Identification of Genetic Factors in Atherosclerosis Using an Apoe Mouse Model

Andrew Todd Grainger
Los Gatos, California

Bachelors of Science in Molecular Biology, University of California San Diego, San Diego, California, 2013
Masters of Science in Biology, University of California San Diego, San Diego, California, 2014

A Dissertation Presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Doctor of Philosophy

Department of Biochemistry and Molecular Genetics

University of Virginia
December 2019

Dr. Weibin Shi

Dr. Charles Farber
Dr. Aakrosh Ratan
Dr. Norbert Leitinger

Abstract

Atherosclerosis is the primary cause of coronary artery disease (CAD), ischemic stroke and peripheral arterial disease. Despite major achievements made in the past few decades, CAD and atherosclerosis-related events remain the number one cause of death in the United States and other developed countries. Therefore, there is a critical medical need to develop novel and effective therapies.

An effective way to find new targets for intervention is through conducting genetic studies in animal models. When deficient in *Apoe*, mouse strains BALB/cJ and SM/J exhibit distinct differences in atherosclerosis and its associated risk factors. We hypothesized that linkage analysis of progeny derived from these inbred strains would lead to the discovery of new genes and new pathways in atherosclerosis and its associated cardiometabolic phenotypes. F2 mice were generated from an intercross between the two *Apoe*$^{-/-}$ strains and fed 12 weeks of Western diet. Many QTL loci were mapped for plasma lipids and glucose, carotid lesion size, and aortic lesion size. This included a significant QTL for aortic atherosclerosis, *Ath49*, which was mapped to the major histocompatibility region. Moreover four novel QTLs for carotid atherosclerosis, two significant QTLs named *Cath7* on chromosome 5 and *Cath8* on chromosome 9 and two suggestive QTLs, *Cath5* and *Cath6* on chromosomes 15 and 18 respectively, were mapped. Through a combination of haplotype analysis and a novel strategy employing gene expression, aortic lesion size correlation, and eQTL data, we prioritized *Mep1α* as a promising candidate gene for *Ath49*. We generated double knockouts and found that *Mep1α* is a novel gene negatively affecting plaque formation. Finally, we developed a method utilizing machine learning-based segmentation to accurately quantify subcutaneous and visceral fat volumes in mice using MRI and humans using CT. We found that BMI, a commonly used measure for diagnosing obesity, is only moderately associated with subcutaneous fat and has no association with visceral fat volume in humans.

Dedication

To my parents, Jim and Claire: For giving me unconditional love and providing me with the care and support I needed to follow my dreams. I could never have gotten where I am today without you, and I love you.

To my mentor, Weibin Shi: For believing in me and providing me with the platform and trust to thrive. You have given me the ability to grow into a scientist I can be proud of.

Table of Contents

List of Figures

List of Abbreviations

CAD	coronary artery disease
CVD	cardiovascular disease
T2D	type 2 diabetes
LDL-C	low-density lipoprotein cholesterol
TG	triglyceride
HDL-C	high-density lipoprotein cholesterol
VCAM-1	vascular cell adhesion molecule 1
MCP-1	monocyte chemotractant protein-1
oxLDL	oxidized low-density lipoprotein
IL-1β	interleukin 1 beta
SMC	smooth muscle cell
GWAS	genome-wide association study
SNP	single-nucleotide polymorphism
QTL	quantitative trait locus
Apoe	apolipoprotein E
LDLR	low-density lipoprotein receptor
Mb	megabase
eQTL	expression quantitative trait locus
CI	confidence interval
WGCNA	weighted gene co-expression network analysis
HDMP	hybrid diversity mouse panel
MHC	major histocompatibility complex
Mep1α	meprin 1 alpha
MRI	magnetic resonance imaging

CT	computed tomography
B6	C57BL/6 inbred mouse strain
C3H	C3H/HeJ inbred mouse strain
BALB	BALB/cJ inbred mouse strain
SM	SM/J inbred mouse strain
LOD	logarithm of odds
HbA1c	glycated hemoglobin
OR	odds ratio
H&E	hematoxylin and eosin
SMC-A	smooth muscle cell actin
MAC3	galectin 3
LY6G	leukocyte antigen-6
CXCL5	C-X-C Motif Chemokine Ligand 5
MDA	malondialdehyde
ECM	extracellular matrix
PFA	paraformaldehyde
BMI	body mass index
ANTs	advanced normalization tools
RMSE	root mean squared error
ANOVA	analysis of variance
SAT	subcutaneous adipose tissue
VAT	visceral adipose tissue
RSE	residual squared error
HU	hounsfield units
PACS	picture archiving and communication system

TIFF	tagged image file format

Chapter 1

Introduction

1.1 Atherosclerosis Pathogenesis

1.1.1 Atherosclerosis Is The Leading Health Burden Worldwide

Atherosclerosis is a complex disease that is the primary contributor to coronary artery disease (CAD) and cardiovascular disease (CVD), the two major causes of heart attack, ischemic stroke and peripheral arterial disease. Despite decades of work combating CAD and CVD, cardiovascular-related events remain the number one cause of death in the United States and throughout the world. In 2013, they accounted for approximately 17.3 million of the 54 million total deaths, or 31.5% (95% uncertainty interval, 30.3%–32.9%) of all global deaths[1]. In 2016, heart disease accounted for 635,260 deaths, or 23.1% of all deaths in the United States[2]. This number is comparable to the number of all deaths from malignant neoplasms and accidents, the second and third leading causes of death, combined (27.6%). Despite it being the world's largest health problem, death rates attributable to CVD have declined 25.3% from 2004 to 2014. However by 2030, 43.9% of the US adult population is projected to have some form of CVD. So while the percentage of deaths has been declining, the number of cardiovascular related mortalities will still remain the most significant public health risk.

Common risk factors for CAD include high low-density lipoprotein (LDL) levels, obesity, and diabetes[3]. In the United States, in adults aged ≥ 20 years, the co-occurrence of age-adjusted hypertension is 30.0%, the co-occurrence of high cholesterol is 27.8% (>240 mg/dL or high-cholesterol medication), and diagnosed and undiagnosed type-2 diabetes mellitus (T2D) is 11.4%.

The largest amount of co-occurrence is with individuals considered overweight or obese, where 68.8% have both high body mass index (BMI) and CAD. The diagnosis of obesity has primarily relied on a few anthropometric measurements, such as body mass index (BMI), waist circumference, or waist-to-hip ratio. A BMI of ≥30 kg/m^2 is considered

obese, and a BMI of 25 to <30 kg/m^2 is defined as overweight. However, these indirect measurements neither allows for distinguishing fat from skeletal muscle nor distinguishing between visceral and subcutaneous fat. Excessive body fat rather than skeletal muscle is related to both increased plasma levels of free fatty acids and proinflammatory cytokines as well as endoplasmic reticulum stress, all of which contribute to development of insulin resistance, type 2 diabetes, and atherosclerosis[4]. Central or abdominal obesity has been shown to be more closely associated with risk of coronary artery disease and type 2 diabetes[5,6].

Thus, there is a medical demand for accurately measuring the amount and distribution of body fat to better understand its impact on health and disease. Such imaging modalities as magnetic resonance imaging (MRI) and computed tomography (CT) can clearly distinguish fat from other tissues and thus allow for accurate measurement of fat and non-fat tissue amounts[7]. Quantification of body fat volume using MRI or CT involves analysis of many cross-sectional or longitudinal slices across the region of interest. Thus, manual measurement of fat volume with MRI or CT images is a tedious and time-consuming task. To save time and also reduce subjective influences from observers, several semi-automated algorithms have been developed for quantifying body fat[8–11]. However, nearly all of the algorithms are still dependent on expert knowledge for tuning the features of images and their accuracy and reliability are often low. Thus, there is a large demand for techniques that can feasibly quantify these fat deposits and accelerate our understanding of the pathogenesis of obesity and it's relationship with T2D and atherosclerosis.

There is strong evidence to support the claim that insulin resistance, hyperinsulinemia, and parental and familial history of diabetes mellitus or impaired glucose tolerance is associated with a higher probability in adult offspring of exhibiting a CAD risk profile or experiencing myocardial infarction[12–15]. Moreover, diabetic patients have a 2∼4

fold increased risk of developing CAD in general. Therefore while T2D has a lower co-occurrence rate than other risk factors, it is also included in the risk profile for CAD.

While each of these risk factors individually increases a person's risk for CAD, the presence of multiple risk factors translates to an even higher risk level. These observations have led to the widespread popularity of the metabolic syndrome concept, a constellation of risk factors that confers an elevated risk of CAD and CVD. This metabolic syndrome concept has been accepted by the medical community and is the current strategy used by clinicians for predicting overall risk of CAD complications or development of T2D[16]. Therefore understanding hyperlipidemia, obesity, T2D, and hypertension will inevitably also enhance our understanding of atherosclerosis and provide novel avenues for discovering new and better ways of preventing both CAD and it's associated health complications.

1.1.2 Current Treatment Strategies For Atherosclerosis

Numerous therapeutic strategies have been developed over the years for prevention and treatment of CVD[17]. Those predicted to have a mild risk of developing cardiovascular disease are recommended to change lifestyle habits to help with reducing dyslipidemia, obesity, and T2D. The idea behind this is that with reductions in the co-occurring diseases, chances of a CAD-related event will also decrease. These lifestyle changes include eating less fats and sugar and eating more whole grains, fruits, vegetables, and fish, an increase in exercise, a reduction in alcohol intake, and a cessation of smoking[18,19]. If the risk of developing CAD is high enough, additional drug therapy is typically considered. These drugs are primarily focused on attenuating dyslipidemia through reducing LDL-C, NonHDL-C, and TG levels while increasing HDL-C levels (**Table 1.1**). More recently some alternative medications have been explored, including additional application of aspirin and long-term antiplatelet therapies. However, neither these nor lipid medication-based strategies have shown large amounts of success[20,21]. For severe cases in which patients have already

developed CAD, interventional strategies such as angioplasty, endarterectomy, and bypass surgery are employed[22–24]. These strategies are implemented with a high degree of efficacy and a low degree of interventional-related mortalities, and do decrease the overall number of heart attacks and stroke occurrences[1,23]. Despite this, CAD remains the number one cause of death worldwide[2]. Therefore, there is a large window for improvement in our prevention regimens and discovering novel players in atherosclerosis in order to develop more effective therapeutics is of great importance.

Table 1.1 *List of current therapeutics for treatment of dyslipidemia in relation to cardiovascular disease*

Category	**Major drug name**
Statin	Pravastatin, Simvastatin, Fluvastatin, Atorvastatin, Pitavastatin, Rosuvastatin
Intestinal cholesterol transporter inhibitor (Cholesterol absorption inhibitor)	Ezetimibe
Anion exchange resin	Colestimide, Cholestyramine
Probucol	Probucol
PCSK9 inhibitor	Evolocumab, Alirocumab
MTP Inhibitor	Lomitapide
Fibrate	Bezafibrate, Fenofibrate, Clinofibrate, Clofibrate
Selective peroxisome proliferator-	Pemafibrate

activated receptor α modulator (SPPARMα)	
Nicotinic acid derivative	Niceritrol, Nicomol, Tocopheryl Nicotinate
N-3 polyunsaturated fatty acid	Ethyl icosapentate, Omega-3-acid ethyl ester

Table adapted from [19]*.*

1.1.3 Atherosclerotic Plaque Formation

Atherosclerosis refers to the development of atheromatous plaques in the inner lining of large and medium-sized arteries. Though murine experiments and observations in human specimens, we have greatly increased our understanding of how these plaques develop. Most schemes of atherogenesis highlight the initial instigator as a change in the monolayer of endothelial cells that lines the inner arterial surface[25–27]. This process is initiated by the deposition of lipids within in the aortic wall. This lipid deposition, despite a systemic rise in risk factors such as increased LDL levels or blood pressure, is likely due to differing haemodynamics in specific parts of the arterial tree. In these specific locations, the ability of normal laminar shear stress to cause the endothelium to express an atheroprotective gene expression profile is inhibited[28,29]. This change in endothelial gene expression then permits the lipids to transmigrate and deposit into the intimal space at these locations.

Once within the intimal space between the endothelial and medial layers, these lipids become oxidized due to oxygen free radicals which have been hypothesized to be produced by resident arterial macrophages, endothelial cells, and smooth muscle cells[30].

Once the oxidized lipids are recognized by the resident macrophages' and mast cells' scavenger receptors, they express pro-inflammatory cytokines. These cytokines cause arterial endothelial cells to express adhesion molecules that capture leukocytes on their surfaces, the primary of which is vascular cell adhesion molecule 1 (VCAM-1)[31,32]. After monocytes adhere to the endothelium, they migrate into the intima under the influence of produced chemokines such as monocyte chemotractant protein-1 (MCP-1)[33]. These chemokines are typically produced by endothelial cells, macrophages and smooth muscle cells[34,35].

This process of endothelial cell activation and monocyte invasion is already occurring in individuals where risk factors such as hypertension and endothelial cell damage are present[27]. Added strain from hypertension or the wound healing response from endothelial cell damage can both cause an innate immune response that will bring about the pro-atherosclerotic changes in the surrounding endothelial cells. These changes will subsequently permit easier uptake of circulating LDL, even if it is not present in increased amounts[32].

Once in the intima, monocytes differentiate into M1 macrophages in order to process the oxLDL[36]. The uptake of large amounts of oxLDL results in excess cholesterol deposition in macrophages and ultimately leads to the formation of foam cells. These foam cells produce a variety of pro-inflammatory cytokines, the primary of which is IL-1β[37,38]. These pro-inflammatory cytokines are released into the blood and promote the recruitment of other leukocytes, inducing a feed-forward mechanism of immune cell recruitment and a chronic local and systemic pro-inflammatory environment. The foam cells collect in the intimal space and over time a large percentage of these foam cells die, primarily through apoptosis. Upon death they release the lipids and cholesterol into the surrounding intimal

space. This collection of apoptotic cells forms a lipid-rich pool called the necrotic core of the plaque[39].

The other major factor involved in plaque progression is smooth muscle cell (SMC) migration and fibrous cap development [40]. In the arterial intima layer, SMCs migrate and produce extracellular matrix proteins, including interstitial collagen and elastin. These migratory cells primarily locate to the intimal space proximal to the endothelial cell layer. Once there, these SMCs form a fibrous cap that typically covers the necrotic core and protects the plaque from rupture. SMCs can also play an inflammatory role akin to macrophages. An alternative gene expression profile is activated when the SMCs uptake oxidized lipids. This alternate gene expression involves an activation of macrophage markers and a gain of phagocytic activity[41–43]. In this way, these SMCs act in a similar fashion to transmigrated M1 macrophages and promote the pro-inflammatory environment of the plaques.

Thrombi are the final consequence of the growth of atherosclerotic plaques and are the ultimate cause of heart attack and stroke. Thrombi often occur after the plaque is physically disrupted, most commonly through a combination of erosion and rupture of the fibrous cap. This rupture causes pro-coagulant material in the plaque to be exposed to coagulation proteins in the blood. This clot enters the bloodstream and triggers a heart attack or stroke, depending on the location of the plaque. Plaques that rupture typically have thin, collagen-poor fibrous caps with few SMCs and abundant macrophages with a large necrotic core[44]. Large amounts of inflammatory cells may hasten plaque disruption due to their expression and release of collagenolytic enzymes that can degrade collagen. They can also generate mediators that promote the death of SMCs, the source of arterial collagen[45]. Overall, plaque formation is primarily mediated through the abundance of circulating lipids and the pro-inflammatory response induced by both M1 macrophages and

migrating SMCs. The likelihood of plaque rupture and the occurrence of heart attack or stroke is primarily determined through the plaque composition. While our knowledge of atherosclerosis has greatly improved over the last few decades, a large amount of work remains in understanding the intricate processes that influence it. One promising tactic for discovering novel pathways and players involved is to discover the underlying genetic architecture influencing an individual's susceptibility to atheromic plaque formation.

1.2 Genetics of Atherosclerosis

1.2.1 Before GWAS

The important role of genetic factors in atherosclerosis has been clearly demonstrated in numerous early studies, including prospective studies of twins, families and cohorts[46–48]. Through measures of atherosclerosis such as myocardial infarction, carotid intima-media thickness or coronary angiography, and measurements of risk factors such as plasma lipid levels or blood pressure, these studies have found the heritable component to be somewhere in the range of 30-60%[25,46,49]. While the influence of genetics was apparent, finding the genetic components involved remained difficult with the limited genetic resources available. Most candidate gene experiments in mice were underpowered and were met with little success, with the exception of a select few genes including *ABO, LDLR, APOB, CBS, and ATPA1*[50,51]. Therefore more powerful techniques for uncovering the heritable components were needed, and these came in the form of Genome-Wide Association Studies (GWAS).

1.2.2 Human Genome-Wide Association Studies (GWAS) for Coronary Artery Disease

Over the past decade and a half, genome-wide association studies (GWAS) have yielded remarkable advances in the understanding of complex traits. Hundreds of thousands of genetic risk variants have been identified in humans for a wide variety of heritable traits and complex diseases[52–54]. As of 10-14-2019, the GWAS Catalog contains 7796 publications and 159202 associations (https://www.ebi.ac.uk/gwas/). A typical GWAS analyzes on average hundreds of thousands to millions of single-nucleotide polymorphisms (SNPs) from each individual to test for an association between each variant and a phenotype of interest[55(p20)]. GWAS have confirmed the heritability of many human traits[56], clarified their underlying genetic architecture[57], and have identified novel biological mechanisms and drug targets[58]. Over the years, a number of GWAS have been performed for case-control diagnosis of coronary artery disease (CAD) and cardiovascular disease (CVD) and have managed to identify hundreds of associated loci[59–61]. With increasing availability of genotyped individuals and a larger number of individuals being introduced into the analyses, the heritable component underlying CAD/CVD that has been identified has been continually improving. Novel strategies using a 5% FDR cutoff instead of the typical –log(P) cutoff for significant associations of 10^8 have uncovered the most associations, for a current number of 304 independent variants at 243 loci[62]. However even with this strategy, only 21.2% of the heritable component of CAD/CVD has been identified. This does not come close to the 30-60% heritability predicted through early twin studies, and highlights the fact that a large amount of genetic influence remains to be uncovered.

One novel hypothesis for how to find the total influence of genetics on a phenotype is called the omnigenic model[63]. In essence, it describes a situation in which a network of genes with small effect sizes work through primary core genes, or the genes having a direct influence on the phenotype. The SNPs influencing CAD are working either in *cis* through direct regulation of the core genes, or in *trans* through peripheral genes of small effect size

that are in turn affecting the core genes. From this hypothesis, the 304 variants for CAD and CVD that have been identified to date[62] are *cis*-effectors for the core genes with large effect sizes involved in CAD. However, lots of *trans*-effectors with very small effect sizes that influence these core genes remain to be discovered. The way to find the missing heritability is to find these *trans*-effector SNPs and understand the genomic landscapes of how all of these SNPs are contributing. According to the model, this would most likely consist of the majority of the genome. However, these omnigenic model ideas are new concepts, and strategies to find these *trans*-effector SNPs remain undiscovered to date. Therefore to uncover this missing heritability that current human GWAS has not found, the option we and others have chosen is to use model organisms.

1.2.3 Linkage Analysis in Mice for Aortic Atherosclerosis

Model organisms like mice can facilitate the discovery of the missing genes by taking advantage of the genetic differences between inbred strains. By using these genetic differences, one can discover genomic regions linked to atherosclerosis progression and susceptibility using quantitative trait locus (QTL) analysis. The QTL analysis process typically involves generating a cross between inbred strains that differ in one or more phenotypes of interest (typically an F2 intercross). These F2 mice are then genotyped with genetic markers evenly spaced across the genome and phenotyped for the traits of interest. The final step is the linking of the genotypes at those markers with the phenotype to find specific parts of the genome that are likely influencing the phenotype[64–66]. This linkage analysis is primarily performed using either the qtl package for R, a program commonly used for the statistical analysis of rodent QTL crosses, or a Java-based graphical interface for R/qtl, J/qtl[67,68].

There are major benefits to using mice for genetic studies compared to humans. Firstly, it allows for strict control over environmental influence to minimize variability between individuals. Second, it allows for accurate phenotypic characterization of the atherosclerotic lesions through direct quantification of the lesion size. While these two things can help uncover loci human models cannot, one important limitation to using mice is that wild-type inbreds or outbreds do not naturally produce atherosclerotic lesions. To circumvent this, models were created in which Apolipoprotein E (*Apoe*) or LDL receptor (*Ldlr)* were knocked out to induce an extremely hyperlipidemic state[69]. This large influx of circulating lipids, primarily low-density lipoprotein (LDL), inevitably causes LDL to be deposited in the aorta and for atherosclerotic plaques to reliably develop. This increase in LDL co-occurs with a decrease in the athero-protective high-density lipoprotein (HDL). This hyperlipidemia is exacerbated when mice are fed a high-fat, high-cholesterol diet, termed Western diet. Using this model, a large number of QTLs for LDL and HDL have been discovered from 10 separate studies involving 23 different inbred strains of mice spanning every chromosome in the mouse genome[70].

Intercrosses of inbred strains on an *Apoe*$^{-/-}$or LDL *Lldlr*$^{-/-}$ background reproduce all phases of atherosclerotic lesions seen in humans and have therefore been the models of choice for QTL generation for aortic atherosclerosis[71,72]. Depending on the inbred strain, *Apoe*$^{-/-}$ mice display a large range of susceptibility to developing atherosclerosis, with females having larger lesion sizes than males (**Figure 1.2**). For this reason, females have been primarily used for aortic QTL studies. This diversity is observed even amongst the more commonly used inbred strains such as C57BL/6 and C3H/HeJ and remains consistent when fed either a normal (chow) diet or Western diet (**Figure 1.3**). When choosing two inbred strains to generate an intercross, one should pick two with a large difference in atherosclerosis susceptibility. This will cause the F2 mice to have a large amount of

phenotypic variation, typically forming a normal distribution of aortic lesion sizes. This large amount of F2 phenotype variation yields the highest likelihood of discovering genetic loci linked to the phenotype of interest. This has held true for aortic atherosclerosis QTL studies we and others have performed. To date, 119 suggestive and significant QTLs have been discovered over 15 male and female intercrosses spanning 13 inbred strains and every chromosome on the mouse genome, many of which overlap on similar regions of the mouse genome (**Figure 1.4**) (**Table 1.1**) [73–90].

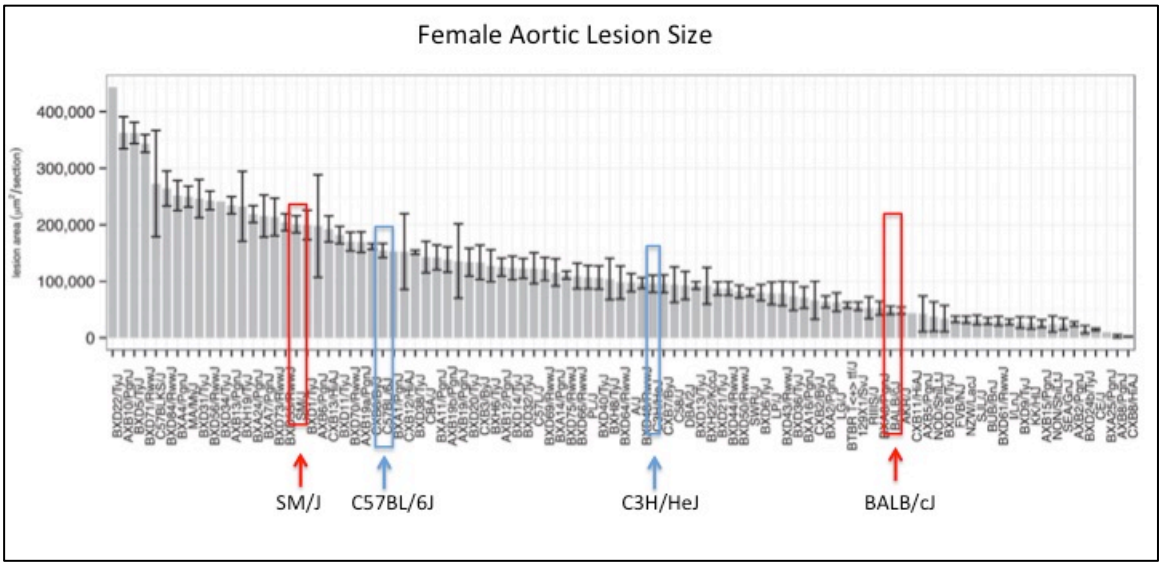

Figure 1.1 *Aortic lesion sizes of inbred and recombinant inbred strains on the Apoe$^{-/-}$ background fed Western diet for 16 weeks. Strains highlighted in blue or red are inbred strains we have previously (blue + BALB/cJ) or are currently using (BALB/cJ + SM/J) for F2 intercrosses to discover QTLs for aortic atherosclerosis. Figure adapted from* [91].

