heterogeneous stocks.* Genetics, 2002. **160**(4): p. 1609-18.

80. Solberg, L.C., et al., *A protocol for high-throughput phenotyping, suitable for quantitative trait analysis in mice.* Mammalian Genome, 2006. **17**(2): p. 129-146.

81. Valdar, W., et al., *Genome-wide genetic association of complex traits in heterogeneous stock mice.* Nature Genetics, 2006. **38**(8): p. 879-887.

82. Bennett, B.J., et al., *A high-resolution association mapping panel for the dissection of complex traits in mice.* Genome Research, 2010. **20**(2): p. 281-290.

83. Ghazalpour, A., et al., *Hybrid mouse diversity panel: a panel of inbred mouse strains suitable for analysis of complex genetic traits.* Mammalian Genome, 2012. **23**(9-10): p. 680-692.

84. Park, C.C., et al., *Gene networks associated with conditional fear in mice identified using a systems genetics approach.* BMC Syst Biol, 2011. **5**: p. 43.

85. Farber, C.R., et al., *Mouse genome-wide association and systems genetics identify Asxl2 as a regulator of bone mineral density and osteoclastogenesis.* PLoS Genet, 2011. **7**(4): p. e1002038.

86. Furlotte, N.A., et al., *Increasing association mapping power and resolution in mouse genetic studies through the use of meta-analysis for structured populations.* Genetics, 2012. **191**(3): p. 959-67.

87. Smolock, E.M., et al., *Genetic locus on mouse chromosome 7 controls elevated heart rate.* Physiol Genomics, 2012. **44**(13): p. 689-98.

88. Orozco, Luz D., et al., *Unraveling Inflammatory Responses using Systems Genetics and Gene-Environment Interactions in Macrophages.* Cell, 2012. **151**(3): p. 658-670.

89. Calabrese, G., et al., *Systems genetic analysis of osteoblast-lineage cells.* PLoS Genet, 2012. **8**(12): p. e1003150.

90. Plaisier, C.L., et al., *Zbtb16 has a role in brown adipocyte bioenergetics.* Nutr Diabetes, 2012. **2**: p. e46.

91. Davis, R.C., et al., *Genome-wide association mapping of blood cell traits in mice.* Mamm Genome, 2013. **24**(3-4): p. 105-18.

92. Bennett, B.J., et al., *High-resolution association mapping of atherosclerosis loci in mice.* Arterioscler Thromb Vasc Biol, 2012. **32**(8): p. 1790-8.

93. Hartiala, J., et al., *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, 2014. **34**(6): p. 1307-13.

94. Churchill, G.A., et al., *The Collaborative Cross, a community resource for the genetic analysis of complex traits.* Nat Genet, 2004. **36**(11): p. 1133-7.

95. Roberts, A., et al., *The polymorphism architecture of mouse genetic resources elucidated using genome-wide resequencing data: implications for QTL discovery and systems genetics.* Mamm Genome, 2007. **18**(6-7): p. 473-81.

96. Aylor, D.L., et al., *Genetic analysis of complex traits in the emerging Collaborative Cross.* Genome Res, 2011. **21**(8): p. 1213-22.

97. Churchill, G.A., et al., *The diversity outbred mouse population.* Mammalian Genome, 2012. **23**(9-10): p. 713-718.

98. Svenson, K.L., et al., *High-resolution genetic mapping using the Mouse Diversity outbred population.* Genetics, 2012. **190**(2): p. 437-47.

99. Logan, R.W., et al., *High-precision genetic mapping of behavioral traits in the diversity outbred mouse population.* Genes Brain Behav, 2013. **12**(4): p. 424-37.

100. Recla, J.M., et al., *Precise genetic mapping and integrative bioinformatics in Diversity Outbred mice reveals Hydin as a novel pain gene.* Mamm Genome, 2014. **25**(5-6): p. 211-22.

101. Gatti, D.M., et al., *Quantitative trait locus mapping methods for diversity outbred mice.* G3 (Bethesda), 2014. **4**(9): p. 1623-33.

102. Church, R.J., et al., *Sensitivity to hepatotoxicity due to epigallocatechin gallate is affected by genetic background in diversity outbred mice.* Food Chem Toxicol, 2015. **76**: p. 19-26.

103. French, J.E., et al., *Diversity Outbred Mice Identify Population-Based Exposure Thresholds and Genetic Factors that Influence Benzene-Induced Genotoxicity.* Environ Health Perspect, 2014.

104. Holdt, L.M. and D. Teupser, *From genotype to phenotype in human atherosclerosis--recent findings.* Curr Opin Lipidol, 2013. **24**(5): p. 410-8.

105. Johansen, C.T., S. Kathiresan, and R.A. Hegele, *Genetic determinants of plasma triglycerides.* J Lipid Res, 2011. **52**(2): p. 189-206.

106. Swerdlow, D.I., et al., *The genetics of coronary heart disease.* Br Med Bull, 2012. **102**: p. 59-77.

107. Hardy, J. and A. Singleton, *Genomewide association studies and human disease.* N Engl J Med, 2009. **360**(17): p. 1759-68.

108. Lucas, G., et al., *Hypothesis-based analysis of gene-gene interactions and risk of myocardial infarction.* PLoS One, 2012. **7**(8): p. e41730.

109. Musameh, M., et al., *Analysis of gene-gene interactions among common variants in coronary artery disease.* Heart Lung Circ, 2014. **23 Suppl 2**: p. e19.

110. Sayols-Baixeras, S., et al., *Pathogenesis of coronary artery disease: focus on genetic risk factors and identification of genetic variants.* Appl Clin Genet, 2014. **7**: p. 15-32.

111. Aylor, D.L., et al., *Genetic analysis of complex traits in the emerging Collaborative Cross.* Genome Research, 2011. **21**(8): p. 1213-1222.

112. Yang, H., et al., *Subspecific origin and haplotype diversity in the laboratory mouse.* Nat Genet, 2011. **43**(7): p. 648-55.

113. Welsh, C.E., et al., *Status and access to the Collaborative Cross population.* Mamm Genome, 2012. **23**(9-10): p. 706-12.

114. Yalcin, B., et al., *Sequence-based characterization of structural variation in the mouse genome.* Nature, 2011. **477**(7364): p. 326-9.

115. Sen, S. and G.A. Churchill, *A statistical framework for quantitative trait mapping.* Genetics, 2001. **159**(1): p. 371-87.

116. Lander, E. and L. Kruglyak, *Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results.* Nat Genet, 1995. **11**(3): p. 241-7.

117. Nakamuta, M., et al., *Complete phenotypic characterization of apobec-1 knockout mice with a wild-type genetic background and a human apolipoprotein B transgenic background, and restoration of apolipoprotein B mRNA editing by somatic gene transfer of Apobec-1.* J Biol Chem, 1996. **271**(42): p. 25981-8.

118. Blanc, V., et al., *Intestine-specific expression of Apobec-1 rescues apolipoprotein B RNA editing and alters chylomicron production in Apobec1 -/- mice.* J Lipid Res, 2012. **53**(12): p. 2643-55.

119. Hassan, M.A., et al., *The genetic basis for individual differences in mRNA splicing and APOBEC1 editing activity in murine macrophages.* Genome Research, 2013. **24**(3): p. 377-389.

120. Teng, B., et al., *Effective lowering of plasma, LDL, and esterified cholesterol in LDL receptor-knockout mice by adenovirus-mediated gene delivery of ApoB mRNA editing enzyme (Apobec1).* Arterioscler Thromb Vasc Biol, 1997. **17**(5): p. 889-97.

121. Guenther, C.A., et al., *A molecular basis for classic blond hair color in Europeans.* 2014. **46**(7): p. 748-52.

122. Rosenberg, B.R., et al., *Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3′ UTRs.* Nature Structural & Molecular Biology, 2011. **18**(2): p. 230-236.

123. Suto, J. and K. Sekikawa, *Quantitative trait locus analysis of plasma cholesterol and triglyceride levels in KK x RR F2 mice.* Biochem Genet, 2003. **41**(9-10): p. 325-41.

124. Charlton-Menys, V. and P.N. Durrington, *Apolipoprotein A5 and hypertriglyceridemia.* Clin Chem, 2005. **51**(2): p. 295-7.

125. Sinha, E., et al., *LDL-R AvaII and NcoI polymorphisms: an indirect risk factor for coronary heart disease among a Mendelian population of Delhi, India.* Biochem Genet, 2010. **48**(9-10): p. 807-15.

126. Ramakrishnan, L., et al., *Relationship of APOA5, PPARgamma and HL gene variants with serial changes in childhood body mass index and coronary artery disease risk factors in young adulthood.* Lipids Health Dis, 2011. **10**: p. 68.