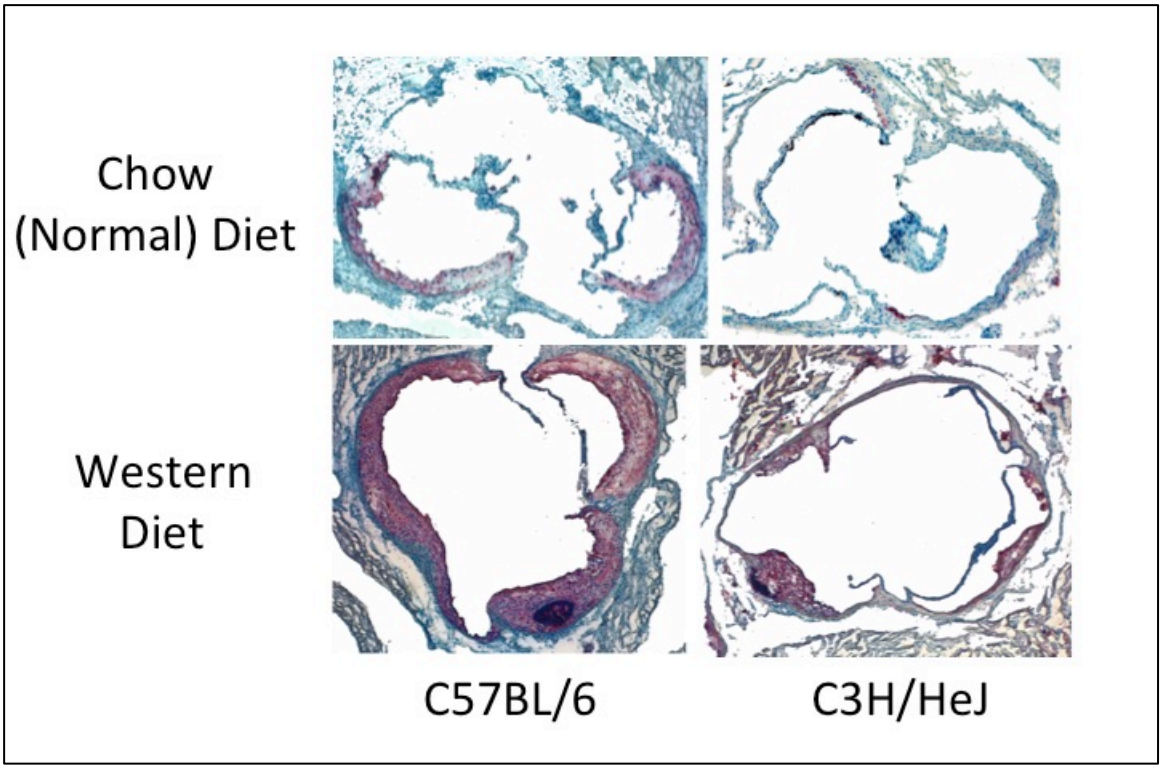

Figure 1.2 *Cross-sections of aortic lesions in the aortic root from inbred strains with different atherosclerosis susceptibility on an Apoe$^{-/-}$ background. Aortic lesions are stained with Oil-Red-O (red) when C57BL.6 (left) or C3H/HeJ (right) mice were fed either chow (top) or Western (bottom) diet.*

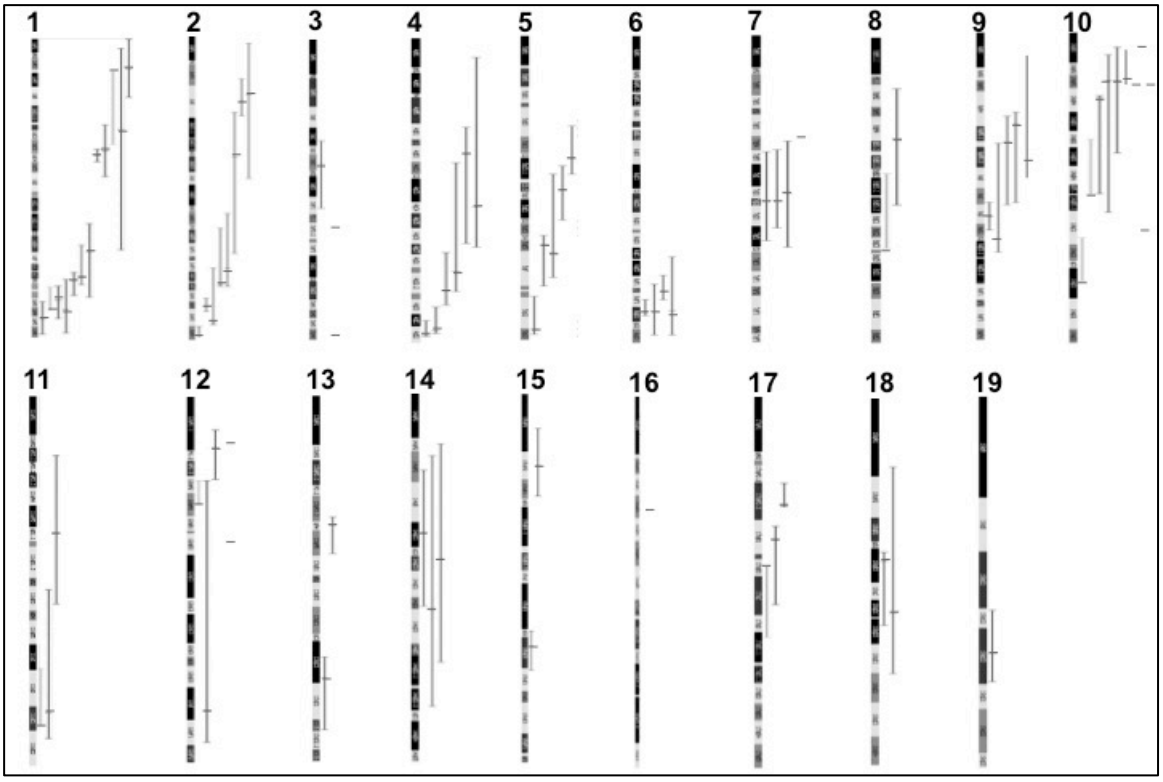

Figure 1.3 *All published quantitative trait loci (QTLs) for aortic lesion size in Apoe$^{-/-}$ or Ldlr$^{-/-}$ mice fed Western diet. Horizontal dashes indicate recorded QTL peak and vertical lines indicate the 95% confidence intervals (CI). QTLs were visualized using a custom track imported into the UCSC genome browser using mouse genome mm10. The custom track is available as* ***S1.1****.*

Table 1.2: *QTLs identified for aortic lesion area in Apoe$^{-/-}$ and Ldlr$^{-/-}$ mice*

Chr	QTL[a]	peak (Mb)[b]	CI (Mb)[c]	Cross	High Allele[d]	M/F	PMID
1	Suggestive	19,292,010	1 - 37,906,000	BALB x SMJ	BALB	F	27736672
	Suggestive	20,871,663	20,871,663 - 68,723,270	B6 x FVB	?	M + F	11973313
	Suggestive	59,000,000	6,000,000 - 138,000,000	DBA x 129S	DBA	M + F	24586312
	Suggestive	73,000,000	56,000,000 - 90,000,000	DBA x 129S	DBA	F	24586312
	Ath30	75,956,444	72,000,000 - 80,000,000	C3H x B6	B6	F	17641228
	Ath43	139,000,000	121,000,000 - 169,000,000	DBA x 129S	DBA	F	24586312
	Ath1	155,924,281	135,000,000 - 161,000,000	C3H x B6	B6	M + F	17641228
	Ath44	158,300,000	153,000,000 - 168,000,000	DBA x 129S	DBA	M + F	24586312

	Suggestive	171,000,000	162,000,000 - 183,000,000	DBA x 129S	DBA	M	24586312
	Suggestive	177,269,841	162,250,311 - 177,269,841	BALB x SMJ	SMJ	F	27736672
	Ath9	177,285,317	157,361,794 - 192,529,115	B6 x FVB	FVB	M	11973313
	Ath1	183,092,234	172,000,000 - 193,241,301	BALB x B6	B6	F	22294616
2	Ath47	34,400,000	3,886,211 - 84,817,798	BALB x B6	B6	F	21252064
	Ath39	38,239,859	25,000,000 - 47,000,000	BALB x B6	B6	M	21252064
	Ath41	71,964,058	45,000,000 - 130,000,000	BALB x B6	B6	F	22294616
	Ath46	139,000,000	106,000,000 - 150,000,000	DBA x 129S	DBA	M	24586312
	Athla1	148,680,255	115,000,000 - 150,000,000	PERA x B6	B6	M + F	15514201
	Ath45	162,000,000	157,000,000 - 165,000,000	DBA x 129S	DBA	M + F	24586312
	Ath35	172,000,000	139,000,000 - 173,000,000	DBA x 129S	DBA	M	24586312
	Ath28	178,800,659	162,355,057 - 181,873,487	B6 x C3H	C3H	M	23938286
	Ath28	179,574,406	174,000,000 - 179,000,000	DBA x AKR	DBA	M	16373612
3	Ath23	66,450,410	54,000,000 - 90,000,000	DBA x AKR	AKR	F	16373612
	Ascla4	115,830,548	?	B6 x FVB	B6	F	16380418
	Ascla3	148,143,886	?	B6 x FVB	B6	F	16380418
4	Ath8	55,251,078	45,230,241 - 150,592,401	B6 x C3H	B6	M	23938286
	Ath8	59,552,657	45,658,566 - 105,299,798	NZB x SMJ	SMJ	F	15530921
	Ath4	86,450,248	9,702,715 - 107,112,550	B6 x C3H	B6	F	17916774
	Suggestive	119,721,657	63,898,275 - 129,953,782	BALB x SMJ	Heterosis	F	27736672
	Athsq1	129,572,122	110,000,000 - 136,927,309	BALB x B6	BALB	F	22294616
	Athsq1	149,926,199	137,890,684 - 151,570,093	B6 x MOLF	MOLF	F	11438740
	Athsq3	151,077,699	145,000,000 - 153,000,000	B6 x C3H	B6	F	17641228
5	Suggestive	61,506,687	44,971,670 - 69,163,000	BALB x SMJ	SMJ	F	27736672
		77,028,226	65,000,000 - 92,000,000	B6 x C3H	B6	M + F	17641228
	Ath42	104,351,018	99,690,738 - 124,829,348	BALB x B6	B6	F	22294616
	Suggestive	108,384,913	69,163,000 - 120,578,000	BALB x SMJ	SMJ	F	27736672
	Ath24	146,583,524	130,000,000 - 149,000,000	DBA x AKR	DBA	M + F	16373612
6	Ath37	124,468,231	116,665,031 - 128,519,025	CAST x B6	B6	M + F	16624897
	Artles	133,853,029	107,795,846 - 145,656,013	CAST x B6	B6	M + F	11463718
	Athsq2	134,083,123	121,115,387 - 145,656,013	B6 x MOLF	B6	M + F	11438740
	Ath38	134,184,819	128,519,025 - 136,316,178	CAST x B6	B6	M + F	16624897

7	Aorls2	49,744,933	?	B6 x DBA	DBA	M + F	12925895
	Ath31	74,629,649	50,000,000 - 100,000,000	B6 x C3H	B6	F	17641228
	Suggestive	78,000,000	55,000,000 - 97,000,000	DBA x 129S	129S	M + F	24586312
	Suggestive	78,000,000	54,000,000 - 91,000,000	DBA x 129S	129S	F	24586312
8	Ath40	42,935,531	21,959,500 - 71,045,600	BALB x B6	B6	M	21252064
	Suggestive	89,351,811	57,763,542 - 89,351,811	BALB x SMJ	SMJ	F	27736672
	Suggestive	?	?	B6 x DBA	B6	M + F	12925895
9	Ath29	37,133,055	32,500,594 - 68,829,021	BALB x B6	B6	F	22294616
	Ath29	44,340,026	34,067,647 - 69,831,153	B6 x C3H	B6	M + F	17916774
	Ath29	49,500,000	9,140,000 - 59,140,000	BALB x SM	Heterosis	M	27736672
	Ath29	74,700,643	68,829,021 - 80,091,050	B6 x C3H	B6	F	16387874
	Ath29	80,091,050	80,091,050 - 96,397,888	B6 x C3H	B6	M	23938286
	Ath29	83,693,797	45,000,000 - 89,000,000	B6 x C3H	B6	M + F	17641228
10		6,202,049	?	B6 x FVB	?	M	11973313
	Ascla2	18,806,860	?	B6 x FVB	FVB	M	16380418
	Ath11	20,407,588	6,202,049 - 50,788,902	B6 x FVB	FVB	M + F	11973313
	Ath11	20,203,863	9,160,000 - 23,160,000	BALB x SMJ	BALB	M	27736672
	Ath20	20,407,588	9,413,521 - 75,681,262	129S x B6	129S	F	14592847
	Ascla1	21,715,442	?	B6 x FVB	FVB	F	16380418
	Suggestive	28,450,065	26,797,957 - 67,821,343	BALB x SMJ	BALB	F	27736672
	Ath17	68,255,809	45,000,035 - 68,255,809	129S x B6	129S	F	14592847
	Aorls1	82,655,287	?	B6 x DBA	DBA	M + F	12925895
	Suggestive	104,516,306	86,301,142 - 104,516,306	BALB x SMJ	BALB	F	27736672
11		45,275,149	19,745,966 - 68,609,504	B6 x C3H	C3H	F	16387874
	Ath19	104,398,884	64,000,000 - 113,000,000	B6 x C3H	B6	F	17641228
	Ath19	108,713,906	89,927,907 - 108,713,906	129S x B6	129S or B6	F	14592847
12	Ascla5	14,347,013	?	B6 x FVB	B6	F	16380418
	Ath6	16,749,351	10,870,733 - 27,065,736	BKS x B6	BKS	F	10393218
	Ath18	35,210,068	27,570,810 - 35,210,068	129S x B6	B6 or 129S	F	14592847
	Ascla6	53,972,026	?	B6 x FVB	B6	M	16380418
	Ath21	102,995,424	27,570,810 - 113,361,762	129S x B6	B6	F	14592847

13	Ath25	42,683,303	40,000,000 - 52,000,000	DBA x AKR	DBA	M + F	16373612
	Ath32	92,906,389	86,000,0000 - 110,000,000	B6 x C3H	B6	F	17641228
14	Ath13	47,097,924	25,856,310 - 71,917,665	B6 x FVB	B6	M + F	11973313
	Suggestive	56,000,000	17,000,000 - 91,000,000	DBA x 129S	129S	M + F	24586312
	Ath36	73,000,000	21,000,000 - 106,000,000	DBA x 129S	129S	M	24586312
15	Ath22	20,464,816	10,000,000 - 29,000,000	DBA x AKR	DBA	F	16373612
	Ath33	71,152,437	67,000,000 - 78,000,000	B6 x C3H	C3H	M + F	17641228
	Ath33	76,258,033	52,153,986 - 90,057,489	B6 x C3H	C3H	M	23938286
16	Suggestive	31,984,214	?	B6 x FVB	?	M + F	11973313
17	Ath34	27,579,594	22,000,000 - 28,000,000	B6 x C3H	B6	M + F	17641228
	Ath26	36,360,361	33,000,000 - 53,000,000	DBA x AKR	AKR	M + F	16373612
		43,089,482	43,089,482 - 61,294,637	BALB x SMJ	BALB	F	27736672
18	Ath27	40,324,813	38,000,000 - 56,000,000	DBA x AKR	Heterosis	M	16373612
	Suggestive	53,000,000	17,000,000 - 68,000,000	DBA x 129S	DBA	M + F	24586312
19	Ath16	42,438,566	35,317,981 - 47,152,941	B6 x FVB	FVB	M + F	11973313

[a] Published name of QTL. QTLs found to be suggestive in the original publication are labeled as "Suggestive".
[b] Published peak of QTL in megabases (Mb). Peak location was determined as the physical location of the marker closest to the peak.
[c] Published 95% confidence interval (CI) in Mb. CI location in Mb was determined as the location of the markers closest to the edges of the reported CI in centimorgans (cM). QTLs that do not have a published CI are denoted with "?".
[d] Allele linked to larger aortic lesion size. QTLs that do not have a published high allele are denoted with "?".

1.3 Going from QTL to gene

These linkage QTLs are large, often spanning 10 megabases (Mb) or more, and finding the causal genes remains a significant challenge. Multiple strategies have been developed in other complex traits to try and narrow down these windows to a more manageable set of genes. These strategies include human-mouse comparative genomics, combined cross analysis, haplotype analysis, gene expression analysis, in silico mapping, and fine mapping[70,92,93]. Therefore when trying to find candidate genes for aortic lesion size QTLs, we apply these successful methods. Each of these methods has their strengths and weaknesses, so a combination of multiple methods is the most reliable way to ensure all the candidate genes are found for a locus. The three methods that have been the most successful for plasma lipids are comparative genomics, haplotype analysis, and gene expression analysis. Thus the combination of these three provides the highest chance of discovering novel influences on atherosclerosis.

1.3.1 Comparative Genomics Between Mouse and Human

The first method used involves human-mouse comparative genomics, or the identification of overlapping loci between mouse QTLs and human GWAS data. A large percentage of the genome is conserved between humans and mice, particularly at loci containing genes. The assumption of this analysis is that when mouse linkage QTLs and GWAS associations both appear in the same conserved region, there is a shared genetic component affecting both mice and humans[70]. Therefore, using bioinformatics data to find candidate genes in one species could also help discover candidate genes in the other[94]. For atherosclerosis, comparative genomics between the aortic lesion QTLs in mice and the syntenic regions in humans using GWAS data for CAD[61] uncovers a large number of suggestive or significant GWAS loci that overlap with significant or suggestive QTLs (**Table 1.2**). Of the significant aortic QTLs, 53 (44.5%) contained a GWAS association that could be

considered potentially suggestive (-log(p) > 5), and 29 (24.4%) of these contained a significant GWAS association. These significant GWAS associations account for 24 (40%) of the recorded genes published from the meta-analysis. This strategy has the potential to provide novel insight into QTLs or GWAS loci whose causal genes have previously gone undiscovered. Moreover, regions in which a significant linkage QTL in mice overlaps a suggestive association in human GWAS for CAD can provide a way to discern whether suggestive GWAS loci have potential phenotypical significance. This list compiled here can be used by future studies to help find things previously undiscovered in mice or humans and should be the first step in generating candidate genes for aortic QTL loci.

Table 1.3: *Overlap between published mouse aortic QTLs and human coronary artery disease (CAD) GWAS*

Chr	QTL	Peak Mouse (Mb)	CI Mouse (Mb)	CI Human Syntenic Chromosome	CI Human Syntenic Bp (Mb)	GWAS Hit (-log(p) >5)	Called GWAS Gene (-log(p) >8)	Cross	High Allele	MF	PMID
1	Suggestive	19,292,010	1 - 37,906,000	Chr2	97-99	-		BALB x SMJ	BALB	F	27736672
					128-132	-					
				Chr6	49-52	-					
					56-57	-					
					62-64	-					
					69-73	-					
				Chr8	50-56	-					
					67-75	-					
	Suggestive	20,871,663	20,871,663 - 68,723,270					B6 x FVB	?	M + F	11973313
	Suggestive	59,000,000	6,000,000 - 138,000,000					DBA x 129S	DBA	M + F	24586312
	Suggestive	73,000,000	56,000,000 - 90,000,000					DBA x 129S	DBA	F	24586312
	Ath30	75,956,444	72,000,000 - 80,000,000	Chr2	216-224	+		C3H x B6	B6	F	17641228
	Ath43	139,000,000	121,000,000 - 169,000,000	Chr1	150.2	-		DBA x 129S	DBA	F	24586312
					164-186	-					
					190-193	-					
					196-207	-					

				Chr2	114.4-114.6	-					
					133-137	-					
	Ath1	155,924,281	135,000,000 - 161,000,000	Chr1	150.2	-		C3H x B6	B6	M + F	17641228
					164-186	-					
					190-193	-					
					196-202	+					
	Ath44	158,300,000	153,000,000 - 168,000,000	Chr1	150.2	-		DBA x 129S	DBA	M + F	24586312
					164-183	-					
	Suggestive	171,000,000	162,000,000 - 183,000,000					DBA x 129S	DBA	M	24586312
		177,269,841	162,250,311 - 177,269,841	Chr1	145.5	-		BALB x SMJ	SMJ	F	27736672
					158-171	+					
					240-243	-					
	Ath9	177,285,317	157,361,794 - 192,529,115	Chr1	145.5	-		B6 x FVB	FVB	M	11973313
					150.2	-					
					158-178	+					
					210-227	*	MIA3				
					241-246	-					
	Ath1	183,092,234	172,000,000 - 193,241,301	Chr1	145.5	-		BALB x B6	B6	F	22294616
					158-160	+					
					209-227	*	MIA3				
					240-246	-					
2	Ath47	34,400,000	3,886,211 - 84,817,798	Chr2	113-114	-		BALB x B6	B6	F	21252064
					138-140	-					
					143.8	-					
					145.1	-					
					148-188	+					
				Chr9	123-140.13	+ , *	ABO				
				Chr10	6-8Mb	-					
					11-18Mb	-					
					20-27	-					
				Chr11	57-58	-					
	Ath39	38,239,859	25,000,000 - 47,000,000	Chr2	141	-		BALB x B6	B6	M	21252064

				145.1	-					
			Chr9	123-140.13	+ , *	ABO				
Ath41	71,964,058	45,000,000 - 130,000,000	Chr2	87.1	-		BALB x B6	B6	F	22294616
				95-97	-					
				110-114	-					
				145.1	-					
				148-184	+					
				187-189	-					
			Chr11	27-28	-					
				32-37	-					
				43-49	-					
				55-58	-					
			Chr15	32-52	-					
			Chr20	1.9Mb	-					
Ath46	139,000,000	106,000,000 - 150,000,000	Chr1	203.8	-		DBA x 129S	DBA	M	24586312
			Chr2	87.1	-					
				95-97	-					
				110-114	-					
			Chr11	27-28	-					
			Chr15	32-52	-					
			Chr20	1.8-2	-					
				2.6-24.5	+					
Athla1	148,680,255	115,000,000 - 150,000,000	Chr2	87.1	-		PERA x B6	B6	M + F	15514201
				95-97	-					
				110-114	-					
			Chr15	36-52	-					
			Chr20	1.8-2	-					
				2.6-24.5	+					
Ath45	162,000,000	157,000,000 - 165,000,000	Chr20	34-45	+		DBA x 129S	DBA	M + F	24586312
Ath35	172,000,000	139,000,000 - 173,000,000	Chr3	175.4	-		DBA x 129S	DBA	M	24586312
			Chr20	.2-1.4	-					
				13.3-14.3	-					
				16.23-25.4	+					
				30-56	+					
Ath28	178,800,659	162,355,057 - 181,873,487	Chr3	175.4	-		B6 x C3H	C3H	M	23938286
			Chr16	0.097	-					

				Chr19	11.72-12.39	-					
				Chr20	40.7-62.89	+					
	Ath28	179,574,406	174,000,000 - 179,800,000	Chr20	57-60	-		DBA x AKR	DBA	M	16373612
3	Ath23	66,450,410	54,000,000 - 90,000,000	Chr1	154-157	*	IL6R	DBA x AKR	AKR	F	16373612
				Chr3	149-162	+					
					164.9	-					
					167-168	-					
				Chr4	150-161	+ , *	GUCY1A3				
				Chr13	36-38.5	-					
				ChrX	0.6	-					
	Ascla4	115,830,548	?					B6 x FVB	B6	F	16380418
	Ascla3	148,143,886	?					B6 x FVB	B6	F	16380418
4	Ath8	55,251,078	45,230,241 - 150,592,401	Chr1	8.41-67.47	+, + * , *	PCSK9; PAP2B	B6 x C3H	B6	M	23938286
				Chr9	13	-					
					15-27	*	9p21				
					86	-					
					100-105	-					
					107-123.37	-					
	Ath8	59,552,657	45,658,566 - 105,299,798	Chr1	57-67	*	PAP2B	NZB x SMJ	SMJ	F	15530921
				Chr9	13	-					
					15-27	*	9p21				
					86	-					
					100-105	-					
					107-123.37	-					
	Ath4	86,450,248	9,702,715 - 107,112,550	Chr1	54.6-67.5	+ , *, *	PCSK9; PAP2B	B6 x C3H	B6	F	17916774
				Chr6	87.79-90.6	-					
					96.65-100.05	+					
				Chr8	87.38-97.15	-					
				Chr9	13	-					
					15-27	*	9p21				
					32.4-37.91	-					
					86	-					
					100.68-123.36	+					