127. Weissglas-Volkov, D., et al., *Genomic study in Mexicans identifies a new locus for triglycerides and refines European lipid loci.* J Med Genet, 2013. **50**(5): p. 298-308.

128. Zhou, L., et al., *A genome wide association study identifies common variants associated with lipid levels in the Chinese population.* PLoS One, 2013. **8**(12): p. e82420.

129. Hu, M.S., Z.K. Li, and D.Z. Fang, *[The association study of the LIPC -250g/A polymorphism and high-carbohydrate/low-fat diet induced serum lipid and apolipoprotein concentration changes in healthy youth]*. Sichuan Da Xue Xue Bao Yi Xue Ban, 2013. **44**(5): p. 727-30, 735.

130. Welch, C.L., et al., *Localization of atherosclerosis susceptibility loci to chromosomes 4 and 6 using the Ldlr knockout mouse model.* Proceedings of the National Academy of Sciences, 2001. **98**(14): p. 7946-7951.

131. Zhang, Z., et al., *Genetic analysis of atherosclerosis and glucose homeostasis in an intercross between C57BL/6 and BALB/cJ apolipoprotein E-deficient mice.* Circ Cardiovasc Genet, 2012. **5**(2): p. 190-201.

132. Corbin, K.D., et al., *Genetic signatures in choline and 1-carbon metabolism are associated with the severity of hepatic steatosis.* Faseb j, 2013. **27**(4): p. 1674-89.

133. Lever, M. and S. Slow, *The clinical significance of betaine, an osmolyte with a key role in methyl group metabolism.* Clin Biochem, 2010. **43**(9): p. 732-44.

134. Wang, Z., et al., *Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease.* Nature, 2011. **472**(7341): p. 57-63.

135. Bennett, B.J., et al., *Trimethylamine-N-oxide, a metabolite associated with atherosclerosis, exhibits complex genetic and dietary regulation.* Cell Metabolism, 2013. **17**: p. 49-60.

136. Tang, W.H.W., et al., *Intestinal Microbial Metabolism of Phosphatidylcholine and Cardiovascular Risk.* New England Journal of Medicine, 2013. **368**(17): p. 1575-1584.

137. Koeth, R., et al., *Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis.* Nature Medicine, 2013. **19**(5): p. 576-585.

138. Rak, K. and D.J. Rader, *The diet-microbe morbid union.* Nature, 2012. **472**: p. 40-41.

139. Loscalzo, J., *Lipid Metabolism by Gut Microbes and Atherosclerosis.* Nature, 2011. **472**: p. 57-63.

140. Tang, W.H.W., et al., *Intestinal Microbial Metabolism of Phosphatidylcholine and Cardiovascular Risk.* New England Journal of Medicine, 2013. **368**(17): p. 1575-1584.

141. Shih, D.M., et al., *Flavin containing monooxygenase 3 exerts broad effects on glucose and lipid metabolism and atherosclerosis.* J Lipid Res, 2015. **56**(1): p. 22-37.

142. Koeth, R.A., et al., *Intestinal microbiota metabolism of l-carnitine, a nutrient in red meat, promotes atherosclerosis.* Nature Medicine, 2013. **19**(5): p. 576-585.

143. Ussher, J.R., G.D. Lopaschuk, and A. Arduini, *Gut microbiota metabolism of l-carnitine and cardiovascular risk.* Atherosclerosis, 2013. **231**(2): p. 456-461.

144. O'Connor, A., et al., *Responsiveness of cardiometabolic-related microbiota to diet is influenced by host genetics.* Mamm Genome, 2014. **25**(11-12): p. 583-99.

145. Wei, J., et al., *The clathrin adaptor proteins ARH, Dab2, and numb play distinct roles in Niemann-Pick C1-Like 1 versus low density lipoprotein receptor-mediated cholesterol uptake.* J Biol Chem, 2014. **289**(48): p. 33689-700.

146. Sabatine, M.S., et al., *Metabolomic identification of novel biomarkers of myocardial ischemia.* Circulation, 2005. **112**(25): p. 3868-75.

147. Sreekumar, A., et al., *Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression.* Nature, 2009. **457**(7231): p. 910-4.

148. Chen, H.W., et al., *VH1-44 gene usage defines a subset of canine B-cell lymphomas associated with better patient survival.* Vet Immunol Immunopathol, 2014. **157**(3-4): p. 125-30.

149. Oishi, Y., et al., *Regulatory polymorphism in transcription factor KLF5 at the MEF2 element alters the response to angiotensin II and is associated with human hypertension.* Faseb j, 2010. **24**(6): p. 1780-8.