		119,721,657	63,898,275 - 129,953,782	Chr1	32.37-67.46	+, + *, *	PCSK9; PAP2B	BALB x SMJ	Heterosis	F	27736672
				Chr9	12.77-27.1	*	9p21				
					86	-					
					117.78-123.36	-					
	Athsq1	129,572,122	110,000,000 - 136,927,309	Chr1	22.97-50.57	+		BALB x B6	BALB	F	22294616
	Athsq1	149,926,199	137,890,684 - 151,570,093	Chr1	6.84-12.8	-		B6 x MOLF	MOLF	F	11438740
					15.48-21.13	-					
	Athsq3	151,077,699	145,000,000 - 153,000,000	Chr1	5.92-12.29	-		B6 x C3H	B6	F	17641228
5		61,506,687	44,971,670 - 69,163,000	Chr1	64.05	-		BALB x SMJ	SMJ	F	27736672
				Chr4	17.51- 17.88	-					
					24.52-26.86	-					
					30.72	-					
					36.06-42.89	+					
		77,028,226	65,000,000 - 92,000,000	Chr4	39.06-48.88	-		B6 x C3H	B6	M + F	17641228
					52.70-57.89	*	REST-NOA1				
					62.36-76.40	-					
	Ath42	104,351,018	99,690,738 - 124,829,348	Chr1	89.99-93.81	-		BALB x B6	B6	F	22294616
				Chr4	.49-1.1	-					
					83.40-88.72	-					
				Chr5	138.94	-					
				Chr12	108.91-124.24	*, +	SH2B3				
					132.37-133.41	-					
				Chr22	25.33-28.37	-					
	Suggestive	108,384,913	69,163,000 - 120,578,000	Chr1	89.99-93.81	-		BALB x SMJ	SMJ	F	27736672
				Chr4	.49-1.1	-					
					44.62-48.88	-					
					52.70-57.89	*	REST-NOA1				
					62.36	-					
					65.14	-					
					68.42-88.72	+, +					
				Chr5	138.94	-					
				Chr12	108.91-124.24	+					

					132.37-133.41	-					
				Chr22	25.33-28.37						
	Ath24	146,583,524	130,000,000 - 149,000,000	Chr7	.19-6.68	-		DBA x AKR	DBA	M + F	16373612
					66.23-66.45	-					
					69.06-76.1	-					
					97.61-102.22	-					
				Chr13	27.21-30.78	+					
6	Ath37	124,468,231	116,665,031 - 128,519,025	Chr2	132.23	-		CAST x B6	B6	M + F	16624897
				Chr10	43.57-45.45	*	Cxcl12				
				Chr12	.17-9.07	-					
				Chr22	17.59-18.27	-					
	Artles	133,853,029	107,795,846 - 145,656,013	Chr2	130.94	-		CAST x B6	B6	M + F	11463718
					132.23	-					
				Chr3	4.53-12.87	-					
					129.15-129.27	-					
				Chr10	38.12	-					
					43.57-51.82	*	Cxcl12				
				Chr12	.17-25.35	+					
					118.57	-					
				Chr22	17.59-18.27	-					
	Athsq2	134,083,123	121,115,387 - 145,656,013	Chr2	130.94	-		B6 x MOLF	B6	M + F	11438740
					132.23	-					
				Chr12	.175-.299	-					
					3.31-25.34	+					
					118.57	-					
				Chr22	18.27	-					
	Ath38	134,184,819	128,519,025 - 136,316,178	Chr12	9.38-13.71	-		CAST x B6	B6	M + F	16624897
					118.57	-					
7	Aorls2	49,744,933	?					B6 x DBA	DBA	M + F	12925895
	Ath31	74,629,649	50,000,000 - 100,000,000	Chr1	28.42	-		B6 x C3H	B6	F	17641228
					156.69	-					
					247.83-248.52	-					
				Chr2	71.33	-					
				Chr11	20.69-22.84	-					

					24.51	-					
					49.16	-					
					74.45-78.36	+					
					82.44-89.05	-					
				Chr15	22.89-32.32	-					
					80.41-94.84	*, *	MFGE8-ABHD2; FURIN				
					98.5-101.84	-					
	Suggestive	78,000,000	55,000,000 - 97,000,000					DBA x 129S	129S	M + F	24586312
	Suggestive	78,000,000	54,000,000 - 91,000,000					DBA x 129S	129S	F	24586312
8	Ath40	42,935,531	21,959,500 - 71,045,600	Chr4	164.24-178.35	-		BALB x B6	B6	M	21252064
					183.81-190.86	-					
				Chr8	8.17-20.10	+					
					29.19-42.99	-					
				Chr13	41.3	-					
					41.36	-					
					52.50-52.98	-					
				Chr19	12.49	-					
					17.83-19.62	+					
		89,351,811	57,763,542 - 89,351,811	Chr1	89.44	-		BALB x SMJ	SMJ	F	27736672
				Chr4	141.26-172.73	+, *, *	EDNRA; GUCY1A3				
				Chr8	18.07-20.10	+					
				Chr16	46.69-51.17	-					
				Chr19	11.72-19.75	+, +					
				Chr22	33.66-35.94	-					
	Suggestive	?	?					B6 x DBA	B6	M + F	12925895
9	Ath29	37,133,055	32,500,594 - 68,829,021	Chr11	107.46-128.62	-		BALB x B6	B6	F	22294616
				Chr15	51.34-51.74	-					
					60.78	-					
					62.14-78.91	*	SMAD3				
	Ath29	44,340,026	34,067,647 - 69,831,153	Chr11	107.46-126.29	-		B6 x C3H	B6	M + F	17916774
				Chr15	51.34-51.74	-					
					60.78	-					

					62.14-78.91	*	SMAD3				
	Ath29	74,700,643	68,829,021 - 80,091,050	Chr6	52.60-55.62	-		B6 x C3H	B6	F	16387874
					74.07-76.31	-					
				Chr15	51.97-60.79	-					
	Ath29	80,091,050	80,091,050 - 96,397,888	Chr3	141.59-147.13	-		B6 x C3H	B6	M	23938286
				Chr6	76.31-86.33	+					
				Chr15	78.28-80.25	*	ADAMTS7				
	Ath29	83,693,797	45,000,000 - 89,000,000	Chr6	52.60-55.62	-		B6 x C3H	B6	M + F	17641228
					74.07-86.32	+					
				Chr11	107.46-118.18	-					
				Chr15	51.34-78.11	*	SMAD3				
10		6,202,049	?					B6 x FVB	?	M	11973313
	Ascla2	18,806,860	?					B6 x FVB	FVB	M	16380418
	Ath11	20,407,588	6,202,049 - 50,788,902	Chr6	100.83-101.84	-		B6 x FVB	FVB	M + F	11973313
					105.17-116.95	-					
					123.31-139.69	*	TCF21				
					142.39-154.72	+					
				ChrY	21.15	-					
	Ath20	20,407,588	9,413,521 - 75,681,262	Chr2	109.06-110.37	-		129S x B6	129S	F	14592847
				Chr6	100.83-101.84	-					
					105.17-116.95	-					
					123.31-139.69	*	TCF21				
					142.39-147.52	+					
				Chr10	55.56	-					
					59.95-74.76	-					
				Chr22	18.76-25.01	*	POM121L9P-ADORA2A				
				ChrY	21.15	-					
	Ascla1	21,715,442	?					B6 x FVB	FVB	F	16380418
		28,450,065	26,797,957 - 67,821,343	Chr2	109.06-110.37	-		BALB x SMJ	BALB	F	27736672
				Chr6	100.83-101.84	-					
					105.17-129.98	-					
				Chr10	64.56-74.76	-					
				ChrY	21.15	-					

	Ath17	68,255,809	45,000,035 - 68,255,809	Chr2	109.06-110.37	-		129S x B6	129S	F	14592847
				Chr6	100.83-101.84	-					
					105.17-123.11	-					
				Chr10	63.66-74.76	-					
	Aorls1	82,655,287	?					B6 x DBA	DBA	M + F	12925895
		104,516,306	86,301,142 - 104,516,306	Chr12	56.66	-		BALB x SMJ	BALB	F	27736672
					85.25-104.32	*	ATP2B1				
				Chr22	32.89-33.20	-					
11		45,275,149	19,745,966 - 68,609,504					B6 x C3H	C3H	F	16387874
	Ath19	104,398,884	64,000,000 - 113,000,000	Chr3	186.5	-		B6 x C3H	B6	F	17641228
				Chr17	.062-20.35	+, +					
					25.62-70.11	+, +, *	BCAS3				
				ChrX	152.85	-					
	Ath19	108,713,906	89,927,907 - 108,713,906	Chr17	32.36-53.82	+, +		129S x B6	129S or B6	F	14592847
					60.45-66.03	-					
12	Ascla5	14,347,013	?					B6 x FVB	B6	F	16380418
	Ath6	16,749,351	10,870,733 - 27,065,736	Chr2	6.98-18.11	-		BKS x B6	BKS	F	10393218
	Ath18	35,210,068	27,570,810 - 35,210,068	Chr2	.21-3.71	-		129S x B6	B6 or 129S	F	14592847
				Chr7	17.83-19.73	*	HDAC9				
					105.24-107.56	+					
	Ascla6	53,972,026	?					B6 x FVB	B6	M	16380418
	Ath21	102,995,424	27,570,810 - 113,361,762	Chr2	.21-3.71	-		129S x B6	B6	F	14592847
				Chr7	12.61-19.73	*	HDAC9				
					105.24-112.12	+					
				Chr14	25.28-47.31	-					
					50.05-52.15	-					
					58.66-81.94	+					
					85.99-105.95	*	HHIPL1				

13	Ath25	42,683,303	40,000,000 - 52,000,000	Chr6	10.40-19.84	*	PHACTR1	DBA x AKR	DBA	M + F	16373612
				Chr9	91-96.87	-					
	Ath32	92,906,389	86,000,000 - 110,000,000	Chr5	58.26-85.91	-		B6 x C3H	B6	F	17641228
14	Ath13	47,097,924	25,856,310 - 71,917,665	Chr3	15.24-16.34	-		B6 x FVB	B6	M + F	11973313
					52.35-57.74	+					
				Chr8	9.91-12.17	-					
					21.55-29	+					
				Chr10	46.55-51.59	-					
					81.31-88.85	+					
				Chr13	19.77-26.62	-					
					49.34-52.23	-					
				Chr14	20.24-25.10	-					
					52.73-58.47	-					
	Suggestive	56,000,000	17,000,000 - 91,000,000					DBA x 129S	129S	M + F	24586312
	Ath36	73,000,000	21,000,000 - 106,000,000	Chr3	15.24-16.34	-		DBA x 129S	129S	M	24586312
					52.35-57.74	+					
				Chr8	9.91-12.17	-					
					21.55-29	+					
					39.6	-					
				Chr10	46.55-51.59	-					
					75.75-88.85	+					
				Chr13	19.77-26.62	-					
					41.50-53.61	+					
					58.20-61.99	-					
					66.87-80.91	-					
				Chr14	20.24-25.10	-					
					52.73-58.47	-					
15	Ath22	20,464,816	10,000,000 - 29,000,000	Chr5	13.69-26.88	-		DBA x AKR	DBA	F	16373612
					31.19-35.07	-					
	Ath33	71,152,437	67,000,000 - 78,000,000	Chr8	134.46-146.27	+		B6 x C3H	C3H	M + F	17641228
				Chr22	36-36.96	-					

	Ath33	76,258,033	52,153,986 - 90,057,489	Chr8	118.14-136.47	+		B6 x C3H	C3H	M	23938286
					139.14-146.28	+					
				Chr12	33.52	-					
				Chr22	36-51.18	-					
16	Suggestive	31,984,214	?					B6 x FVB	?	M + F	11973313
17	Ath34	27,579,594	22,000,000 - 28,000,000	Chr2	95.53	-		B6 x C3H	B6	M + F	17641228
				Chr5	172.06-172.66	-					
				Chr6	33.36-34.86	-					
				Chr16	.23-3.19	-					
	Ath26	36,360,361	33,000,000 - 53,000,000	Chr3	16.62-19.19	-		DBA x AKR	AKR	M + F	16373612
				Chr6	29.07-33.37	+, +					
					39.82-49.80	-					
				Chr19	8.36-8.81	-					
				Chr21	18.96	-					
		43,089,482	43,089,482 - 61,294,637	Chr2	108.6-108.91	-		BALB x SMJ	BALB	F	27736672
					124.78	-					
				Chr3	16.62-20.11	-					
				Chr6	39.82-46.97	-					
				Chr19	4.2-6.77	-					
				Chr21	18.96	-					
18	Ath27	40,324,813	38,000,000 - 56,000,000	Chr5	112.31-115.78	-		DBA x AKR	Heterosis	M	16373612
					118.17-123.97	-					
					137.2	-					
					141.23-147.45	-					
				Chr8	101.71	-					
	Suggestive	53,000,000	17,000,000 - 68,000,000					DBA x 129S	DBA	M + F	24586312
19	Ath16	42,438,566	35,317,981 - 47,152,941	Chr10	92.5-105.21	*	CYP17A1-CNNM2-NT5C2	B6 x FVB	FVB	M + F	11973313

The regions of the human genome corresponding to the mouse genomic regions within the 95% confidence interval of an aortic QTL were probed for suggestive or significant associations with coronary artery disease in human GWAS meta-analysis studies[61].
- = No GWAS association
+ = Suggestive association (-log(p) > 5)
** = Significant association (-log(p) > 8)*

1.3.2 Haplotype Analysis

The second common strategy for prioritizing candidate genes at QTL loci is through the use of haplotype analysis. This method is performed at loci where mouse linkage QTLs identified in separate intercrosses overlap. The assumption of haplotype analysis is that at a locus, inbred strains linked to higher amounts of the phenotype (in this case aortic lesion size) share a common ancestor (high alleles) while those linked to lower amounts of the phenotype share a different common ancestor (low alleles). Using this assumption, comparison of the genomic sequence between the two groups of inbred strains is used to find SNPs that are shared between the high allele strains and differ from the SNPs shared by the low allele strains. These sequence differences can cause potential changes in gene expression through changing an enhancer or promoter site, or protein sequence through nonsynonymous changes in amino acid sequence. By looking at these two types of SNP changes, one can highlight candidate genes for further studies.

This strategy has been implemented numerous times for HDL, Triglyceride, and aortic lesion QTLs with a high degree of success[70,92,95,96]. With the availability of high-density SNP mapping and whole-genome sequencing by the Sanger Mouse Genomes Project (https://www.sanger.ac.uk/sanger/Mouse_SnpViewer/rel-1505), the ability to find SNPs that could have a potential role at a QTL locus has significantly improved. Thus this method remains effective and should be considered the second step in generating candidate genes.

1.3.3 Gene Expression Changes Through eQTLs

The third primary way of prioritizing candidate genes for overlapping QTLs is through the discovery of SNPs linked to changes in gene expression. The improvements in RNA quantification methodologies have permitted fast, cheap, high-throughput generation of gene expression data from large numbers of mice. If gene expression data and phenotypic data are collected from a large intercross of mice, one can determine the correlation of gene expression with phenotypes of interest. With the inclusion of high-density SNP data, linkage QTLs linking specific SNPs to the gene expression of a gene can be found (eQTLs). This is repeated for each gene, and a comprehensive list of all SNPs influencing gene expression can be generated. This list also contains which genes each individual SNP on that list is affecting. Finally, QTL data is obtained from linkage analysis for the phenotype. The integration of the correlation of gene expression with the phenotype, the eQTLs, and the phenotype's QTLs allows for discovery of high-confidence candidate genes at the phenotype's QTL peaks. A SNP in this list is a SNP in the linkage QTL affecting a gene's expression, which is in turn affecting the phenotype.

An additional strategy involves weighted gene co-expression network analysis (WGCNA). This method can be performed to create modules of genes that are being co-expressed and provide additional insights into the pathways or clusters of genes that could be the most influential to the phenotype. This has been done for aortic lesion size and other phenotypes to successfully find new pathways involved in atherosclerosis and osteoporosis[91,97–99]. If a gene implicated through the integration of phenotype QTLs, eQTLs, and gene expression resides in one of these modules, it has an even higher likelihood of being a gene affecting the phenotype.

Gene expression correlation analysis, eQTL generation, and WGCNA have been done for aortic atherosclerosis on a large cohort of inbred and recombinant inbred strains of mice called the Hybrid Diversity Mouse Panel (HDMP)[91]. The mice were fed 16 weeks of

Western diet, aortic lesion areas were quantified (n= 4-5 mice per strain), and both aortic and hepatic gene expression data was generated. Correlations between aortic and hepatic gene expression and aortic lesion size were calculated. WGCNA was done to identify multiple modules significantly affecting atherosclerosis through either the aorta or the liver. This study provides a large repository of information with which to investigate other linkage QTLs both our lab and others have discovered. Through grouping of high and low alleles in a way akin to haplotype analysis, we can find gene expression differences in the aorta and liver between the two groups. We can then filter these genes to those with a significant correlation to aortic lesion size. The final filter determines whether there are any significant eQTLs in the aortic QTL's 95% confidence interval (CI) and if so, whether there are any sequence differences between the two groups at said SNPs. Through this strategy, we can significantly improve our ability to find gene expression-based candidate genes in a way that was previously not possible. This strategy should be the third strategy used when generating candidate genes for QTLs. This strategy, in combination with haplotype analysis and overlap with human GWAS for CAD/CVD, provides a hitherto unprecedented ability to uncover the genes influencing mouse aortic QTL loci whose candidate genes have remained elusive for years.

1.4 Aims

While studies on atherosclerosis have been ongoing for decades, cardiovascular-related death still remains the most significant cause of mortality the world over. Current treatment strategies leave much to be desired, and thus discovering novel influencers and targets for therapy of cardiovascular disease and its related cardiometabolic disorders remains a pressing matter. The best way to accomplish this task is to discover the genetic underpinnings of said diseases. Over the past decade, genetic studies in mice using linkage

analysis for aortic lesion size and in humans using GWAS for CAD/CVD have provided unprecedented understanding of how genetic variation influences both atherosclerotic plaque progression and other cardiometabolic disorders. Hundreds of significant loci have been identified in humans that have influences on CAD susceptibility, and over a hundred QTLs have been discovered in mice for aortic lesion size. However, discerning the casual genes underlying these genetic associations remains a difficult task. This is especially true in mice where QTLs have large confidence intervals. More recent advances in technology has allowed for high-throughput dense genotyping and gene expression generation in mice. These advances have vastly improved our abilities to prioritize candidate genes for aortic lesion QTLs. Despite the quantity of QTLs, a large amount of work remains uncovering the genes causing the observed linkage and a large amount of the heritable component of atherosclerosis remains undiscovered. We hypothesize that through generating novel linkage data using rarely used inbred strains, we will be able to discover novel insights into the heritable components of atherosclerosis and its associated cardiometabolic disorders. We plan to accomplish this through QTL analysis of aortic lesion size, carotid lesion size, hyperlipidemia, and type 2 diabetes. Furthermore, we hypothesize that through modern strategies utilizing the large amounts of publicly available data, we will be able to uncover novel genes influencing atherosclerosis. Finally, we hypothesize that by developing novel, more accurate ways to quantify abdominal fat deposits, we will be able to create a platform in which to improve our understanding of the genetics underlying obesity, the cardiometabolic disease with the highest co-occurrence with atherosclerosis. We aim to investigate these hypotheses in the following studies:

(1) In Chapter 2, we generate QTLs for plasma lipids and fasting plasma glucose in an intercross between a commonly used inbred strain of mice BALB/cJ, which has lower LDL, lower glucose, and higher HDL, and a rarely used inbred strain

SM/J, which has higher LDL, higher glucose, and lower HDL. From performing linkage analysis on plasma lipids and glucose from F2 mice, we find a significant QTL for fasting glucose, named Bglu17, that coincides with a significant QTL for HDL and a suggestive QTL for non-HDL cholesterol levels. Haplotype analysis reveals that "lipid genes" Sik3, Apoa1, and Apoc3 are probable candidates for Bglu17. In this chapter, we find that the colocalization of QTLs for multiple phenotypes and the sharing of potential candidate genes demonstrates genetic connections between dyslipidemia and type 2 diabetes.

(2) In Chapter 3, we use the same intercross as Chapter 2 to generate QTLs for carotid lesion size in order to find novel genetic influences on stroke. QTL analysis reveals eight loci for carotid lesion sizes. Combined cross-linkage analysis using data from this cross and two previous F2 crosses derived from BALB, C57BL/6J and C3H/HeJ strains, identifies five significant and nine suggestive QTLs. Of them, the QTL on chromosome 12 is the most prominent. Bioinformatic analysis prioritizes nine probable candidate genes for this QTL. In this chapter, we demonstrate the polygenic control of carotid atherosclerosis in mice.

(3) In Chapter 4, we use the same intercross as Chapters 2 and 3 to generate QTLs for aortic lesion size in order to find novel genetic influences on aortic atherosclerosis. A significant locus, named *Ath49,* for atherosclerosis maps to the H2 complex [mouse major histocompatibility complex (MHC)] on chromosome 17. Bioinformatics-based analysis identifies a single likely candidate gene *Mep1α*. In this chapter, we identify the MHC as a major genetic determinant of atherosclerosis, highlighting the importance of inflammation in atherogenesis.

(4) In Chapter 5, we further investigate *Mep1α* as a novel influencer of atherosclerosis by generating *Mep1α*$^{-/-}$.*Apoe*$^{-/-}$ double knockout mice. Double knockout mice show decreased lesion size, changes in plaque stability markers, and changes in lesion cell content. In this chapter, we discover that *Mep1α* is a novel gene negatively influencing atherosclerosis.

(5) In Chapter 6, we present a novel machine learning-based methodology for quantifying abdominal subcutaneous and visceral fat volume in mice using magnetic resonance imaging (MRI). Analysis with this automated method of C57BL/6 mice and chromosome 9 congenic mice, in which a small region of chromosome 9 in C57BL/6 mice was replaced with the sequence from C3H/HeJ mice, shows significant reductions in the volumes of visceral and subcutaneous fat but not non-fat tissues. In this chapter, we show the accuracy of deep learning in quantification of abdominal fat and its significance in determining body weight.

(6) In Chapter 7, we adapt the machine learning-based methodology for quantifying abdominal subcutaneous and visceral fat volume in mice for use with human computed tomography (CT). Our data demonstrates the accuracy and efficiency of deep learning in quantifying abdominal fat, particularly in the central abdominal region containing more fat. Manual segmentation of total and subcutaneous fat on 271 CT slices took 22 hours while automated segmentation took 5 min. In this chapter, we provide a strategy for feasibly quantifying fat volume in humans and using a phenotype more accurate than the current metric of BMI.

Ultimately, this work contributes to our understanding of the genetic architecture of atherosclerosis, both in the aorta and the carotid, in addition to the other cardiometabolic

disorders influencing it. Moreover we have outlined a strategy here utilizing publicly available data for prioritizing candidate genes at overlapping QTL loci that can successfully discover novel genes involved in atherosclerosis. Finally, we provide new tools to study a more accurate phenotype for abdominal fat in both mice and humans in order to enhance our understanding of the pathogenesis of obesity.

Chapter 2

Genetic linkage of hyperglycemia and dyslipidemia in an intercross between BALB-cJ and SM-J Apoe-deficient mouse strains

Quian Wang*, Andrew T. Grainger*, Ani Manichaikul, Emily Farber, Suna Onengut-Gumuscu and Weibin Shi

*Co-First Authors

Published in: Wang Q, Grainger AT, Manichaikul A, Farber E, Onengut-Gumuscu S, Shi W. Genetic linkage of hyperglycemia and dyslipidemia in an intercross between BALB/cJ and SM/J Apoe-deficient mouse strains. *BMC Genet*. 2015;16:133. doi:10.1186/s12863-015-0292-y

2.1 Abstract

Individuals with dyslipidemia often develop type 2 diabetes, and diabetic patients often have dyslipidemia. It remains to be determined whether there are genetic connections between the 2 disorders. A female F2 cohort, generated from BALB/cJ (BALB) and SM/J (SM) Apoe-deficient (*Apoe*$^{-/-}$) strains, was started on a Western diet at 6 weeks of age and maintained on the diet for 12 weeks. Fasting plasma glucose and lipid levels were measured before and after 12 weeks of Western diet. 144 genetic markers across the entire genome were used for quantitative trait locus (QTL) analysis. One significant QTL on chromosome 9, named Bglu17 [26.4 cM, logarithm of odds ratio (LOD): 5.4], and 3 suggestive QTLs were identified for fasting glucose levels. The suggestive QTL near the proximal end of chromosome 9 (2.4 cM, LOD: 3.12) was replicated at both time points and named Bglu16. Bglu17 coincided with a significant QTL for HDL (high-density lipoprotein) and a suggestive QTL for non-HDL cholesterol levels. Plasma glucose levels were inversely correlated with HDL but positively correlated with non-HDL cholesterol levels in F2 mice on either chow or Western diet. A significant correlation between fasting glucose and triglyceride levels was also observed on the Western diet. Haplotype analysis revealed that "lipid genes" Sik3, Apoa1, and Apoc3 were probable candidates for Bglu17. In this study, we have identified multiple QTLs for fasting glucose and lipid levels. The co-localization of QTLs for both phenotypes and the sharing of potential candidate genes demonstrate genetic connections between dyslipidemia and type 2 diabetes.