150. McConnell, B.B., et al., *The diverse functions of Kruppel-like factors 4 and 5 in epithelial biology and pathobiology.* Bioessays, 2007. **29**(6): p. 549-57.

151. Bell, S.M., et al., *Kruppel-like factor 5 controls villus formation and initiation of cytodifferentiation in the embryonic intestinal epithelium.* Dev Biol, 2013. **375**(2): p. 128-39.

152. Guo, M., L.Y. Jan, and Y.N. Jan, *Control of daughter cell fates during asymmetric division: interaction of Numb and Notch.* Neuron, 1996. **17**(1): p. 27-41.

153. Li, P.S., et al., *The clathrin adaptor Numb regulates intestinal cholesterol absorption through dynamic interaction with NPC1L1.* Nat Med, 2014. **20**(1): p. 80-6.

154. Ermann, J. and L.H. Glimcher, *After GWAS: mice to the rescue?* Curr Opin Immunol, 2012. **24**(5): p. 564-70.

155. Flint, J., et al., *Strategies for mapping and cloning quantitative trait genes in rodents.* Nature Reviews Genetics, 2005. **6**(4): p. 271-286.

156. Logan, R.W., et al., *High-precision genetic mapping of behavioral traits in the diversity outbred mouse population.* Genes, Brain and Behavior, 2013. **12**(4): p. 424-437.

157. Smallwood, T.L., et al., *High-Resolution Genetic Mapping in the Diversity Outbred Mouse Population Identifies Apobec1 as a Candidate Gene for Atherosclerosis.* G3 (Bethesda), 2014.

158. Berglund, E.D., et al., *Glucose metabolism in vivo in four commonly used inbred mouse strains.* Diabetes, 2008. **57**(7): p. 1790-9.

159. Rivera, L.R., et al., *Damage to enteric neurons occurs in mice that develop fatty liver disease but not diabetes in response to a high-fat diet.* Neurogastroenterol Motil, 2014. **26**(8): p. 1188-99.

160. Berry, R., et al., *Esrrg functions in early branch generation of the ureteric bud and is essential for normal development of the renal papilla.* Hum Mol Genet, 2011. **20**(5): p. 917-26.

161. Matthews, D.R., et al., *Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man.* Diabetologia, 1985. **28**(7): p. 412-9.

162. Kurosu, H., et al., *Tissue-specific expression of betaKlotho and fibroblast growth factor (FGF) receptor isoforms determines metabolic activity of FGF19 and FGF21.* J Biol Chem, 2007. **282**(37): p. 26687-95.

163. Yang, C., et al., *Differential specificity of endocrine FGF19 and FGF21 to FGFR1 and FGFR4 in complex with KLB.* PLoS One, 2012. **7**(3): p. e33870.

164. Sonoda, J. and Y. Wu, *Fgfr1 agonists and methods of use*. 2012, Google Patents.

165. Hughes, S.E., *Differential expression of the fibroblast growth factor receptor (FGFR) multigene family in normal human adult tissues.* J Histochem Cytochem, 1997. **45**(7): p. 1005-19.

166. Dailey, L., et al., *Mechanisms underlying differential responses to FGF signaling.* Cytokine Growth Factor Rev, 2005. **16**(2): p. 233-47.

167. Miki, T., et al., *Determination of ligand-binding specificity by alternative splicing: two distinct growth factor receptors encoded by a single gene.* Proc Natl Acad Sci U S A, 1992. **89**(1): p. 246-50.

168. Navaratnam, N., et al., *The p27 catalytic subunit of the apolipoprotein B mRNA editing enzyme is a cytidine deaminase.* J Biol Chem, 1993. **268**(28): p. 20709-12.

169. Chester, A., et al., *The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay.* Embo j, 2003. **22**(15): p. 3971-82.

170. Morrison, J.R., et al., *Apolipoprotein B RNA editing enzyme-deficient mice are viable despite alterations in lipoprotein metabolism.* Proc Natl Acad Sci U S A, 1996. **93**(14): p. 7154-9.

171. Okamoto, Y., et al., *Adiponectin reduces atherosclerosis in apolipoprotein E-deficient mice.* Circulation, 2002. **106**(22): p. 2767-70.

172. Okamoto, Y., *Adiponectin provides cardiovascular protection in metabolic syndrome.* Cardiol Res Pract, 2011. **2011**: p. 313179.