2.2 Introduction

Individuals with dyslipidemia have an increased risk of developing type 2 diabetes (T2D), and diabetic patients often have dyslipidemia, which includes elevations in plasma triglyceride and low-density lipoprotein (LDL) cholesterol levels and reductions in high-

density lipoprotein (HDL) cholesterol levels[100]. Part of the increased diabetic risk associated with dyslipidemia is due to genetic variations that influence both lipoprotein homeostasis and the development of T2D. Indeed, a few rare gene mutations result in both dyslipidemia and T2D, which include ABCA1[101], LIPE[102], LPL[103], and LRP6[104]. Genome-wide association studies (GWAS) have identified >150 loci associated to variation in plasma lipids[105,106] and >70 loci associated with T2D, fasting plasma glucose, glycated hemoglobin (HbA1c), or insulin resistance[107–109]. Nearly a dozen of the loci detected are associated with both lipid and T2D-related traits at the genome-wide significance level, including GCKR, FADS1, IRS1, KLF14, and HFE (http://www.genome.gov/GWAStudies/). Surprisingly, half of them have shown opposite allelic effect on dyslipidemia and glucose levels[110], and this is in contrary to the positive correlations observed at the clinical level. Furthermore, it is challenging to establish causality between genetic variants and complex traits in humans due to small gene effects, complex genetic structure, and environmental influences.

A complementary approach to finding genetic components in human disease is to use animal models. Apolipoprotein E-deficient (*Apoe*$^{-/-}$) mice are a commonly used mouse model of dyslipidemia, with elevations in non-HDL cholesterol levels and reductions in HDL levels, even when fed a low fat chow diet[111,112]. High fat diet feeding aggravates dyslipidemia. Moreover, these mice develop all phases of atherosclerotic lesions seen in humans[71] and are extensively used for atherosclerosis research[113–116]. We have found that Apoe$^{-/-}$ mice with certain genetic backgrounds develop significant hyperglycemia and T2D when fed a Western-type diet but become resistant with some other genetic backgrounds[114,117,118]. BALB/cJ (BALB) and SM/J (SM) *Apoe*$^{-/-}$ mice exhibit differences in dyslipidemia and T2D-related phenotypes[114]. The objective of the present study was to explore potential genetic connections between dyslipidemia and T2D through quantitative

trait locus (QTL) analysis of a female cohort derived from an intercross between BALB-*Apoe*$^{-/-}$ and SM-*Apoe*$^{-/-}$ mice.

2.3 Results

2.3.1 Trait value distributions

Fasting plasma glucose and lipid levels of F_2 mice were measured before and after 12 weeks of Western diet. Values of fasting plasma glucose, non-HDL cholesterol and triglyceride levels of F_2 mice on both chow and Western diets and of HDL cholesterol level on the chow diet were normally or approximately normally distributed (**Fig. 2.1**). Values of square root-transformed HDL cholesterol levels on the Western diet showed a normal distribution. These data were then analyzed to search for QTLs affecting the traits. Loci with a genome-wide suggestive or significant *P* value are presented in **Table 2.1**.

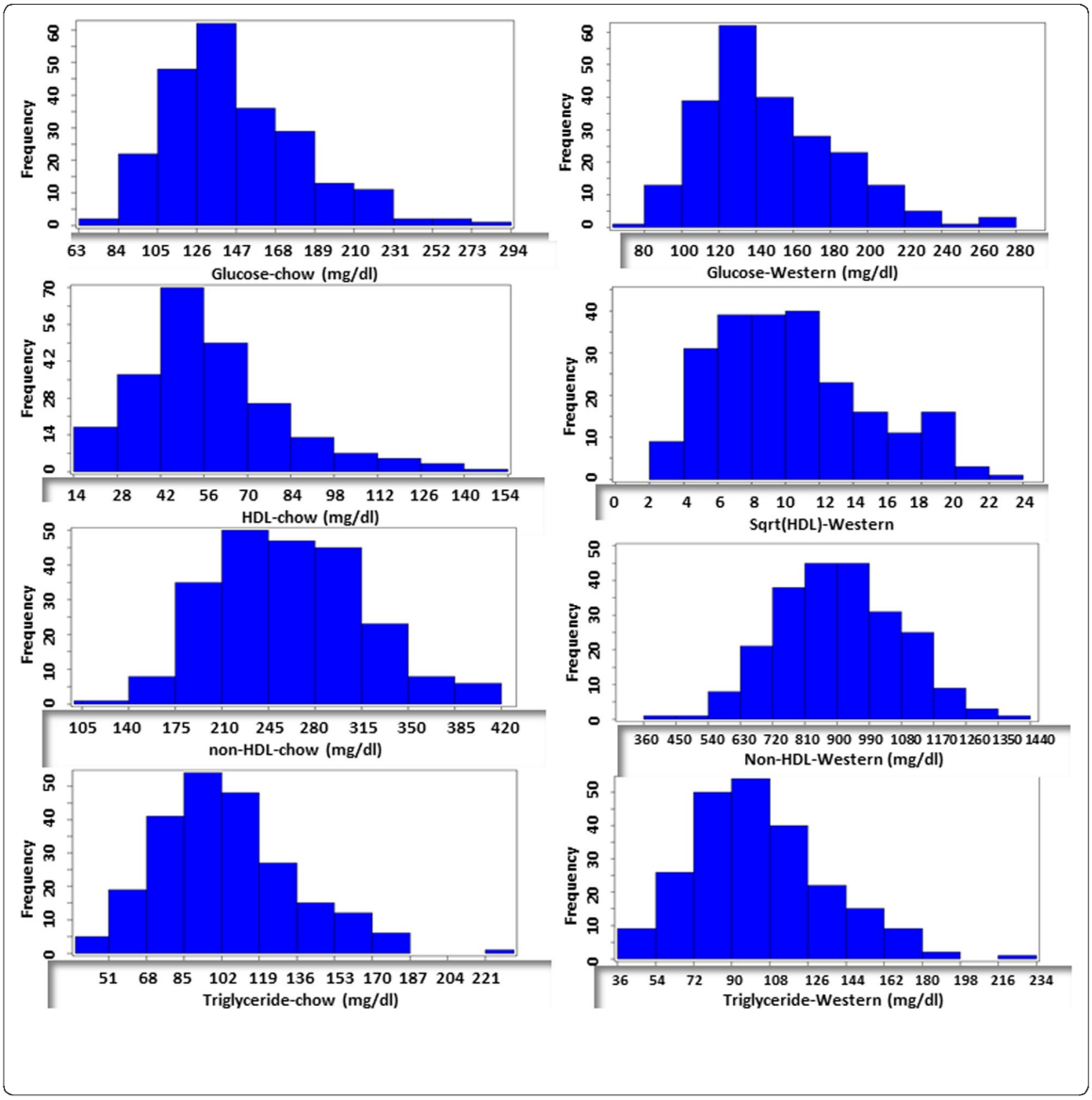

Figure 2.1 *The distributions of trait values for fasting plasma glucose, HDL, non-HDL cholesterol and triglyceride of 228 female F2 mice derived from an intercross between BALB-Apoe$^{-/-}$ and SM-Apoe$^{-/-}$ mice. Fasting blood was collected once before initiation of the Western diet (left panel) and once after 12 weeks on the Western diet (right panel). Graphs were created using a plotting function of J/qtl software.*

Table 2.1 *Significant and suggestive QTLs for plasma glucose and lipid levels in female F2 mice derived from BALB-Apoe$^{-/-}$ and SM-Apoe$^{-/-}$ mice.*

Locus	Chr	Trait	LOD[a]	*p*-value[b]	Peak (cM)	95% CI[c]	High allele	Mode of inheritence[d]
Bglu16	9	Glucose-C	2.214	0.549	2.37	0.37–30.37	B	Additive
Bglu13	5	Glucose-W	2.1.8	<0.63	67.4	45.4–80.03	S	Recessive
-	5	Glucose-W	3.198	0.097	101.24	29.40–101.24	S	Additive
Bglu16	9	Glucose-W	3.12	<0.63	2.37	0–10.37	B	Additive
Bglu17	**9**	**Glucose-W**	**5.425**	**0.001**	**26.37**	**16.37–40.37**	**B**	**Additive**
Hdlq5	**1**	**HDL-C**	**8.64**	**0.000**	**93.52**	**87.52–97.02**	**S**	**Additive**
Hdlcl1	7	HDL-C	2.668	0.321	61.33	35.57–89.57	S	Dominant
Hdlq17	**9**	**HDL-C**	**4.614**	**0.014**	**30.37**	**16.37–32.37**	**S**	**Additive**
Hdlq26	**10**	**HDL-C**	**2.181**	**0.591**	**61.22**	**25.03–61.22**	**S**	**Dominant**
Hdlq5	**1**	**HDL-W**	**13.944**	**0.000**	**87.52**	**83.52–93.52**	**S**	**Additive**
Hdlcl1	**7**	**HDL-W**	**3.658**	**0.034**	**85.57**	**77.57–89.67**	**S**	**Additive**
Hdlq17	**9**	**HDL-W**	**10.625**	**0.000**	**30.42**	**24.37–30.53**	**S**	**Additive**
Chol7	1	non-HDL-C	2.093	0.626	66.95	9.52–74.56	B	Recessive
Nhdlq15	2	non-HDL-C	2.56	0.321	23.86	8.73–38.73	B	Additive
Hdlq34	5	non-HDL-C	2.106	0.614	19.4	19.4–30.5	S	Additive
Pnhdlc1	6	non-HDL-C	2.489	0.362	57.53	1.53–77.53	B	Recessive
Nhdlq1	8	non-HDL-C	2.221	0.537	44.14	10.14–60.14	B	Additive
Nhdlq12	12	non-HDL-C	2.73	0.245	39.41	15.41–59.41	B	Additive
Nhdlq15	**2**	**non-HDL-W**	**4.79**	**0.002**	**31.80**	**22.73–40.73**	**B**	**Dominant**
Nhdlq11	9	non-HDL-W	2.136	0.585	32.37	0.37–75.33	B	Additive
-	11	non-HDL-W	2.332	0.436	1.99	1.99–17.99	B	Dominant
Nhdlq16	**16**	**non-HDL-W**	**3.99**	**0.011**	**46.66**	**35.43–46.66**	**S**	**Dominant**
Tgq11	2	Triglyceride-C	2.952	0.169	26.73	12.73–60.83	B	Additive
-	5	Triglyceride-C	2.759	0.234	80.03	73.40–93.40	S	Heterosis
Trglyd	1	Triglyceride-W	3.291	0.091	97.02	79.24–97.02	S	Additive

[a]LOD scores were obtained from genome-wide QTL analysis using J/qtl software. The significant LOD scores were highlighted in bold. The suggestive and significant LOD score thresholds were determined by 1,000 permutation tests for each trait. Suggestive and significant LOD scores were 2.116 and 3.429, respectively, for glucose on the chow diet; 2.056 and 3.569 for glucose on the Western diet; 2.127 and 3.725 for HDL cholesterol, 2.09 and 3.662 for non-HDL cholesterol, and 2.102 and 3.522 for triglyceride on the chow diet; 2.10 and 3.486 for HDL, 2.123 and 3.628 for non-HDL, and 2.123 and 3.628 for triglyceride on the Western diet.

[b]*The p-values reported represent the level of genome-wide significance.*

[c]*95 % Confidence interval in cM defined by a whole genome QTL scan.*

[d]*Mode of inheritance was defined according to allelic effect at the nearest marker of a QTL.*

2.3.2 Fasting glucose levels

A genome-wide scan for main effect QTLs revealed a suggestive QTL near the proximal end of Chr9 for fasting glucose when mice were fed the chow diet (2.37 cM, LOD: 2.21) (**Fig. 2.2 and Table 2.1**). As this QTL was replicated on the Western diet, it was named *Bglu16.* For fasting glucose levels on the Western diet, a significant QTL on Chr9 and 3 suggestive QTLs, including *Bglu16* on Chr9, were identified. The significant QTL on Chr9 peaked at 26.37 cM and had a LOD score of 5.425. It was named *Bglu17.* The suggestive QTL near the middle portion of Chr5 (67.4 cM, LOD 2.18) replicated *Bglu13*, initially mapped in a B6 x BALB *Apoe*$^{-/-}$ intercross[76] . The suggestive QTL on distal Chr5 (101.24 cM, LOD 3.198) was novel. The BALB allele conferred an increased glucose level for both of the Chr9 QTLs while the SM allele conferred increased glucose levels for the 2 Chr5 QTLs (**Table 2.2**).

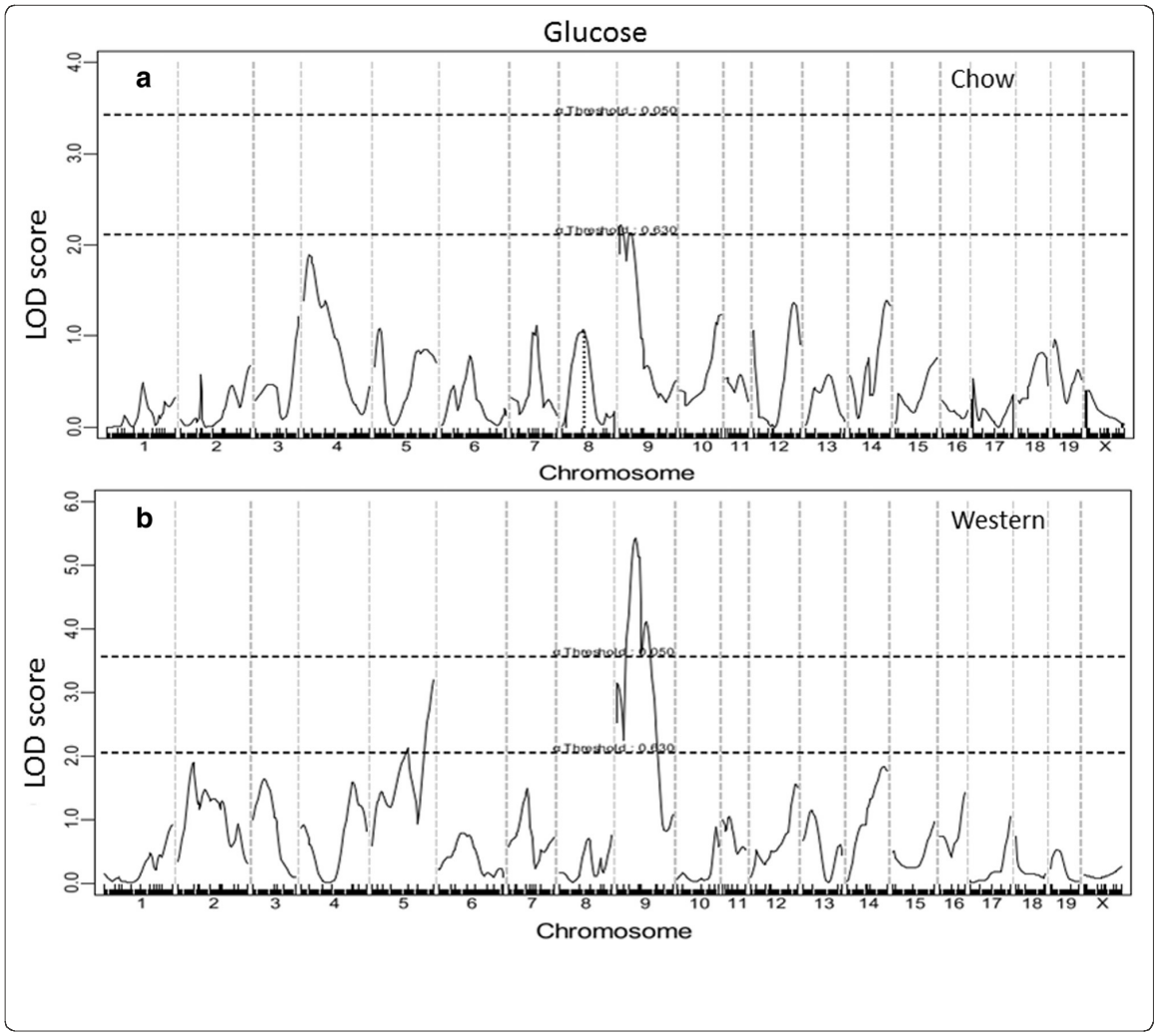

Figure 2.2 *Genome-wide scans to search for main effect loci influencing fasting plasma glucose levels of female F2 mice when fed a chow (a) or Western diet (b). Chromosomes 1 through X are represented numerically on the X-axis. The Y-axis represents the LOD score. Two horizontal dashed lines denote genome-wide empirical thresholds for suggestive (P = 0.63) and significant (P = 0.05) linkage.*

Table 2.2 *Allelic effects in different QTLs on plasma glucose and lipids of female F2 mice derived from BALB and SM Apoe$^{-/-}$ mice.*

Locus name	Chr	Trait	LOD	Peak (cM)	Closest marker	BB	SS	SB
Bglu16	9	Glucose-C	2.214	2.37	rs13480073	109.0±28.7 (*n*=44)	93.9±22.9 (*n*=43)	97.4±22.6 (*n*=141)
Bglu13	5	Glucose-W	2.1.8	67.4	rs3726547	144.4±30.6 (*n*=43)	153.5±40.4 (*n*=88)	142.9±35.3 (*n*=97)
-	5	Glucose-W	3.198	101.24	rs13478578	132.7±31.3 (*n*=51)	158.8±41.8 (*n*=63)	147.5±34.0 (*n*=113)
Bglu16	9	Glucose-W	3.12	2.37	rs13480073	165.3±40.9 (*n*=54)	138.3±30.0 (*n*=43)	144.4±35.7 (*n*=141)
Bglu17	**9**	**Glucose-W**	**5.425**	**26.37**	**CEL.9_4918363**	**168.0±39.8 (*n*=42)**	**134.7±26.5 (*n*=62)**	**146.1±37.1 (*n*=**
Hdlq5	**1**	**HDL-C**	**8.64**	**93.52**	**rs13476259**	**49.5±20.9 (*n*=60)**	**73.2±26.5 (*n*=62)**	**55.1±19.4 (*n*=106)**
Hdlcl1	7	HDL-C	2.668	61.33	rs3724711	49.0±20.0 (*n*=50)	58.3±21.0 (*n*=63)	62.9±25.5 (*n*=115)
Hdlq17	**9**	**HDL-C**	**4.614**	**30.37**	**CEL.9_4918363**	**49.5±15.9 (*n*=42)**	**69.4±26.8 (*n*=62)**	**56.2±22.5 (*n*=124)**
Hdlq26	10	HDL-C	2.181	61.22	rs3688351	50.7±19.2 (*n*=60)	59.4±21.9 (*n*=53)	62.4±25.8 (*n*=114)
Hdlq5	**1**	**sqrtHDL-W**	**13.944**	**87.52**	**rs3685643**	**66.6±50.2 (*n*=57)**	**201.1±118.8 (*n*=**	**117.0±97.1 (*n*=**
Hdlcl1	**7**	**sqrtHDL-W**	**3.658**	**85.57**	**rs6216320**	**95.5±87.6 (*n*=63)**	**173.3±126.7 (*n*=**	**122.4±98.4 (*n*=**
Hdlq17	**9**	**sqrtHDL-W**	**10.625**	**30.42**	**CEL.9_4918363**	**57.6±49.3 (*n*=42)**	**183.3±115.0 (*n*=**	**122.8±101.7 (*n*=**
Chol7	1	non-HDL-C	2.093	66.95	rs6354736	279.5±62.8 (*n*=56)	257.8±56.9 (*n*=57)	251.2±52.1 (*n*=114)
Nhdlq15	2	non-HDL-C	2.56	23.86	mCV23209429	273.9±56.1 (*n*=55)	238.0±47.0 (*n*=53)	262.7±59.0 (*n*=120)
Hdlq34	5	non-HDL-C	2.106	19.4	rs3658401	244.5±54.7 (*n*=63)	276.0±53.2 (*n*=61)	259.3±58.3 (*n*=104)
Pnhdlc1	6	non-HDL-C	2.489	57.53	rs13478909	279.6±51.3 (*n*=51)	252.0±65.2 (*n*=57)	254.8±53.5 (*n*=120)
Nhdlq1	8	non-HDL-C	2.221	44.14	D8Mit50	275.0±54.5 (*n*=60)	242.4±57.4 (*n*=57)	262.9±55.6 (*n*=96)
Nhdlq12	12	non-HDL-C	2.73	39.41	rs6195664	278.6±52.3 (*n*=62)	243.8±57.8 (*n*=59)	257.4±56.5 (*n*=107)
Nhdlq15	**2**	**non-HDL-W**	**4.79**	**31.8**	**rs13476507**	**954.1±156.0 (*n*=**	**806.9±158.2 (*n*=**	**915.6±166.1 (*n*=**
Nhdlq11	9	non-HDL-W	2.136	32.37	rs3709825	958.4±211.4 (*n*=42)	856.8±165.3 (*n*=62)	906.6±149.5 (*n*=
-	11	non-HDL-W	2.332	1.99	rs4222040	927.3±149.8 (*n*=67)	849.0±165.1 (*n*=69)	917.0±170.5 (*n*=85)
Nhdlq16	**16**	**non-HDL-W**	**3.99**	**46.66**	**rs3721202**	**820.2±152.7 (*n*=**	**931.9±146.4 (*n*=**	**928.4±174.6 (*n*=**
Tgq11	2	Triglyceride-C	2.952	26.73	mCV23209429	123.7±35.7 (*n*=55)	101.9±34.6 (*n*=53)	107.3±31.8 (*n*=120)
-	5	Triglyceride-C	2.759	80.03	gnf05.120.578	110.2±33.2 (*n*=43)	119.3±35.8 (*n*=88)	101.6±31.3 (*n*=97)
Trglyd	1	Triglyceride-W	3.291	97.02	rs13476259	94.0±28.6 (*n*=59)	115.3±33.0 (*n*=62)	100.7±30.8 (*n*=106)
Bglu16	9	Glucose-C	2.214	2.37	rs13480073	109.0±28.7 (*n*=44)	93.9±22.9 (*n*=43)	97.4±22.6 (*n*=141)
Bglu13	5	Glucose-W	2.1.8	67.4	rs3726547	144.4±30.6 (*n*=43)	153.5±40.4 (*n*=88)	142.9±35.3 (*n*=97)
-	5	Glucose-W	3.198	101.24	rs13478578	132.7±31.3 (*n*=51)	158.8±41.8 (*n*=63)	147.5±34.0 (*n*=113)
Bglu16	9	Glucose-W	3.12	2.37	rs13480073	165.3±40.9 (*n*=54)	138.3±30.0 (*n*=43)	144.4±35.7 (*n*=141)
Bglu17	**9**	**Glucose-W**	**5.425**	**26.37**	**CEL.9_4918363**	**168.0±39.8 (*n*=42)**	**134.7±26.5 (*n*=62)**	**146.1±37.1 (*n*=**
Hdlq5	**1**	**HDL-C**	**8.64**	**93.52**	**rs13476259**	**49.5±20.9 (*n*=60)**	**73.2±26.5 (*n*=62)**	**55.1±19.4 (*n*=106)**
Hdlcl1	7	HDL-C	2.668	61.33	rs3724711	49.0±20.0 (*n*=50)	58.3±21.0 (*n*=63)	62.9±25.5 (*n*=115)
Hdlq17	**9**	**HDL-C**	**4.614**	**30.37**	**CEL.9_4918363**	**49.5±15.9 (*n*=42)**	**69.4±26.8 (*n*=62)**	**56.2±22.5 (*n*=124)**
Hdlq26	10	HDL-C	2.181	61.22	rs3688351	50.7±19.2 (*n*=60)	59.4±21.9 (*n*=53)	62.4±25.8 (*n*=114)
Hdlq5	**1**	**sqrtHDL-W**	**13.944**	**87.52**	**r[55]s3685643**	**66.6±50.2 (*n*=57)**	**201.1±118.8 (*n*=**	**117.0±97.1 (*n*=**
Hdlcl1	**7**	**sqrtHDL-W**	**3.658**	**85.57**	**rs6216320**	**95.5±87.6 (*n*=63)**	**173.3±126.7 (*n*=**	**122.4±98.4 (*n*=**
Hdlq17	**9**	**sqrtHDL-W**	**10.625**	**30.42**	**CEL.9_4918363**	**57.6±49.3 (*n*=42)**	**183.3±115.0 (*n*=**	**122.8±101.7 (*n*=**
Chol7	1	non-HDL-C	2.093	66.95	rs6354736	279.5±62.8 (*n*=56)	257.8±56.9 (*n*=57)	251.2±52.1 (*n*=114)
Nhdlq15	2	non-HDL-C	2.56	23.86	mCV23209429	273.9±56.1 (*n*=55)	238.0±47.0 (*n*=53)	262.7±59.0 (*n*=120)
Hdlq34	5	non-HDL-C	2.106	19.4	rs3658401	244.5±54.7 (*n*=63)	276.0±53.2 (*n*=61)	259.3±58.3 (*n*=104)
Pnhdlc1	6	non-HDL-C	2.489	57.53	rs13478909	279.6±51.3 (*n*=51)	252.0±65.2 (*n*=57)	254.8±53.5 (*n*=120)

Nhdlq1	8	non-HDL-C	2.221	44.14	D8Mit50	275.0±54.5 (*n*=60)	242.4±57.4 (*n*=57)	262.9±55.6 (*n*=96)
Nhdlq12	12	non-HDL-C	2.73	39.41	rs6195664	278.6±52.3 (*n*=62)	243.8±57.8 (*n*=59)	257.4±56.5 (*n*=107)
Nhdlq15	**2**	**non-HDL-W**	**4.79**	**31.8**	**rs13476507**	**954.1±156.0 (*n*= 56)**	**806.9±158.2 (*n*= 47)**	**915.6±166.1 (*n*= 125)**
Nhdlq11	9	non-HDL-W	2.136	32.37	rs3709825	958.4 ± 211.4 (n = 42)	856.8 ± 165.3 (n = 62)	906.6 ± 149.5 (n = 124)
-	11	non-HDL-W	2.332	1.99	rs4222040	927.3 ± 149.8 (n = 67)	849.0 ± 165.1 (n = 69)	917.0 ± 170.5 (n = 85)
Nhdlq16	**16**	**non-HDL-W**	**3.99**	**46.66**	**rs3721202**	**820.2 ± 152.7 (n = 56)**	**931.9 ± 146.4 (n = 52)**	**928.4 ± 174.6 (n = 120)**
Tgq11	2	Triglyceride-C	2.952	26.73	mCV23209429	123.7 ± 35.7 (n = 55)	101.9 ± 34.6 (n = 53)	107.3 ± 31.8 (n = 120)
-	5	Triglyceride-C	2.759	80.03	gnf05.120.578	110.2 ± 33.2 (n = 43)	119.3 ± 35.8 (n = 88)	101.6 ± 31.3 (n = 97)
Trglyd	1	Triglyceride-W	3.291	97.02	rs13476259	94.0 ± 28.6 (n = 59)	115.3 ± 33.0 (n = 62)	100.7 ± 30.8 (n = 106)

Chr = chromosome, LOD = logarithm of odds, C = chow diet, W = Western diet, BB = homozygous BALB allele, SS = homozygous SM allele, SM = heterozygous allele.
Data are mean ± SD. The units for these measurements are mg/dL for plasma glucose or lipid levels. The number in the brackets represents the number of progeny with a specific genotype at a peak marker. The significant QTLs and their LOD scores were highlighted in bold.