173. Ratcliffe, N.R., S.M. Kennedy, and P.M. Morganelli, *Immunocytochemical detection of Fcgamma receptors in human atherosclerotic lesions.* Immunol Lett, 2001. **77**(3): p. 169-74.

174. Krimbou, L., et al., *Interaction of lecithin:cholesterol acyltransferase (LCAT).alpha 2-macroglobulin complex with low density lipoprotein receptor-related protein (LRP). Evidence for an alpha 2-macroglobulin/LRP receptor-mediated system participating in LCAT clearance.* J Biol Chem, 2001. **276**(35): p. 33241-8.

175. Silva, J., et al., *Nanog is the gateway to the pluripotent ground state.* Cell, 2009. **138**(4): p. 722-37.

176. Bowles, J., et al., *Dppa3 is a marker of pluripotency and has a human homologue that is expressed in germ cell tumours.* Cytogenet Genome Res, 2003. **101**(3-4): p. 261-5.

177. Sluimer, J.C., et al., *Dead or alive: gene expression profiles of advanced atherosclerotic plaques from autopsy and surgery.* Physiol Genomics, 2007. **30**(3): p. 335-41.

178. Aerni-Flessner, L., et al., *GLUT4, GLUT1, and GLUT8 are the dominant GLUT transcripts expressed in the murine left ventricle.* Cardiovasc Diabetol, 2012. **11**: p. 63.

179. Tobin, J.F. and A.J. Celeste, *Bone morphogenetic proteins and growth differentiation factors as drug targets in cardiovascular and metabolic disease.* Drug Discov Today, 2006. **11**(9-10): p. 405-11.

180. Brown, D.A., et al., *Concentration in plasma of macrophage inhibitory cytokine-1 and risk of cardiovascular events in women: a nested case-control study.* Lancet, 2002. **359**(9324): p. 2159-63.

181. Fogelman, A.M., *TMAO is both a biomarker and a renal toxin.* Circ Res, 2015. **116**(3): p. 396-7.

182. Zhao, N., et al., *Mechanism of kidney injury caused by bevacizumab in rats.* Int J Clin Exp Pathol, 2014. **7**(12): p. 8675-83.

183. Eilbeck, K. and S.E. Lewis, *Sequence ontology annotation guide.* Comp Funct Genomics, 2004. **5**(8): p. 642-7.

184. Joss, S., et al., *De novo translocation (1; 2)(q32; p25) associated with bilateral renal dysplasia.* Clin Genet, 2003. **63**(3): p. 239-40.

185. Costantini, F., *Genetic controls and cellular behaviors in branching morphogenesis of the renal collecting system.* Wiley Interdiscip Rev Dev Biol, 2012. **1**(5): p. 693-713.

186. Rampersaud, E., et al., *Identification of novel candidate genes for type 2 diabetes from a genome-wide association scan in the Old Order Amish: evidence for replication from diabetes-related quantitative traits and from independent populations.* Diabetes, 2007. **56**(12): p. 3053-62.

187. Sangrajrang, S., et al., *Genetic polymorphisms of estrogen metabolizing enzyme and breast cancer risk in Thai women.* Int J Cancer, 2009. **125**(4): p. 837-43.

188. Skuse, G.R., et al., *The neurofibromatosis type I messenger RNA undergoes base-modification RNA editing.* Nucleic Acids Res, 1996. **24**(3): p. 478-85.

189. Jia, L., J.L. Betters, and L. Yu, *Niemann-pick C1-like 1 (NPC1L1) protein in intestinal and hepatic cholesterol transport.* Annu Rev Physiol, 2011. **73**: p. 239-59.

190. Albright, J., et al., *Genetic network identifies novel pathways contributing to atherosclerosis susceptibility in the innominate artery.* BMC Med Genomics, 2014. **7**: p. 51.

191. Romano, K.A., et al., *Intestinal microbiota composition modulates choline bioavailability from diet and accumulation of the proatherogenic metabolite trimethylamine-N-oxide.* MBio, 2015. **6**(2).

192. Benson, A.K., et al., *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, 2010. **107**(44): p. 18933-8.

193. Harrison, M.M., et al., *A CRISPR view of development.* Genes Dev, 2014. **28**(17): p. 1859-72.

194. Wijshake, T., D.J. Baker, and B. van de Sluis, *Endonucleases: new tools to edit the mouse genome.* Biochim Biophys Acta, 2014. **1842**(10): p. 1942-1950.