2.3.3 Fasting lipid levels

Genome-wide scans for main effect QTLs showed that HDL, non-HDL cholesterol, and triglyceride levels were each controlled by multiple QTLs (**Figs. 2.3, 2.4, and 2.5; Table 2.1**). For HDL, 3 significant QTLs, located on Chr1, Chr7 and Chr9, and 1 suggestive QTL on Chr10, were identified. All 3 significant QTLs for HDL were detected when mice were fed either chow or Western diet, while the suggestive QTL on Chr10 was found when mice were on the chow diet. The significant QTL on Chr1 replicated *Hdlq5,* which had been mapped in numerous crosses[92]. The Chr7 QTL replicated *Hdlcl1,* initially mapped in (PERA/EiJ x B6-*Ldlr*)) x B6-*Ldlr* backcross[78]. The Chr9 QTL replicated *Hdlq17*, previously mapped in B6 x 129S1/SvImJ F_2 mice[89] . The suggestive QTL on Chr10 overlapped with

Hdlq26 mapped in a SM/J x NZB/BlNJ intercross[82]. For all 4 HDL QTLs, F_2 mice homozygous for the SS allele had higher HDL levels than those homozygous for the BB allele (**Table 2.2**).

For non-HDL cholesterol levels, 6 suggestive QTLs were detected when F_2 mice were fed the chow diet, and 2 significant and 2 suggestive QTLs were detected on the Western diet (**Fig. 2.4**). The 2 significant QTLs on Chr2 and Chr16 and the suggestive QTL on Chr11 were novel. The former 2 QTLs were named *Nhdlq15* and *Nhdlq16*, respectively. *Nhdlq15* peaked at 31.8 cM on Chr2 and affected non-HDL levels in a dominant mode from the BB allele while *Nhdlq16* peaked at 46.66 cM on Chr16 and affected non-HDL levels in a dominant mode from the SS allele. The rest replicated previously identified ones in other mouse crosses: The Chr1 QTL peaked at 66.95 cM, overlapping with *Chol7* mapped in an intercross of 129S1/SvImJ and CAST/Ei mice[119] . The Chr5 QTL overlapped with *Hdlq34* mapped in PERA/EiJ × I/LnJ and PERA/EiJ × DBA/2 J intercrosses[120]. The Chr6 QTL overlapped with *Pnhdlc1*, initially mapped in a B6 x CASA/Rk intercross and then replicated in B6 x C3H *Apoe*$^{-/-}$ F_2 mice[121,122]. The Chr8 QTL replicated *Nhdlq1*, initially mapped in B6 x 129S1/SvImJ F_2 mice[123]. The Chr9 QTL replicated *Nhdlq11*, initially mapped in B6 x C3H *Apoe*$^{-/-}$ F_2 mice[122]. The Chr12 QTL peaked at 44.14 cM, overlapping with *Nhdlq12* mapped in a B6 x C3H *Apoe*$^{-/-}$ F_2 intercross[122].

For triglyceride levels, 3 suggestive QTLs, located on Chr1, 2, and 5, respectively, were identified (**Fig. 2.5**). The Chr1 QTL peaked at 97 cM, 17 cM distal the *Apoa2* gene (80 cM). The Chr2 QTL replicated *Tgq11,* mapped in an intercross between DBA/1J and DBA/2J[124]. The Chr5 QTL was novel.

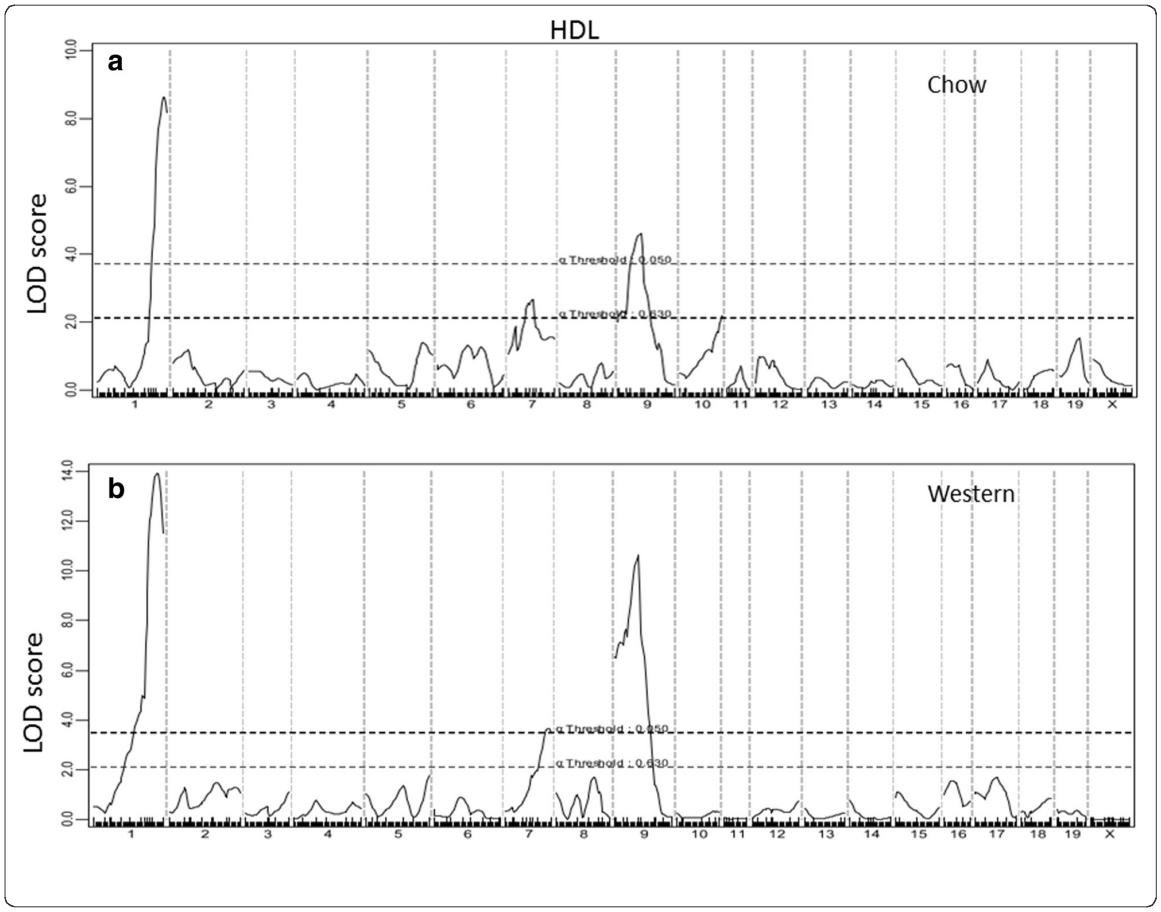

Figure 2.3 *Genome-wide scans to search for loci influencing HDL cholesterol levels of female F2 mice when fed a chow (a) or Western diet (b). Three significant loci on chromosomes 1, 7, and 9 and one suggestive locus on chromosome 10 were detected to affect HDL cholesterol levels of mice.*

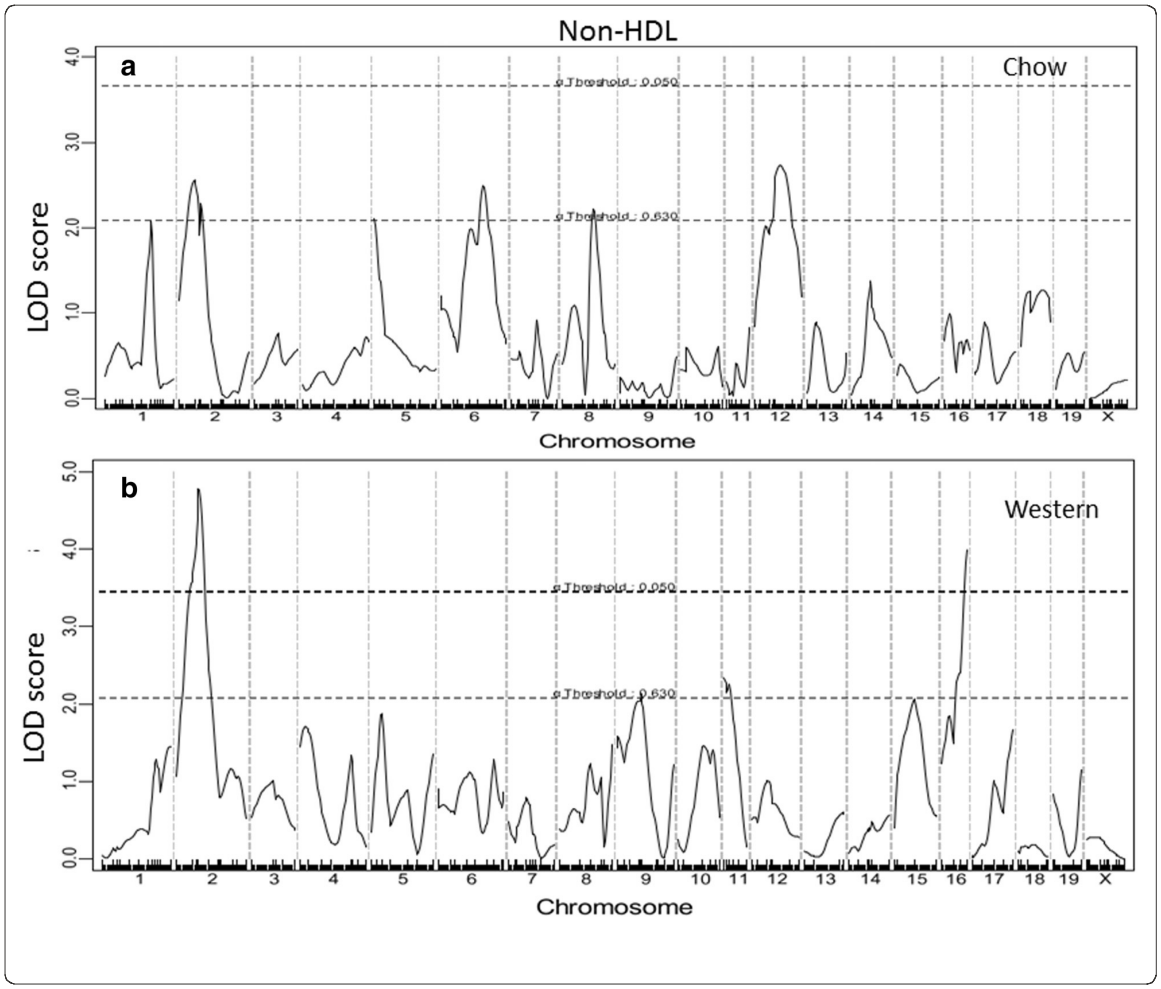

Figure 2.4 *Genome-wide scans to search for loci influencing non-HDL cholesterol levels of female F2 mice fed a chow (a) or Western diet (b). Two significant loci on chromosomes 2 and 16 were identified to affect non-HDL cholesterol levels of mice fed the Western diet.*

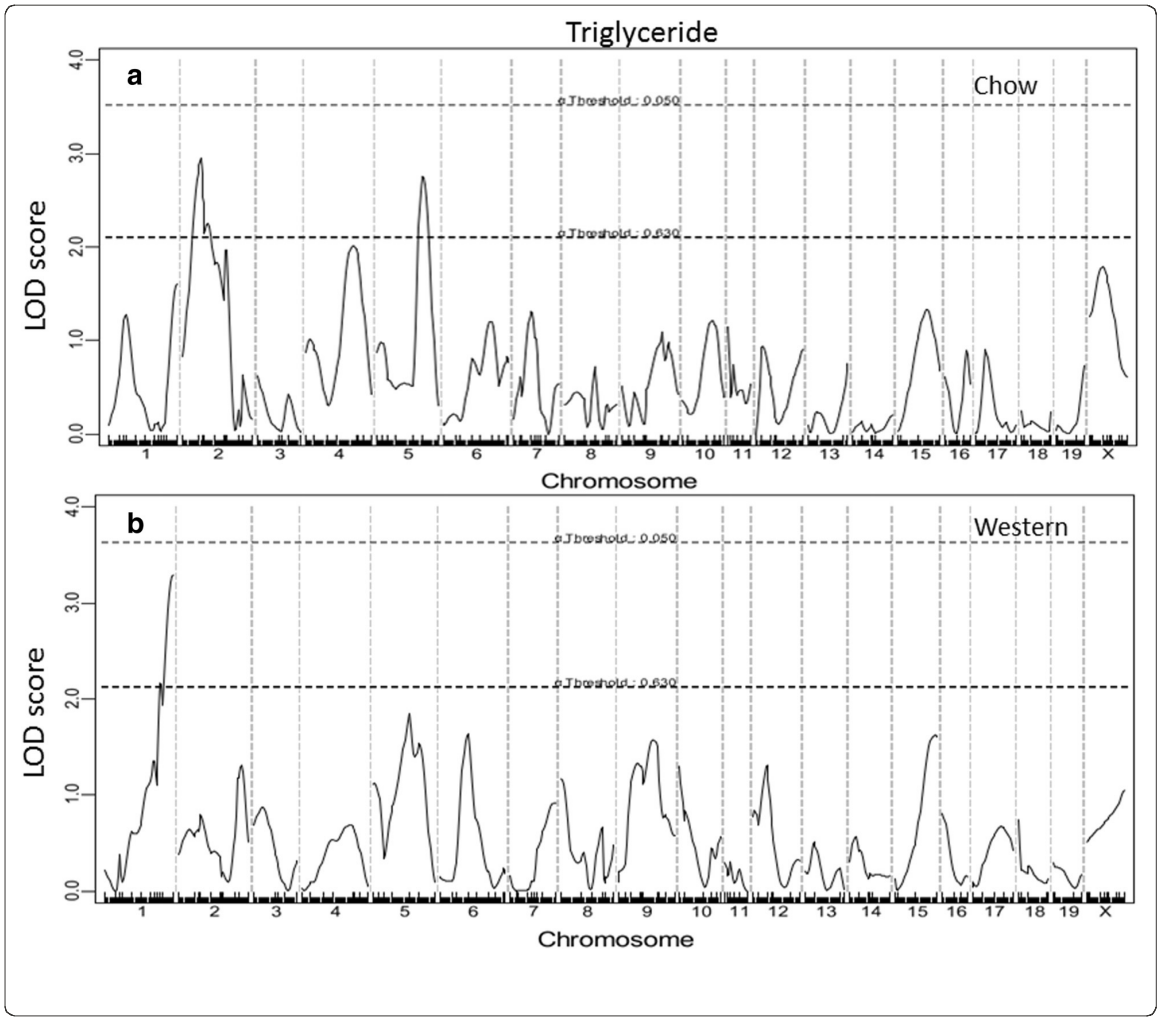

Figure 2.5 *Genome-wide scans to search for loci influencing triglyceride levels of female F2 mice fed a chow (a) or Western diet (b). Three suggestive loci were identified for triglyceride levels.*

2.3.4 Coincident QTLs for fasting glucose and lipids

LOD score plots for Chr9 showed that the QTL for fasting glucose (*Bglu17*) coincided precisely with the QTLs for HDL (*Hdlq17*) and non-HDL (*Nhdlq11*) in the confidence interval (**Fig. 2.6**). F2 mice homozygous for the BB allele exhibited

elevated levels of fasting glucose and non-HDL but decreased levels of HDL, compared to those homozygous for the SS allele (**Table 2.2**). These QTLs affected their respective trait values in an additive manner.

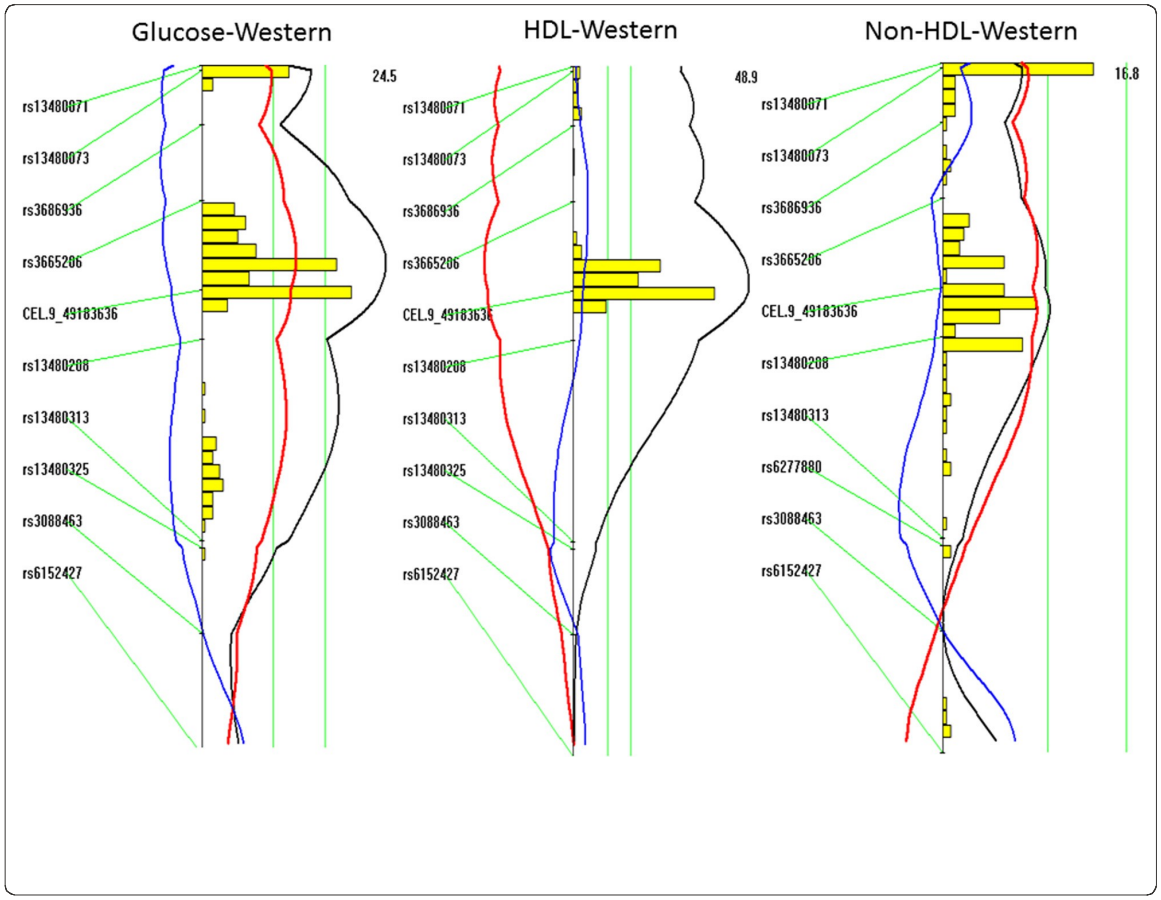

Figure 2.6 *LOD score plots for fasting glucose, HDL, and non-HDL cholesterol of F2 mice fed the Western diet on chromosome 9. Plots were created with the interval mapping function of Map Manager QTX. The histogram in the plot estimates the confidence interval for a QTL. Two green vertical lines represent genome-wide significance thresholds for suggestive or significant linkage (P = 0.63 and P = 0.05, respectively). Black plots reflect the LOD score calculated at 1-cM intervals, the red plot represents the effect of the BALB allele, and the blue plot represents the effect of the SM allele. If BALB represents the high allele, then the red plot will be to the right of the graph; otherwise, it will be to the left.*

2.3.5 Correlations between plasma glucose and lipid levels

The correlations of fasting glucose levels with plasma levels of HDL, non-HDL cholesterol, or triglyceride were analyzed with the F2 population (**Fig. 2.7**). A significant inverse correlation between fasting glucose and HDL cholesterol levels was observed when the mice were fed a chow (R = −0.220; P = 8.1E-4) or Western diet (R = −0.257; P = 8.5E-5). F2 mice with higher HDL cholesterol levels had lower fasting glucose levels. Conversely, significant positive correlations between fasting glucose and non-HDL cholesterol levels were observed when mice were fed either chow (R = 0.194; P = 3.31E-3) or Western diet (R = 0.558; P = 4.7E-20). F2 mice with higher non-HDL cholesterol levels also had higher fasting glucose levels, especially on the Western diet. A significant positive correlation between plasma levels of fasting glucose and triglyceride was observed when mice were fed the Western diet (R = 0.377; P = 3.9E-9) but not the chow diet (R = 0.065; P = 0.330).

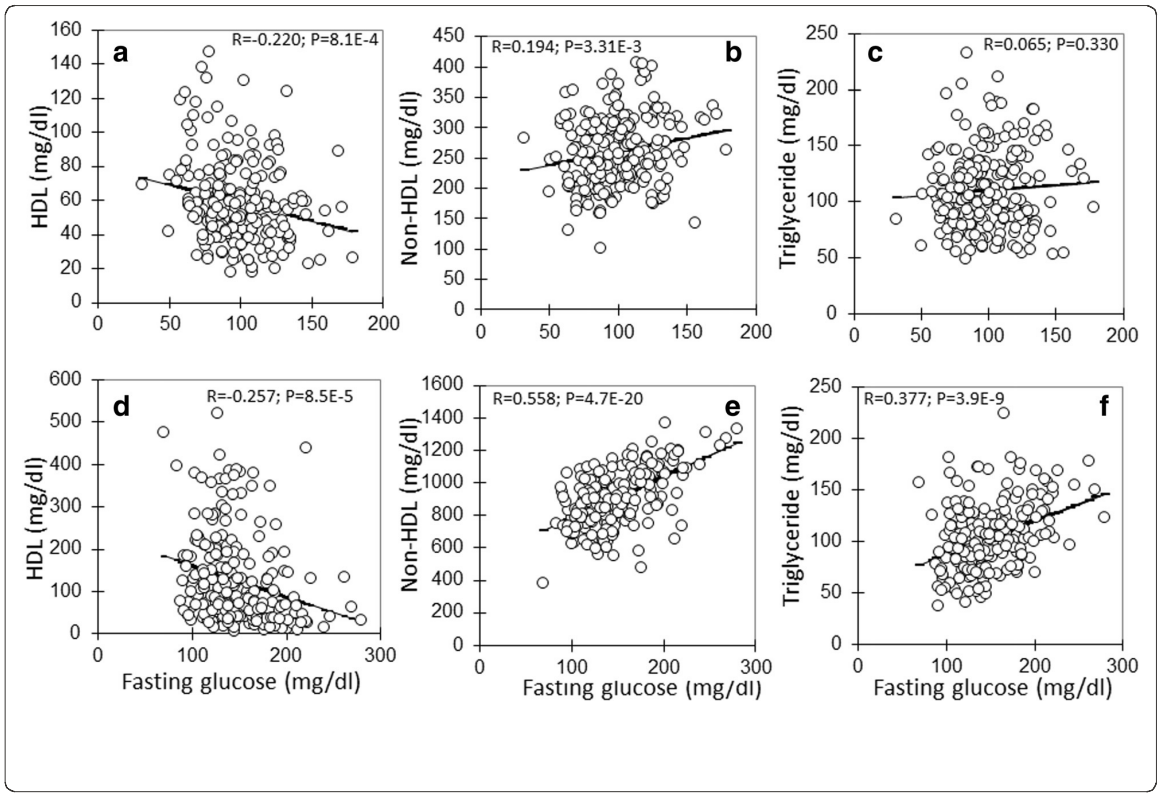

Figure 2.7 *Correlations of fasting plasma glucose levels with plasma levels of HDL, non-HDL cholesterol and triglyceride in the F2 population fed a chow (top row: a, b, c) or Western diet (bottom row: d, e, f). Each point represents values of an individual F2 mouse. The correlation coefficient (R) and significance (P) are shown.*

2.3.6 Prioritization of positional candidate genes for Chr9 coincident QTLs

Bglu17 on Chr9 has been mapped in 3 separate intercrosses, including previously reported C57BLKS x DBA/2 [125] and B6-*Apoe*$^{-/-}$ x BALB-*Apoe*$^{-/-}$ crosses[76]. *Hdlq17* on Chr9 has been mapped in multiple crosses, including B6 x 129, B6 x CAST/EiJ, B6-*Apoe*$^{-/-}$ x C3H-*Apoe*$^{-/-}$, and B6-*Apoe*$^{-/-}$ x BALB-*Apoe*$^{-/-}$ crosses[79,89,121,122,126–128]. We conducted haplotype analyses using Sanger SNP database to prioritize positional

candidate genes for both QTLs. Prioritized candidate genes for *Hdlq17* and candidate genes for *Bglu17* are shown in **S2.1**. Most candidates for *Hdlq17* are also candidate genes for *Bglu17*. These candidates contain one or more non-synonymous SNPs in the coding regions or SNPs in the upstream regulatory region that are shared by the high allele strains but are different from the low allele strains at the QTL. All candidate genes were further examined for associations with relevant human diseases using the NIH GWAS database (http://www.genome.gov/GWAStudies/). *Sik3, Apoa1,* and *Apoc3* have been shown to be associated with variations in total, HDL, LDL-cholesterol or triglyceride levels[105,106,129], and *Cadm1* with obesity-related traits[130].

2.4 Discussion

BALB and SM are two mouse strains that exhibit distinct differences in HDL, non-HDL cholesterol, and type 2 diabetes-related traits when deficient in *Apoe*[114]. BALB-*Apoe*$^{-/-}$ mice have higher HDL, lower non-HDL cholesterol, and lower glucose levels than SM-*Apoe*$^{-/-}$ mice when they are fed a Western diet. To identify the genetic factors responsible for these differences, we performed QTL analysis on a female cohort derived from an intercross between the two *Apoe*$^{-/-}$ strains. We have identified four loci contributing to fasting glucose levels, four loci contributing to HDL cholesterol levels, nine loci for non-HDL cholesterol levels, and three loci for triglyceride levels. Moreover, we have observed genetic connections between dyslipidemia and type 2 diabetes in that the QTL for fasting glucose is co-localized with the QTLs for HDL and non-HDL cholesterol on chromosome 9 and these coincident QTLs share a large fraction of potential candidate genes.

We identified a significant QTL on chromosome 9, peaked at 26 cM, which

affected fasting plasma glucose levels when mice were fed a chow or Western diet. We named it *Bglu17* to represent a novel locus regulating fasting glucose levels in the mouse. This locus is overlapping with a significant QTL (not named) for blood glucose levels on the intraperitoneal glucose tolerance test identified in a BKS-Cg-*Leprdb*$^{+/+m}$ x DBA/2 intercross and a suggested QTL identified in a B6-*Apoe*$^{-/-}$ x BALB-*Apoe*$^{-/-}$ intercross[76,125]. Interestingly, we found that *Bglu17* coincided precisely with *Hdlq17*, a QTL for HDL cholesterol levels, and *Nhdlq11*, a QTL for non-HDL cholesterol levels. The co-localization of two or more QTLs for different traits suggests that these traits are controlled either by the same gene(s) or closely linked but different individual genes. *Hdlq17* has been mapped in multiple crosses derived from inbred mouse strains whose genomes have been re-sequenced by Sanger, including B6, 129, BALB, C3H/HeJ, and CAST/EiJ[79,89,121,122,126–128]. *Nhdlq11* was previously mapped in a NZB/BINJ x SM/JF2 cross and a B6- *Apoe*$^{-/-}$ x C3H-*Apoe*$^{-/-}$ intercross[122,131]. To determine whether *Bglu17* and *Hdlq17* share the same underlying candidate genes, we performed haplotype analyses on those crosses that led to the identification of the QTLs. The number of shared genetic variants between *Bglu17* and *Hdlq17* was surprisingly high. Of them, *Sik3*, *Apoa1*, and *Apoc3* are located precisely underneath the linkage peak of *Bglu17* and *Hdlq17*, and they are also functional candidate genes of *Hdlq17*. Indeed, recent GWAS studies have associated these three genes with dyslipidemia or variations in HDL, LDL cholesterol, and triglyceride levels[105,129,132]. The finding in this study strongly suggests that one or more of these "lipid genes" might be the causal gene(s) of *Bglu17*, contributing to variation in fasting glucose levels. Although it is unknown how they affect glucose homeostasis, one probable effect path is through the influence on plasma lipid levels, which then predispose variation in glucose-related traits. The current observation on

the significant correlations of fasting glucose levels with HDL, non-HDL cholesterol, and triglyceride levels in this cross supports this speculation. Plasma lipid levels, especially non-HDL cholesterol, of the F2 mice were significantly elevated on the Western diet, so were the fasting glucose levels. When fed the Western diet, *Apoe*$^{-/-}$ mice display a rapid rise in non-HDL cholesterol levels, often reaching their peak within a couple of weeks (unpublished data), whereas their blood glucose levels rise more slowly and gradually within 12 weeks[133,134]. This difference in onset suggests a causal role for plasma lipids in the rise of blood glucose in the *Apoe*$^{-/-}$ mouse model.

A significant reverse correlation was observed between plasma HDL cholesterol levels and fasting glucose levels in this cross on either chow or Western diet. This result is consistent with the findings of prospective human studies that low HDL levels can predict the future risk of developing T2D and low HDL levels are more prevalent in diabetic patients than in the normal population[135,136]. HDL can increase insulin secretion from β-cells, improve insulin sensitivity of the target tissues, and accelerate glucose uptake by muscle via the AMP-activated protein kinase[137]. A significant correlation of non-HDL cholesterol levels with fasting glucose levels was also observed in this cross, and the correlation was extremely high when mice were fed the Western diet. Emerging human studies have also revealed associations of non-HDL cholesterol and *ApoB* with fasting glucose levels and incident type 2 diabetes[138–140]. We previously observed that the elevation of non-HDL cholesterol levels in *Apoe*$^{-/-}$ mice during the consumption of a Western diet induces a chronic, low-grade inflammation state characterized by rises in circulating cytokines and infiltration of monocytes/macrophages in various organs or tissues[112,115,118,133]. Inflammation in the islets impairs β-cell function[118]. LDL can also directly affect function and survival of β-cells[141]. In addition, high levels of LDL can

induce insulin resistance due to its lipotoxicity and effect on endoplasmic reticulum stress[100].

Plasma triglyceride levels were strongly correlated with fasting glucose levels in this cross on the Western diet, although no significant correlation was found when mice were fed the chow diet. Despite the strong correlation, no overlapping QTLs were observed for fasting glucose and triglyceride. The reason for the discrepancy between non-HDL cholesterol and triglyceride in terms of the presence or absence of co-localized QTLs is unclear.

A suggestive QTL for fasting glucose near the proximal end of chromosome 9 (2.37 cM) was detected in this cross, initially on the chow diet and then replicated on the Western diet. The LOD score plot for chromosome 9 has shown 2 distinct peaks, one with a suggestive LOD score at the proximal end and one with a significant LOD score at a more distal region, suggesting the existence of two loci for fasting glucose on the chromosome. The bootstrap test, a statistical method for defining the confidence interval of QTLs using simulation[142], also indicated the existence of two QTLs for the trait on chromosome 9. We named the proximal one *Bglu16* to represent a new QTL for fasting glucose in the mouse. Naming a suggestive locus is considered appropriate if it is repeatedly observed[66].

Two suggestive QTLs for fasting glucose on chromo- some 5 were identified when mice were fed the Western diet. The proximal one replicated *Bglu13*, recently mapped in the B6-Apoe$^{-/-}$ x BALB-Apoe$^{-/-}$ cross[76]. One probable candidate gene for this QTL is *Hnf1a*, which encodes hepatocyte nuclear factor 1α. In humans, *Hnf1a* mutations are the most common cause of maturity-onset diabetes of the young (MODY)[143]. The suggestive QTL in the distal region was novel.

Most of the QTLs identified for plasma lipids confirm those identified in

previous studies, whereas two QTLs for non-HDL are new and named *Nhdlq15* and *Nhdlq16*, respectively. The QTLs on distal chromosome 1 for HDL and triglyceride has been mapped in a number of mouse crosses, and *Apoa2* has been identified as the underlying causal gene[144]. However, the QTL (~90 cM) mapped in this study showed that it was more distal to the *Apoa2* gene (80 cM), thus suggesting a different underlying causal gene.

In this study, we have identified multiple QTLs contributing to dyslipidemia and hyperglycemia in a segregating F_2 population. The finding on the co-localization of QTLs for fasting glucose, HDL and non-HDL cholesterol levels and the sharing of probable candidate genes has demonstrated genetic connections between dyslipidemia and type 2 diabetes. The close correlations of fasting glucose with HDL, non-HDL cholesterol, and triglyceride support the hypothesis that dyslipidemia plays a causal role in the development of type 2 diabetes[100]. The haplotype analysis has prioritized candidates for either chromosome 9 QTL down to a handful of genes. Nevertheless, functional studies need to be performed to prove causality.

2.5 Methods

Ethics statement

All procedures were in accordance with current National Institutes of Health guidelines (https://grants.nih.gov/grants/olaw/Guide-for-the-Care-and-use-of-laboratory-animals.pdf) and approved by the institutional Animal Care and Use Committee (protocol #: 3109). Blood was drawn from the retro-orbital plexus of overnight fasted mice with the animals under isoflurane anesthesia.

Animals, experimental design and procedures

BALB and SM *Apoe*$^{-/-}$ mice were created using the classic congenic breeding strategy, as described[114]. BALB- *Apoe*$^{-/-}$ mice were crossed with SM-*Apoe*$^{-/-}$ mice to generate F1s, which were intercrossed by brother-sister mating to generate a female F2 cohort. Mice were weaned at 3 weeks of age onto a rodent chow diet. At 6 weeks of age, female F2 mice were started on a Western diet containing 21 % fat, 34.1 % sucrose, 0.15 % cholesterol, and 19.5 % casein *by weight* (Harlan Laboratories, TD 88137) and maintained on the diet for 12 weeks. Mice were bled twice: once before initiation of the Western diet and once at the end of the 12-week feeding period. Overnight fasted mice were bled into tubes containing 8 μL of 0.5 mol/L ethylenediaminetetraacetic acid. Plasma was prepared and stored at −80 °C before use.

Housing and husbandry

Breeding pairs were housed in a cage of 1 adult male and 2 females, and litters were weaned at 3 weeks of age onto a rodent chow diet in a cage of 5 or less. At 6 weeks of age, F2 mice were switched onto the Western diet and maintained on the diet for 12 weeks. All mice were housed under a 12-h light/dark cycle at an ambient temperature of 23 °C and allowed free access to water and drinking food. Mice were fasted overnight before blood samples were collected.

Measurements of plasma glucose and lipid levels

Plasma glucose was measured with a Sigma glucose (HK) assay kit, as reported with modification to a longer incubation time[76]. Briefly, 6 μl of plasma samples were incubated with 150 μl of assay reagent in a 96-well plate for 30 min at 30 °C. The absorbance at 340 nm was read on a Molecular Devices (Menlo Park, CA) plate reader. The measurements of total cholesterol, HDL cholesterol, and triglyceride

were performed as reported previously[112]. Non-HDL cholesterol was calculated as the difference between total and HDL cholesterol.

Genotyping

Genomic DNA was isolated from the tails of mice by using the phenol/chloroform extraction and ethanol precipitation method. The Illumina LD linkage panel consisting of 377 SNP loci was used to genotype the F2 cohort. Microsatellite markers were typed for chromosome 8 where SNP markers were uninformative in distinguishing the parental origin of alleles. DNA samples from the two parental strains and their F1s served as controls. Uninformative SNPs were excluded from QTL analysis. SNP markers were also filtered based on the expected pattern in the control samples, and F2 mice were filtered based on 95 % call rates in genotype calls. After filtration, 228 F2s and 144 markers were included in genome-wide QTL analysis.

Statistical analysis

QTL analysis was performed using J/qtl and Map Manager QTX software as previously reported[88,117,145]. One thousand permutations of trait values were run to define the genome-wide LOD (logarithm of odds) score threshold needed for significant or suggestive linkage of each trait. Loci that exceeded the 95th percentile of the permutation distribution were defined as significant ($P < 0.05$) and those exceeding the 37th percentile were suggestive ($P < 0.63$).

Prioritization of positional candidate genes

The Sanger SNP database (http://www.sanger.ac.uk/ sanger/Mouse_SnpViewer/rel-

1410) was used to prioritize candidate genes for overlapping QTLs affecting plasma glucose and HDL cholesterol levels on chromosome (Chr) 9, which were mapped in two or more crosses derived from different parental strains for either phenotype. We converted the original mapping positions in cM for the confidence interval to physical positions in Mb and then examined SNPs within the confidence interval. Probable candidate genes were defined as those with one or more SNPs in coding or upstream promoter regions that were shared by the parental strains carrying the "high" allele but were different from the parental strains carrying the "low" allele at a QTL, as previously reported[79].

Chapter 3

Polygenic Control of Carotid Atherosclerosis in a BALB-cJ × SM-J Intercross and a Combined Cross Involving Multiple Mouse Strains

Andrew T. Grainger, Michael B. Jones, Mei-Hua Chen, and Weibin Shi

Published in: Grainger AT, Jones MB, Chen MH, Shi W. Polygenic Control of Carotid Atherosclerosis in a BALB/cJ x SM/J Intercross and a Combined Cross Involving Multiple Mouse Strains. G3 (Bethesda) (2016).

3.1 Abstract

Atherosclerosis in the carotid arteries is a major cause of ischemic stroke, which accounts for 85% of all stroke cases. Genetic factors contributing to carotid atherosclerosis remain poorly understood. The aim of this study was to identify chromosomal regions harboring genes contributing to carotid atherosclerosis in mice. From an intercross between BALB/cJ (BALB) and SM/J (SM) apolipoprotein E-deficient (*Apoe*$^{-/-}$) mice, 228 female F2 mice were generated and fed a "Western" diet for 12 wk. Atherosclerotic lesion sizes in the left carotid artery were quantified. Across the entire genome, 149 genetic markers were genotyped. Quantitative trait locus (QTL) analysis revealed eight loci for carotid lesion sizes, located on chromosomes 1, 5, 12, 13, 15, 16, and 18. Combined cross-linkage analysis using data from this cross, and two previous F2 crosses derived from BALB, C57BL/6J and C3H/HeJ strains, identified five significant QTL on chromosomes 5, 9, 12, and 13, and nine suggestive QTL for carotid atherosclerosis. Of them, the QTL on chromosome 12 had a high LOD score of 9.95. Bioinformatics-based analysis prioritized *Arhgap5*, *Akap6*, *Mipol1*, *Clec14a*, *Fancm*, *Nin*, *Dact1*, *Rtn1*, and *Slc38a6* as probable candidate genes for this QTL. Atherosclerotic lesion sizes were significantly correlated with non-HDL cholesterol levels (r = 0.254; p = 0.00016) but inversely correlated with HDL cholesterol levels (r = 20.134; p = 0.049) in the current cross. Thus, we demonstrated the polygenic control of carotid atherosclerosis in mice. The correlations of carotid lesion sizes with non-HDL and HDL suggest that genetic factors exert effects on carotid atherosclerosis partially through modulation of lipoprotein homeostasis.

3.2 Introduction

Stroke is the leading cause of extended disability and a major cause of mortality in the United States[146]. 800,000 people are estimated to experience a new or recurrent stroke and 131,000 die of stroke annually in this country. Ischemic stroke accounts for ~85% of all stroke cases and a large fraction of them are caused by atheromas in the carotid arteries[147]. Plaque in the carotid arteries directly or indirectly, though thrombus formation, blocks the blood flow to the brain[148,149]. Genetic studies of twins and families indicate that carotid arterial intima-media thickness and plaque, which reflect a thickening of the carotid artery wall and the presence of large irregular arterial wall deposits, respectively, is a genetically determined trait with heritability ranging from 30 to 60%[150–152]. Recent genome-wide association studies (GWAS) have identified over a dozen common variants associated with carotid intima-media thickness and plaque, including LRIG1, EDNRA, SLC17A4, PIK3CG, PINX1, ZHX2, APOC1, LDLR, ANGPT1, ZBTB7C, HDAC9, the BCAR1-CFDP1-TMEM170A locus, EBF1, and PCDH15[153–155]. However, these variants explain only a tiny fraction of the total heritability of the traits, suggesting that many more remain to be discovered. Furthermore, it is challenging to assess causality between a variant and disease in humans due to small gene effects, complex genetic structures, and environmental influences.

Genetic studies of animal models have contributed greatly to the understanding of the genetic basis of human diseases, including atherosclerosis. *Apoe*$^{-/-}$ mice develop all phases of atherosclerotic lesions in large- and medium-sized arteries, including the carotid arteries. QTL analysis for carotid atherosclerosis has been performed on two F2 populations derived from C57BL/6 (B6), C3H/HeJ (C3H), and BALB/cJ (BALB) strains and identified several significant and suggestive loci for the trait[122,128]. Nevertheless, more crosses are needed to identify new QTL and expedite the finding of underlying genes for carotid atherosclerosis. We have recently found that *Apoe*$^{-/-}$ mice with a SM/J (SM) genetic

background developed significantly larger atherosclerotic lesions than those with a BALB background[114]. In the present study, we generated a female F2 cohort from an intercross between the two *Apoe*$^{-/-}$ strains to search for loci contributing to carotid atherosclerosis. The combined cross analysis using data from multiple intercrosses has been shown to improve the resolution of shared QTL and increase the power of identify new QTL not found in an individual cross[156]. Thus, in this study we also performed a combined cross-linkage analysis using data from the current cross and two previously reported B6 x C3H and B6 x BALB intercrosses[122,128].

3.3 Results

3.3.1 Trait value frequency distribution

Values of atherosclerotic lesion sizes in the left carotid arteries of 228 F2 mice were distributed in the Pareto manner: the frequency of F2 mice with a total lesion size of <480x1000 mm^2 was the highest and then decreased with increasing lesion sizes (**Figure 3.1**). After being log2-transformed, these values exhibited a bimodal distribution with 25% of the F2 mice (n = 57) falling under the no or small lesion peak on the left (Log2 value, 2.2) and the remaining 75% falling under the bell- shaped curve on the right.

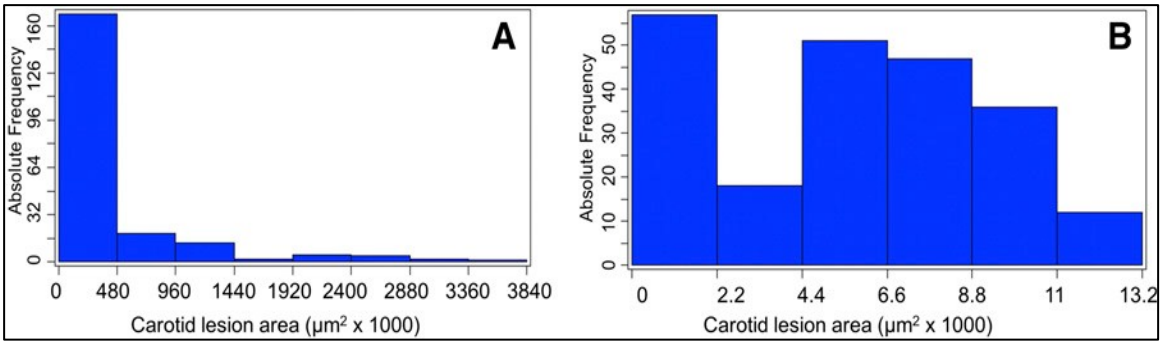

Figure 3.1 *Frequency distributions of untransformed (A) and log2-transformed (B) total carotid lesion areas of 228 female F2 mice derived from BALB-Apoe$^{-/-}$ and SM-Apoe$^{-/-}$ mice. Each histogram indicates the number of individual F2 mice with a certain lesion area. Apoe$^{-/-}$, apolipoprotein E-deficient.*

3.3.2 QTL analysis of carotid lesion sizes

Genome-wide scans for carotid lesion sizes were performed using both a nonparametric algorithm to analyze non-transformed lesion data and a parametric algorithm to analyze Log2-transformed lesion data (**Figure 3.2**). Eight suggestive QTLs, located on chromosomes 1, 5, 12, 13, 15, 16, and 18, were detected. With the exception of the QTLs on distal chromosome 5 and the one on chromosome 15, which were only detected with the nonparametric algorithm, all QTL were detected on both scans (**Table 3.1**). The QTL on chromosome 12 peaked at 30.28 cM and had a LOD score of 2.48. This QTL replicated *Cath1*, a locus for carotid atherosclerosis originally mapped in the B6 x C3H *Apoe*$^{-/-}$ intercross and then replicated in the B6 x BALB *Apoe*$^{-/-}$ intercross[122,128]. Two QTLs on chromosome 5 were detected: the proximal one had a suggestive LOD score of 2.33 and peaked at 63.4 cM, and the distal one had a LOD score of 2.03 and peaked at 99.4 cM. The distal locus overlapped

with *Cath2*, mapped initially in the B6 x C3H *Apoe*$^{-/-}$ intercross as a suggestive QTL for carotid atherosclerosis and then replicated in the B6 x BALB *Apoe*$^{-/-}$ intercross as a highly significant QTL[122,128]. The locus on chromosome 13 peaked at 34.02 cM and had a suggestive LOD score of 2.8. This QTL replicated *Cath3*, mapped in the B6 x BALB *Apoe*$^{-/-}$ intercross[128]. The QTL on chromosome 15 peaked at 46.74 cM and had a suggestive LOD score of 2.24. This QTL overlapped with a suggestive locus for atherosclerosis in the innominate artery and mapped a B6 x C3H *Apoe*$^{-/-}$ intercross[91]. We named it *Cath5* as this QTL was mapped in two separate crosses. The QTL on chromosome 18 had a suggestive LOD score of 2.22 and peaked at 16.27 cM. It replicated a suggestive QTL for carotid atherosclerosis mapped in the B6 x BALB *Apoe*$^{-/-}$ intercross[128], and was named *Cath6*.

The QTL on chromosome 1 peaked at 91.52 cM and had a LOD score of 2.17. It overlapped with *Ath1*, a QTL for aortic atherosclerosis mapped in a number of crosses[76,157,158]. The QTL on chromosome 16 peaked at 46.66 cM and had a score of 2.58, and this QTL was novel. The SM allele was associated with increased lesion sizes for chromosome 12, 13, 15, and 18 QTL, while the BALB allele was associated with increased lesion sizes for the chromosome 5 and 16 QTL (**Table 3.2**). The chromosome 1 QTL affected lesion formation in a heterotic manner in that F2 mice with heterozygous alleles exhibited increased lesion size over those with homozygous alleles.

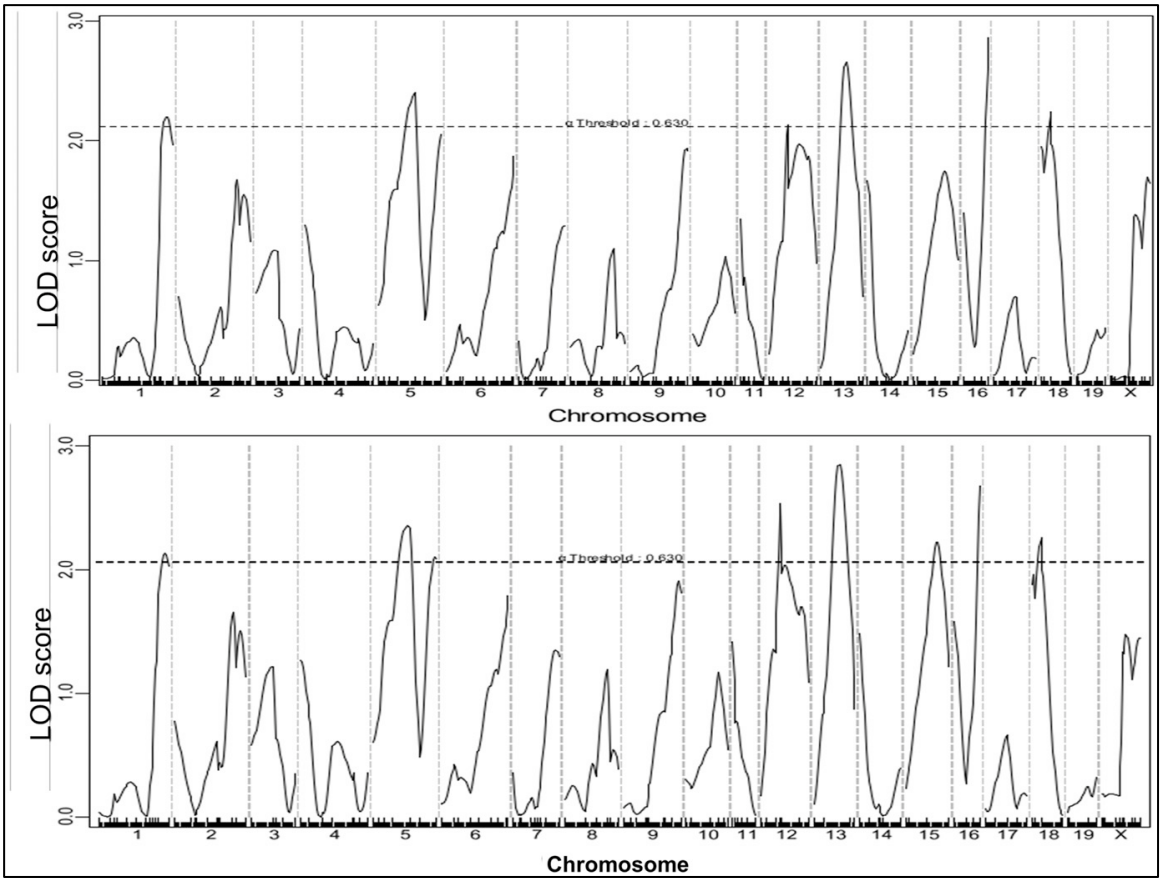

Figure 3.2 *Genome-wide QTL analysis for carotid lesion sizes in the F2 population. Chromosomes 1 through 20 are represented numerically on the x-axis. y-axis represents LOD score. The horizontal dashed line denotes the genome-wide threshold for suggestive linkage, which was determined by 1000 permutations. Top panel: a genome-wide scan using untransformed carotid lesion data performed with the nonparametric algorism; bottom panel: a genome-wide scan using log2-transformed carotid lesion data performed with the parametric mode. LOD, logarithm of the odds; QTL, quantitative trait locus.*

Table 3.1 *QTL identified for carotid lesion areas in female F2 mice derived from an intercross between BALB-Apoe$^{-/-}$ and SM-Apoe$^{-/-}$ mice*

Locus	Chr	Analysis	LOD[a]	*p*-Value[b]	Peak (cM)	95% C.I.[c]	High Allele	Mode of Inheritance[d]
	1	Nonparametric	2.17	0.535	91.52	75.52–97.02	—	Heterosis
	5	Nonparametric	2.33	0.422	63.4	34.19–101.24	BALB	Additive
Cath2	5	Nonparametric	2.03	0.630	99.4	79.4–101.2	BALB	Dominant
Cath1	12	Nonparametric	2.48	0.324	30.28	19.41–63.41	SM	Additive
Cath3	13	Nonparametric	2.8	0.163	34.02	22.02–46.02	SM	Dominant
Cath5	15	Nonparametric	2.24	0.474	46.74	26.74–62.74	SM	Recessive
	16	Nonparametric	2.58	0.274	44.66	13.43–46.66	BALB	Dominant
Cath6	18	Nonparametric	2.22	0.497	16.27	3.73–27.73	SM	Additive
	1	Parametric	2.23	0.545	87.52	77.52–97.02	—	Heterosis
	5	Parametric	2.38	0.413	67.27	33.4–101.4	BALB	Additive
Cath2	5	Parametric	2.03	0.630	99.4	79.4–101.2	BALB	Additive
Cath1	12	Parametric	2.1	0.644	30.28	23.41–65.41	SM	Additive
Cath3	13	Parametric	2.64	0.267	32.02	22.02–47.99	SM	Dominant
	16	Parametric	2.81	0.205	46.66	13.43–46.66	BALB	Dominant
Cath6	18	Parametric	2.21	0.552	16.27	3.73–25.73	SM	Additive

Chr, chromosome; LOD, logarithm of the odds; QTL, quantitative trait locus.

[a] *LOD scores were obtained from genome-wide scans using J/qtl. LOD score threshold for suggestive QTL . 2.054; for significance . 3.314 established by 1000 permutation tests.*

[b] *p-values represent genome-wide significance at each locus.*

[c] *95% C.I. was determined through whole-genome scans.*

[d] *Inheritance was determined based on the effect of each parental allele at the nearest genomic marker.*

Table 3.2 *Effects of BALB and SM alleles on carotid lesion area at identified QTL in female F2 mice derived from BALB-Apoe$^{-/-}$ and SM-Apoe$^{-/-}$ mice*

Locus Name	Chr	Analysis	Peak Marker	Peak (cM)	BB	BS	SS	*p*-Value
	1	Nonparametric	rs3685643	91.52	281.5 6 662.4	522.8 6 1049.5	260.9 6 466.9	0.016
	5	Nonparametric	rs3726547	63.4	604.1 6 1250.9	341.5 6 710.1	273.3 6 504.0	0.006953
Cath2	5	Nonparametric	rs13478578	99.4	454.6 6 629.3	412.1 6 1023.7	285.8 6 635.7	0.008146
Cath1	12	Nonparametric	rs13481509	30.28	171.1 6 389.1	374.9 6 952.2	652.9 6 917.3	0.002917
Cath3	13	Nonparametric	rs6259014	34.02	313.9 6 650.0	412.7 6 938.2	427.7 6 793.2	0.143
Cath5	15	Nonparametric	rs13482641	46.74	244.8 6 397.0	294.2 6 664.1	711.2 6 1281.3	0.03
	16	Nonparametric	rs3721202	44.66	426.5 6 1249.9	485.8 6 192.8	192.8 6 405.0	0.002091
Cath6	18	Nonparametric	rs3683699	16.27	256.1 6 440.2	423.9 6 1059.8	539.0 6 759.6	0.005427
	1	Parametric	rs3685643	87.52	4.3 6 3.8	6.1 6 3.7	5.8 6 3.2	0.01199282
	5	Parametric	rs3726547	67.27	6.8 6 3.3	5.2 6 3.6	4.8 6 3.8	0.00624255
Cath2	5	Parametric	rs13478578	99.4	6.4 6 3.6	5.6 6 3.5	4.2 6 3.8	0.00941731
Cath1	12	Parametric	rs13481509	30.28	4.7 6 3.2	5.3 6 3.8	6.8 6 3.8	0.00787988
Cath3	13	Parametric	rs6259014	32.02	4.6 6 3.8	5.7 6 3.7	6.0 6 3.6	0.14731571
	16	Parametric	rs3721202	46.66	5.8 6 3.4	6.2 6 3.7	4.0 6 3.6	0.00150063
Cath6	18	Parametric	rs3683699	16.27	5.3 6 3.5	5.0 6 3.9	7.1 6 3.2	0.00613849

Measurements for carotid lesion areas are expressed as means 6 SD. The unit for these measurements is: mm^2 x 1000 for nonparametric analysis. For parametric analysis, the values are log2-transformed total carotid lesion areas. The Kruskal–Wallis test was used on the nonparametric data and ANOVA on the parametric data to determine the significance (p-value) of the differences among the BB, BS, and SS genotypes. Chr, chromosome; BB, homozygous for the BALB allele at the linked peak marker; BS, heterozygous for both BALB and SMJ; SS, homozygous for the SMJ allele.

3.3.3 Combined cross analysis for overlapping QTL

Combined cross analysis was performed for carotid atherosclerosis using data from the current cross and two previously reported B6 x C3H and B6 x BALB intercrosses [122,128]. Five significant QTL, located on chromosomes 5, 9, 12, and 13, and nine suggestive QTL on chromosomes 2, 3, 6, 11, 15, 16, 18, and 19, were identified (**Figure 3.3 and Table 3.3**). The majority of these QTLs had been identified as significant or suggestive QTL in one or more individual crosses, but the LOD scores for the significant QTL on chromosomes 5, 9, 12, and

13 were higher compared to those determined in an individual cross. The 95% C.I. was relatively smaller than that in an individual cross for most QTL. A LOD score plot for chromosome 5 revealed two disparate peaks, indicating the presence of two QTL for carotid atherosclerosis (**Figure 3.4**). The distal QTL replicated *Cath2*, mapped in all the three crosses[122,128]. The proximal QTL was visible as a distinct peak in the current cross as well as the previously reported B6 x BALB intercross[128] and was named *Cath7* to represent a new locus for carotid atherosclerosis. The significant QTL on chromosome 9 was initially mapped as a suggestive QTL in the B6 x BALB intercross[128], and was named *Cath8*. The suggestive QTL on chromosomes 6, 11, 15, 16, and 18 were each mapped in one or more individual crosses, while the suggestive QTLs on chromosomes 2, 3, and 19 were only detected in the combined cross.

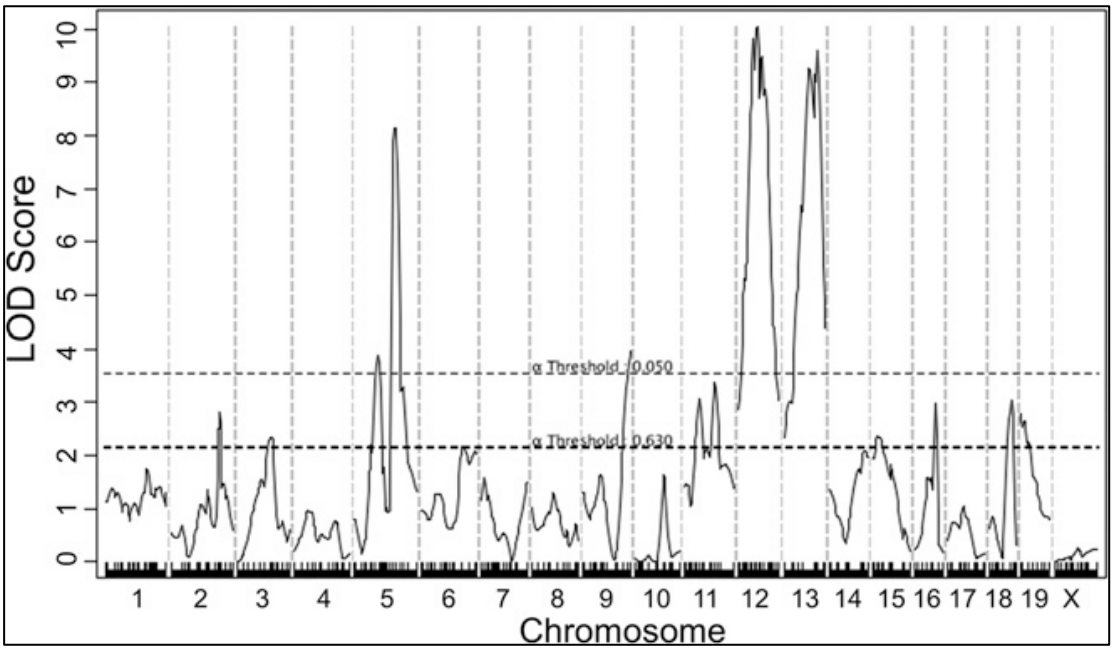

Figure 3.3 *Genome-wide QTL analysis for carotid lesion sizes using combined data from the current cross and two previously reported B6 x BALB and B6 x C3H Apoe*$^{-/-}$

intercrosses. The horizontal dotted lines indicate the thresholds for genome-wide suggestive and significant linkage, as determined by 1000 permutations. Apoe$^{-/-}$, apolipoprotein E-deficient; LOD, logarithm of the odds; QTL, quantitative trait locus.

Table 3.3 *Significant and suggestive QTL for carotid atherosclerosis identified in combined cross analysis of data from the current cross and the two previously reported crosses*

Locus	Chr	Trait	LOD	Peak (cM)	95% C.I.	Peak (Mb)	95% C.I. (Mb)
	2	Carotid lesion	2.77	80.22	44.22–98.22	159.59	71.96–170.59
	3	Carotid lesion	2.31	50.01	26.01–64.01	114.85	56.96–138.77
Cath7	**5**	**Carotid lesion**	**3.84**	**39.05**	**34.19–44.28**	**65.31**	**61.51–69.16**
Cath2	**5**	**Carotid lesion**	**8.06**	**66.35**	**63.84–70.35**	**127.32**	**124.83–131.29**
Cath4	6	Carotid lesion	2.15	66.21	1.53–88.79	120.60	6.44–145.75
Cath8	**9**	**Carotid lesion**	**3.92**	**75.33**	**66.37–75.33**	**114.09**	**103.61–114.09**
	11	Carotid lesion	3.02	26.1	18.2–32.2	45.28	30.91–54.19
	11	Carotid lesion	3.32	51	17.99–69.99	83.84	30.91–105.15
Cath1	**12**	**Carotid lesion**	**9.95**	**32.59**	**23.47–44.59**	**70.23**	**48.06–88.56**
Cath3	**13**	**Carotid lesion**	**9.49**	**53.35**	**36.02–56.02**	**100.5**	**68.40–103.48**
Cath5	15	Carotid lesion	2.35	11.26	3.8–37.8	30.76	7.87–71.92
	16	Carotid lesion	2.94	41.66	28.95–43.66	64.11	37.35–72.76
Cath6	18	Carotid lesion	3.01	39.73	31.73–41.73	62.50	56.27–65.17
	19	Carotid lesion	2.74	2.43	2.43–26.43	3.65	3.65–36.64

LOD score threshold for suggestive QTL was 2.128 and was 3.508 for significant QTL. Significant QTL are highlighted in bold. Chr, chromosome; LOD, logarithm of the odds; QTL, quantitative trait loci.

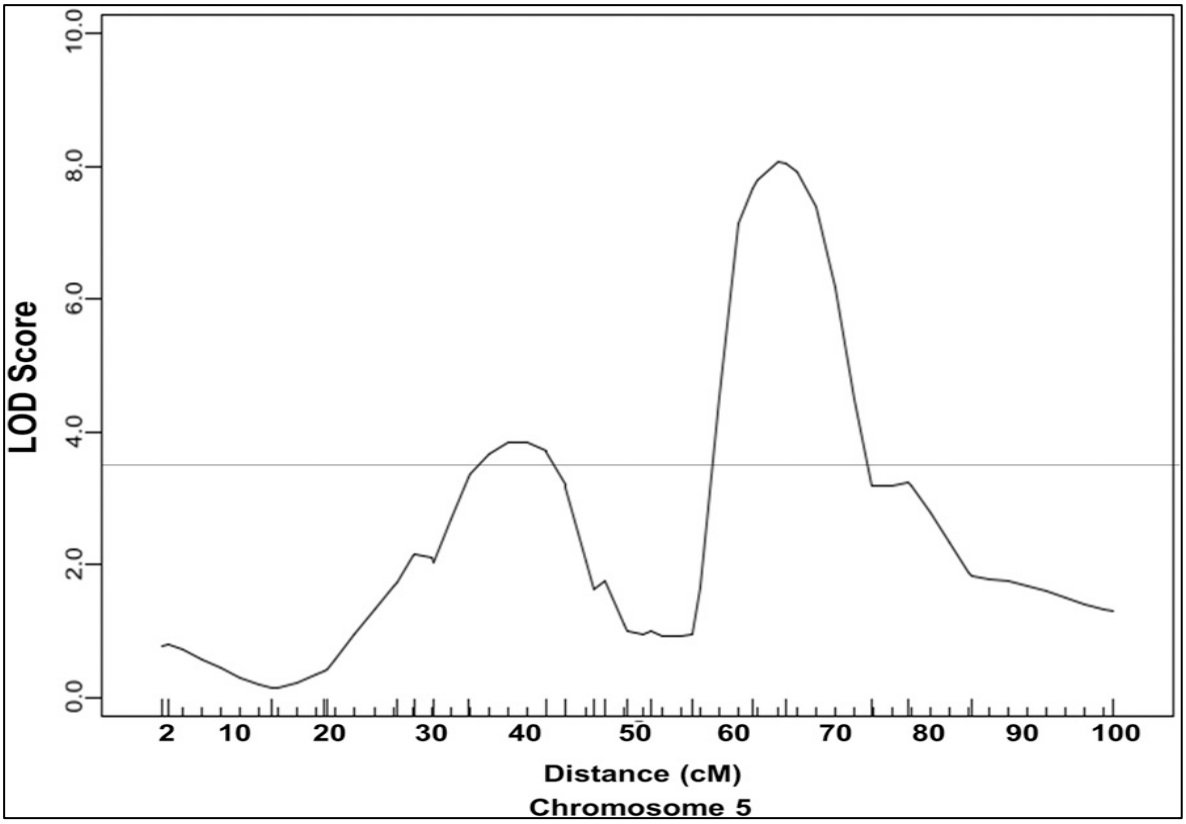

Figure 3.4 *Interval mapping graph for carotid lesion size on chromosome 5 using combined data from the current cross and previously reported B6 x BALB and B6 x C3H Apoe$^{-/-}$ intercrosses. The horizontal line denotes the threshold for significant linkage. Apoe$^{-/-}$ = apolipoprotein E-deficient; LOD = logarithm of the odds.*

3.3.4 Candidate genes for Cath1

Cath1 on chromosome 12 was mapped in the current cross and two previously reported B6 x C3H and B6 x BALB *Apoe*$^{-/-}$ intercrosses[122,128]. For this QTL, the B6 and SM alleles were associated with increased lesion sizes, while the C3H and BALB alleles were associated with smaller lesion sizes. We used the Sanger SNP database to search for

positional candidate genes that contain non-synonymous SNP(s) or SNP(s) in upstream regulatory regions that are shared by the low allele strains (BALB and C3H) but are different from ones carried by the high allele strain (B6) under the linkage peak. The SM strain was not included due to its incomplete genomic sequences for the region. Twenty-four candidate genes were identified (**Table 3.4**). Among them, *Eapp*, *Foxa1*, *Fancm*, *Nin*, *Dact1*, *Rtn1*, and *Trmt5* contained one or more non-synonymous SNPs with a low SIFT (Sorting Intolerant From Tolerant) score, predicting a high likelihood that an amino acid substitution has an adverse effect on protein function.

Table 3.4 *Haplotype analysis for Cath1 on chromosome 12 (52–75 Mb)*

Chr	Position	Gene	dbSNP	High Allele	Low Allele		Consequence	Amino Acid Change	SIFT Score	Tolerated
				C57BL/6	BALB_cJ	C3H_HeH				
12	52006466	Dtd2	rs46701436	A	G	G	Missense variant	Cn 7:V/A	0.92	Yes
12	52023971	Gpr33	rs29173669	A	G	G	Missense variant	Cn 95:V/A	0.71	Yes
12	52027979	Gpr33	rs51561875	T	G	G	5'-UTR variant			
12	52027989	Gpr33	rs49936313	T	A	A	5'-UTR variant			
12	52027993	Gpr33	rs47019843	C	T	T	5'-UTR variant			
12	52519522	Arhgap5	rs29198609	T	C	C	Missense variant	Cn 1092:V/A	1	Yes
12	52887261	Akap6	rs29183247	G	A	A	Missense variant	Cn 512:R/Q	0.2	Yes
12	52887389	Akap6	rs29223294	A	G	G	Missense variant	Cn 555:T/A	0.47	Yes
12	53140291	Akap6	rs48484112	G	A	A	Missense variant	Cn 1496:R/H	1	Yes
12	54203369	Egln3	rs29130898	A	G	G	Missense variant	Cn 65:C/R	0.89	Yes
12	54203615	Egln3	rs29122127	T	G	G	5'-UTR variant			
12	54203690	Egln3	rs13473456	G	A	A	5'-UTR variant			
12	54695720	Eapp	rs29183105	G	A	A	Missense variant	Cn 22:A/V	0.01	No
12	54941453	Baz1a	rs29195192	G	A	A	Missense variant	Cn 88:L/F	0.04	Yes
12	54999084	Baz1a	rs29196908	G	C	C	5'-UTR variant			
12	57303392	Mipol1	rs29163022	G	A	A	5'-UTR variant			
12	57325598	Mipol1	rs46300008	A	G	G	Missense variant	Cn 148:K/E	1	Yes
12	57325623	Mipol1	rs13481473	A	G	G	Missense variant	Cn 156:H/R	1	Yes
12	57542267	Foxa1	rs13481474	T	C	C	Missense variant	Cn 389:H/R	0	No
12	57576142	Ttc6	rs50478178	G	C	C	Missense variant	Cn 109:R/P	1	Yes
12	57725789	Ttc6	rs48534883	T	C	C	Splice region variant			
12	58267790	Clec14a	rs31966428	T	C	C	Missense variant	Cn 349:I/V	0.23	Yes
12	58268339	Clec14a	rs13465063	T	C	C	Missense variant	Cn 166:T/A	1	Yes
12	58268988	Clec14a	rs29162388	C	G	G	5'-UTR variant			
12	58268992	Clec14a	rs29194398	G	C	C	5'-UTR variant			

12	64471729	Fscb	rs13481500	G	A	A	Missense variant	Cn 988:P/S	1	Yes
12	64472091	Fscb	rs29131205	G	C	C	Missense variant	Cn 867:A/G	0.21	Yes
12	64472965	Fscb	rs585463036	C	A	A	Missense variant			
12	64473313	Fscb	rs29220106	G	A	A	Missense variant	Cn 460:P/S	0.2	Yes
12	64950146	Klhl28	rs33846378	C	T	T	Missense variant	Cn 474:V/I	0.07	Yes
12	65113969	Fancm	rs212043559	A	T	T	Missense variant	Cn 1407:N/I	0	No
12	65130342	Fancm	rs29212900	A	C	C	Missense variant	Cn 1987:I/L	0.47	Yes
12	65130397	Fancm	rs29213465	A	T	T	Missense variant	Cn 2005:Q/L	1	Yes
12	65130436	Fancm	rs29184120	A	C	C	Missense variant	Cn 2018:K/T	0	No (low confidence)
12	65149007	Mis18bp1	rs50634267	C	T	T	Missense variant	Cn 661: R/Q	0.27	Yes
12	65152837	Mis18bp1	rs29200949	T	C	C	Splice region variant			
12	65172467	Mis18bp1	rs3695606	T	A	A	5'-UTR variant			
12	65172551	Mis18bp1	rs3696207	A	G	G	5'-UTR variant			
12	69204274	Pole2	rs3704977	T	C	C	Splice region variant			
12	69223117	Pole2	rs29135637	T	C	C	Missense variant	Cn 78:M/V	0.43	Yes
12	69741794	Atp5s	rs29193315	G	A	A	Missense variant	Cn 156: V/I	1	Yes
12	70043177	Nin	rs32225358	C	T	T	Missense	Cn 1155:E/K	0.06	Yes
12	70043386	Nin	rs29192398	C	T	T	Missense	Cn 1085:R/Q	0.01	No
12	70043389	Nin	rs29159683	G	T	T	Missense	Cn 1084:S/Y	0.02	No
12	70043915	Nin	rs29149025	T	C	C	Missense	Cn 909:K/E	1	Yes
12	70180988	Abhd12b	rs29173916	G	T	T	Missense variant	Cn 258:M/I	1	Yes
12	70183081	Abhd12b	rs51691757	A	G	G	Splice region variant			
12	70183205	Abhd12b	rs32247424	A	G	G	Stop retained variant, 3'-UTR variant			
12	70193813	Pygl	rs32246688	G	T	T	Splice region variant			
12	70197551	Pygl	rs29151561	A	G	G	Splice region variant			
12	70201877	Pygl	rs13467444	T	C	C	Missense variant	Cn 323:M/V	1	Yes

Functional candidate genes are denoted in bold. A smaller SIFT score denotes a higher likelihood of protein function change. Chr, chromosome; dbSNP, Single Nucleotide Polymorphism Database identifier; SIFT, Sorting Intolerant From Tolerant; Cn, Coding non-synonymous polymorphism; UTR, untranslated region.

3.3.5 Relationships of carotid atherosclerosis with plasma lipids and glucose

Associations of carotid lesion sizes with plasma HDL, non-HDL cholesterol, triglyceride, and glucose levels were evaluated using the F2 population (**Figure 3.5**). A significant correlation with non-HDL cholesterol levels was observed (r = 0.254; p = 0.00016). F2 mice with higher non-HDL cholesterol levels tended to develop larger carotid

lesions. The value of the correlation coefficient r^2 indicates that non-HDL accounted for 6.45% of the variance in carotid lesion sizes among the F2 population. A marginal, but statistically significant, inverse correlation with HDL cholesterol levels was observed (r = 20.134; p = 0.049). F2 mice with higher HDL cholesterol levels tended to develop smaller carotid lesions. HDL accounted for 1.8% of the variance in lesion sizes of the F2 mice. No correlation with triglyceride levels was found (r = 0.021; p = 0.758). There was a trend toward a significant correlation with plasma glucose levels (r = 0.127; p = 0.062).

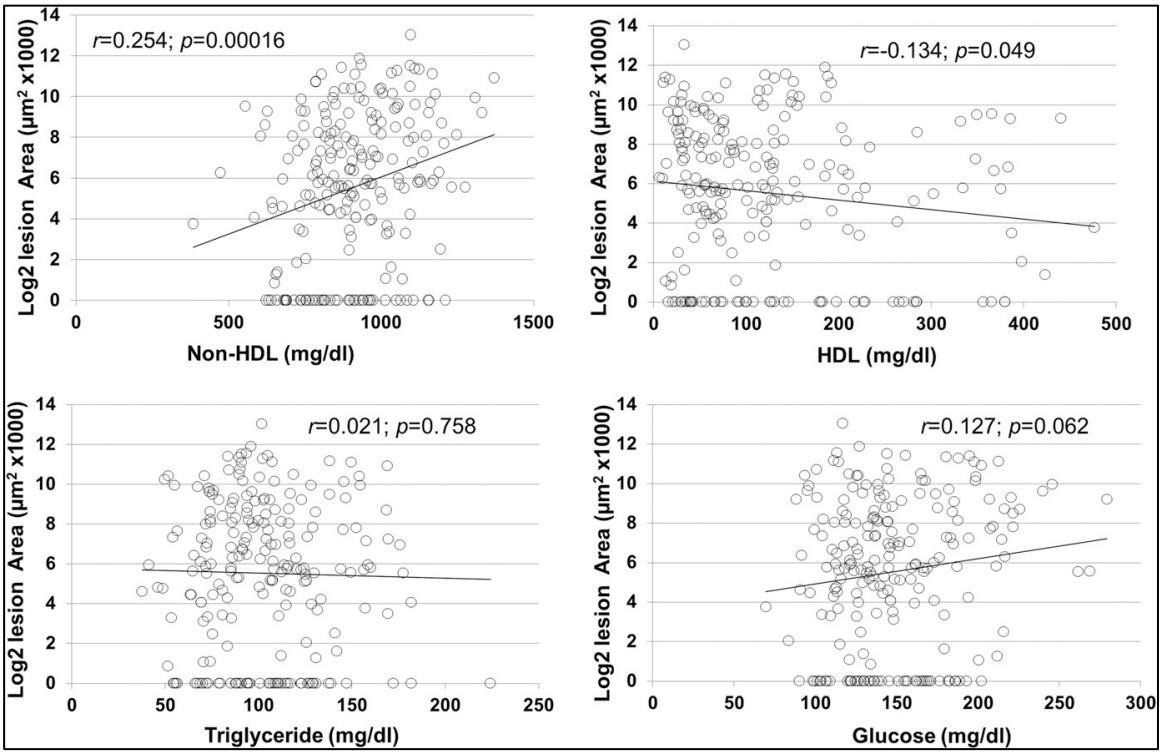

Figure 3.5 *Scatterplots showing the correlations of carotid lesion sizes with plasma non-HDL (A), HDL cholesterol (B), triglyceride (C), and glucose (D) in the F2 population. Each point represents an individual value of a F2 mouse. The correlation*

coefficient (r) and significance (p) are shown. Log2-transformed carotid total areas were used for the analyses. HDL, high-density lipoprotein.

3.4 Discussion

Genetic factors contributing to carotid atherosclerosis, which is a major cause of ischemic stroke, are poorly understood. In this study, we performed QTL analysis using data from a newly generated intercross and combined data from three independent intercrosses to search for QTL contributing to carotid atherosclerosis. Five significant QTLs and 10 suggestive QTLs were identified for the trait. Bioinformatics-based tools were successfully used to reduce the number of candidate genes for *Cath1*. Moreover, plasma non-HDL cholesterol was found to explain 6.5% of the variance in carotid lesion sizes of the F2 population.

Atherosclerotic lesions in the left carotid artery were measured after F2 mice were fed a Western diet for 12 weeks. Under this condition, these mice, which were on the *Apoe*$^{-/-}$ background, developed severe hyperlipidemia[159] (**Chapter 2**). Nevertheless, we found that a large fraction of F2 mice developed little or no atherosclerotic lesions in the carotid artery. The same phenomenon has also been observed in two other intercrosses previously constructed for QTL analysis of carotid atherosclerosis in the mouse[122,128]. In contrast, all the F2 mice developed atherosclerotic lesions in the aortic root[158] (**Chapter 4**). As the aortic root and the carotid arteries are exposed to the same level of lipoproteins and the same type of blood cells, the site-specific difference in the development of atherosclerosis should be attributable to local factors, such as vascular geometry, blood flow dynamics, and vessel wall properties. A genetic study of aortic arch curvature and atherosclerosis in a mouse

cross has linked genetic factors regulating aortic arch geometry to aortic lesion formation[160].

We and others have found that QTL identified for atherosclerotic lesions in the aortic root can be quite different from those mapped in another site of the vasculature, even in the same crosses[76,122,128,158,161]. Because the aortic root is easy to study in mice, genetic studies of atherosclerosis have largely focused on this site. However, this site has little clinical significance to humans. In contrast, the carotid arteries are the most extensively studied vessels in humans with ultrasonography because of their close association with the brain and ready accessibility.

Cath1 on chromosome 12, *Cath2* on chromosome 5, *Cath3* on chromosome 13, and *Cath4* on chromosome 6 are four significant QTLs for carotid atherosclerosis thus far mapped in two *Apoe*$^{-/-}$ mouse intercrosses[122,128]. Three of the four QTLs were replicated in the current BALB x SM *Apoe*$^{-/-}$ intercross, and all of them were replicated in the combined cross analysis. The QTL on chromosome 15 overlapped in the C.I. with a suggestive locus affecting both atherosclerotic lesion size and composition in the innominate artery of *Apoe*$^{-/-}$ mice[161]. We named it *Cath5* to represent a locus for carotid atherosclerosis in the mouse. Naming a suggestive locus is considered appropriate if it is repeatedly observed[66]. The QTL on chromosome 18 overlapped in the CI with a suggestive locus for carotid atherosclerosis mapped in the B6 x BALB *Apoe*$^{-/-}$ intercross, and was named *Cath6*.

Five significant QTLs and nine suggestive QTLs for carotid atherosclerosis were identified in the combined cross analysis. Nearly all of these QTLs were mapped in one or more individual crosses. However, the combined cross analysis had an increased power of detecting shared QTL by two or more crosses. Indeed, all five significant QTLs had a higher LOD score than that achieved in an individual cross. The current and previous B6 x BALB F2 crosses suggested the presence of two QTLs on chromosome 5 for carotid

atherosclerosis[128], while the combined cross analysis clearly demonstrated the presence of two disparate QTL on the chromosome. We named the proximal QTL *Cath7* to represent a new locus for carotid atherosclerosis. The significant QTL on distal chromosome 9 identified by the combined cross analysis overlapped with a suggestive QTL previously mapped in the B6 x BALB F2 cross[128], and was named *Cath8*. Consistent with the conclusion drawn by Li *et al.* (2005)[156], we found that the C.I. defined by the combined cross analysis was smaller than that defined in an individual cross for most of the QTL.

Cath1 has been mapped in three intercrosses derived from mouse strains, including B6, C3H, and BALB, whose genome sequences are publicly available through the Sanger mouse genomes project. By examining genes containing variants that were shared among the low allele strains (BALB and C3H) but different from those carried by the low allele strain (B6), we reduced the number of candidate genes to 24. Because a QTL is yielded from changes in the function or the quantity of a gene product, we concentrated on genes carrying a non-synonymous SNP or a SNP in the upstream regulatory region. *Nin*, *Dact1*, and *Rtn1*, which are located underneath the linkage peak and contain one or more non-tolerated non-synonymous SNPs, have shown suggestive associations with increased risk of ischemic stroke[162] or lipoprotein particle size[163].

A significant correlation was observed between non-HDL cholesterol levels and atherosclerotic lesion sizes in the present cross. Our previous study of a F2 population also showed a correlation between carotid lesion sizes and non-HDL cholesterol levels[128]. A marginal inverse correlation of HDL cholesterol levels with lesion sizes was observed in this cross, and also in two previous crosses[122,128]. These findings are consistent with the observations made in humans[164,165]. No correlation between carotid lesion sizes and plasma triglyceride levels was observed in this cross, nor in previous crosses[122,128]. We have observed a trend toward a significant correlation of carotid lesion sizes with fasting plasma

glucose levels in this cross. Blood glucose levels of the F2 mice were markedly elevated by feeding of a Western diet[159] (**Chapter 2**). In humans, impaired fasting glucose homeostasis has also been associated with preclinical carotid atherosclerosis[166].

In summary, we have identified a number of QTLs for carotid atherosclerosis, demonstrating the polygenic control of the disorder. The significant correlations of carotid lesion sizes with HDL and non-HDL cholesterol levels suggest that some loci exert effects on carotid atherosclerosis partially through action on lipoproteins. Using bioinformatics tools, we have reduced the list of candidate genes for a major atherosclerosis locus.

3.5 Methods

Animals and experimental design

BALB and SM *Apoe*$^{-/-}$ mice were generated in our laboratory using the classic congenic breeding strategy, as done previously[76,114]. The two *Apoe*$^{-/-}$ strains were crossed to generate F1s, which were intercrossed to generate a F2 population. Mice were weaned at 3 weeks of age onto a chow diet. At 6 weeks of age, female F2 mice were switched onto a Western diet containing 21% fat, 34.1% sucrose, 0.15% cholesterol, and 19.5% casein (TD 88137; Envigo) and maintained on the diet for 12 weeks.

Quantification of carotid atherosclerosis

Atherosclerotic lesion sizes in the left common carotid artery and its main branches were measured as previously reported with minor modifications[122]. Briefly, the vasculature of mice was perfused through the heart with 4% paraformaldehyde, then the distal portion of the left common carotid artery and its adjacent branches were dissected *en bloc* and

embedded in OCT compound (Tissue-Tek). Cryosections in 10 um thickness were collected in every three sections, stained with oil red O and hematoxylin, and counterstained with fast green. Lesion areas were measured under a microscope using Zeiss AxioVision 4.8 software. Carotid lesion sizes on all sections were added up for each mouse and this sum was used for statistical analysis.

Measurements of plasma lipids and glucose

Plasma total cholesterol, HDL cholesterol, triglyceride, and glucose were measured using assay kits as reported[112,159] (**Chapter 2**). Non-HDL cholesterol was calculated as the difference between total and HDL cholesterol.

Genotyping

The Illumina mouse LD linkage panel consisting of 377 SNP loci was used to genotype F2 mice, as reported[159] (**Chapter 2**). Microsatellite markers were typed for chromosome 8 where only one SNP marker was informative. DNA from the two parental strains and F1s served as controls. After excluding uninformative and poorly typed SNPs, 149 markers were included in genome-wide QTL analysis.

Statistical analysis

QTL analysis was performed using J/qtl. Genome-wide LOD score thresholds for significant or suggestive linkage were determined through 1000 permutations, as reported[88,145,159] (**Chapter 2**).

Combined cross analysis

A combined cross analysis was performed using data from the current cross and two previously published B6 x C3H and B6 x BALB intercrosses[122,128]. Genotype data for the chromosomal regions where a suggestive or significant QTL was found in an individual cross were recoded as "High" for F2s homozygous for the allele contributing to a larger lesion size, "Low" for F2s homozygous for the allele contributing to a smaller lesion size, and "H" for F2s with heterozygous alleles at each marker. For all other regions where no QTL was found, alleles at each marker were recoded based on the progenitor strain phenotype as reported[167]. Phenotype data on carotid lesion sizes were switched from the total lesion area to the average of the top five lesion sizes for each F2 mouse in all crosses.

Prioritization of candidate genes

Bioinformatics-based tools were used to prioritize candidate genes for major QTL that were mapped in two or more crosses derived from different parental strains. Probable candidate genes were those that contained one or more non-synonymous SNPs or a SNP in the upstream regulatory region, and that SNP was shared by the progenitor strains carrying the high allele but different from the one shared by the progenitor strains carrying the low allele at a QTL, as reported[79,158] (**Chapter 4**).

Data and reagent availability

BALB-Apoe$^{-/-}$ mice are available upon request. Supplemental Material, Supplementary file **S3.1** contains original genotype and phenotype data used for the current study.

Chapter 4

Genetic analysis of atherosclerosis identifies a major susceptibility locus in the major histocompatibility complex of mice

Andrew T. Grainger, Michael B. Jones, Jing Li, Mei-Hua Chen, Ani Manichaikul, and Weibin Shi

Published in part in: A.T. Grainger, M.B. Jones, J. Li, M.H. Chen, A. Manichaikul, W. Shi, Genetic analysis of atherosclerosis identifies a major susceptibility locus in the major histocompatibility complex of mice. Atherosclerosis 254 (2016), 124-132.

4.1 Abstract

Recent genome-wide association studies (GWAS) have identified 304 independent variants 243 loci associated at 5% FDR with coronary artery disease. However, these variants explain only 21.2% of the genetic heritability of the disease, suggesting that many more variants remain to be discovered. Here, we examined the genetic basis underlying the marked difference between SM/J-*Apoe*$^{-/-}$ and BALB/cJ-*Apoe*$^{-/-}$ mice in atherosclerotic lesion formation. 206 female F2 mice generated from an intercross between the two Apoe$^{-/-}$ strains were fed 12 weeks of western diet. Atherosclerotic lesion sizes in the aortic root were measured and 149 genetic markers genotyped across the entire genome. A significant locus, named *Ath49* (LOD score: 4.18), for atherosclerosis was mapped to the H2 complex [mouse major histocompatibility complex (MHC)] on chromosome 17. Bioinformatics-based analysis identified a single likely candidate gene *Mep1α*. Corresponding human genomic regions of Ath49 showed significant association with coronary heart disease. Five suggestive loci on chromosomes 1, 4, 5, and 8 for atherosclerosis were also identified. Atherosclerotic lesion sizes were significantly correlated with HDL but not with non-HDL cholesterol, triglyceride or glucose levels in the F2 cohort. In this study, we have identified the MHC as a major genetic determinant of atherosclerosis, highlighting the importance of inflammation in atherogenesis.

4.2 Introduction

Atherosclerosis, the primary cause of heart attack, ischemic stroke and peripheral arterial disease, is a complex disease resulting from interactions between environmental

and genetic factors[146]. The important role of genetic factors in atherosclerosis has been clearly demonstrated in numerous studies, including prospective studies of twins, families and cohorts[46–48], and genome-wide association studies (GWAS)[59–62]. Apart from rare cases of Mendelian inheritance that are caused by missense or nonsense mutations with large effect, the vast majority of coronary heart disease is complex, probably involving many genes of small effect[84]. Recent meta-analysis of GWAS data have identified 243 loci associated at 5% FDR harboring 304 common variants significantly associated with coronary heart disease[59–62]. These variants implicate pathways in blood vessel morphogenesis as well as lipid metabolism, nitric oxide signaling and inflammation[62]. Together these variants only explain 21.2% of the genetic heritability of coronary heart disease[62], suggesting that many more loci have not been discovered. Furthermore, most of the loci identified have small effect sizes with odds ratios (OR) < 1.25[61]; thus it's extremely challenging to establish causality between a genetic variant and disease in humans.

A complementary approach to finding genes and pathways involved in atherosclerosis is to study animal models. This allows for strict control over environmental influence and accurate phenotypic characterization of atherosclerotic lesions. Apolipoprotein E-null (*Apoe*$^{-/-}$) and LDL receptor-null (*Lldlr*$^{-/-}$) mouse models reproduce all phases of atherosclerotic lesions seen in humans[71,72]. Over a dozen intercrosses or backcrosses have been generated from atherosclerosis-susceptible and -resistant inbred strains with either *Apoe*$^{-/-}$ or *Lldlr*$^{-/-}$ background, leading to identification of 50 unique atherosclerosis susceptibility loci (http://www.informatics.jax.org/allele). Unfortunately, nearly all the crosses generated have chosen C57BL/6 (B6) mice as an atherosclerosis-susceptible strain; thus limiting their mapping power and coverage of allelic diversity. Creation of new genetic crosses using a different susceptible strain may discover new loci and also empower bioinformatics analysis for finding causative genes. When a QTL for the

same trait has been mapped to the same chromosomal location with multiple crosses derived from different inbred strains, whole-genome sequences and SNP databases available for them can be utilized to prioritize candidate genes. We recently have shown that SM/J (SM) and SWR/J *Apoe*$^{-/-}$ mice are susceptible to atherosclerosis compared to BALB/cJ (BALB) or C3H/HeJ *Apoe*$^{-/-}$ mice[114]. In this study, we performed QTL analysis using a female F2 cohort derived from an intercross between BALB/cJ (BALB)-*Apoe*$^{-/-}$ and SM-*Apoe*$^{-/-}$ mice to understand the genetic control of atherosclerosis susceptibility.

4.3 Results

4.3.1 QTL analysis of atherosclerotic lesions

Values of atherosclerotic lesion areas in 206 female F2 mice were calculated by summing up the top eight sections for each mouse. These values display a normal distribution (**Figure 4.1**).

Genome-wide QTL analysis of these data revealed one significant QTL on chromosome 17 and five suggestive QTLs on chromosomes 1, 4, 5, and 8 for atherosclerosis (**Figure 4.2**). Details of the QTLs detected, including locus name, LOD score, peak location, 95% confidence interval (CI), genome-wide P value, high allele, and mode of inheritance, are shown in **Table 4.1**. The QTL on chromosome 17 had a significant LOD score of 4.18 and peaked at 26.08 cM. It exerted effect in a dominant mode of inheritance with the BALB allele conferring susceptibility and the SM allele conferring resistance to atherosclerosis (**Figure 4.3A, Table 4.2**). This QTL was overlapping in the confidence interval with Ath26, mapped in an AKR-*Apoe*$^{-/-}$ × DBA-*Apoe*$^{-/-}$ intercross[80]. Because this QTL was mapped in an intercross derived from distinct parental strains, it was named *Ath49* in accordance to the guideline provided by the International Committee on Standardized Genetic Nomenclature for Mice

(http://www.informatics.jax.org/mgihome/nomen/gene.shtml). The QTL on chromosome 1 had a suggestive LOD score of 2.28 and peaked at 97.02 cM (**Figure 4.3B**). This QTL replicated Ath1, mapped in several crosses[76,83,157,160]. The QTL on chromosome 4 peaked at 71.37 cM and had a LOD score of 2.52. It replicated Athsq1 mapped in a (MOLF/Ei × B6-*Ldlr*$^{-/-}$) × B6-*Ldlr*$^{-/-}$ backcross[84]. Two suggestive QTLs were detected on chromosome 5, peaking at 34.19 and 69.4 cM, respectively (**Figures 4.2 and 4.3C**). The distal QTL replicated Ath42, mapped in a B6-*Apoe*$^{-/-}$ × BALB-*Apoe*$^{-/-}$ intercross[76]. The proximal QTL replicated a suggestive one mapped in a B6-*Apoe*$^{-/-}$ × C3H-*Apoe*$^{-/-}$ intercross[83] and was named *Ath50*. The QTL on chromosome 8 peaked at 46.14 cM and had a suggestive LOD score of 2.18. It was partially overlapping with *Ath40* mapped in a B6 × BALB/cByJ *LDLR*$^{-/-}$ intercross[77]. SM alleles were responsible for increased lesion sizes for the chromosome 1, 5 and 8 QTLs but decreased lesion sizes for *Ath49*. The chromosome 4 QTL affected lesion formation in a heterotic manner as F2 mice with heterozygous alleles exhibited increased lesion size over those with homozygous alleles.

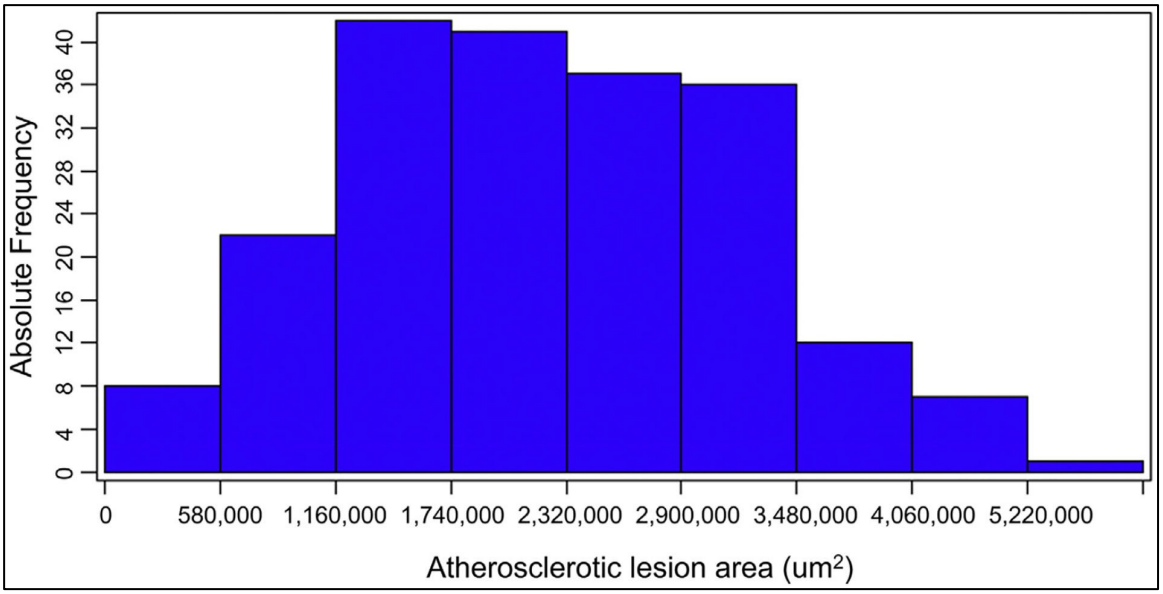

Figure 4.1. *Frequency distributions of atherosclerotic lesion sizes in 206 female F2 mice derived from BALB-Apoe$^{-/-}$ and SM-Apoe$^{-/-}$ mice. The lesion size of each mouse was the sum of lesion areas on 8 cross-sections with the largest readings. The graph was created using a plotting function of J/qtl software.*

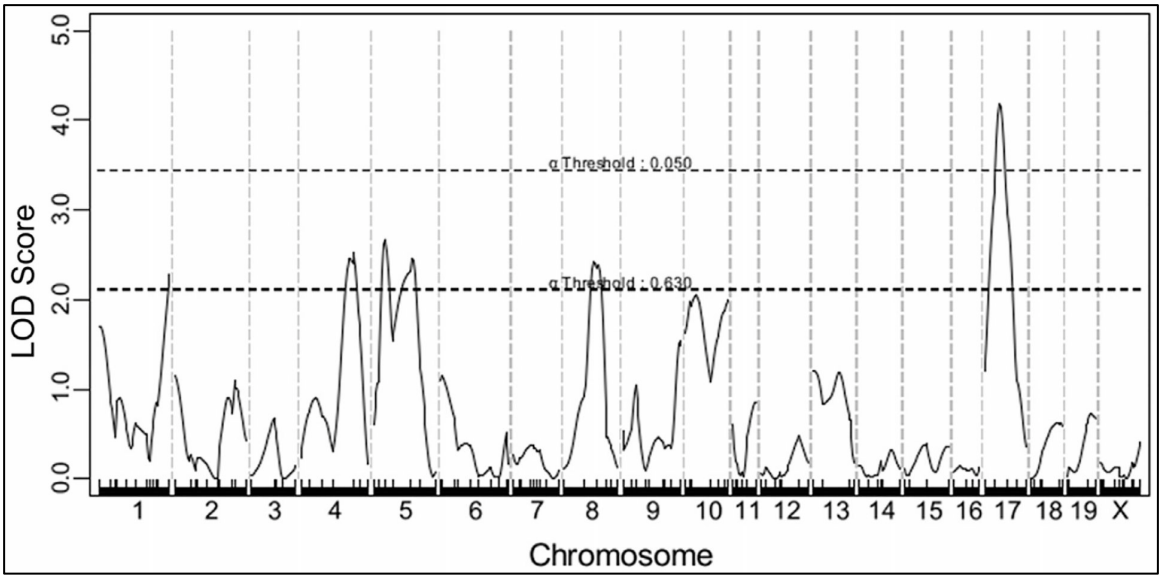

Figure 4.2. *A genome-wide scan to search for main effect QTLs influencing atherosclerotic lesion sizes in the F2 cohort. Chromosomes 1 through X are represented numerically on the X-axis. Each minor tick on the X axis represents one informative genetic marker. The Y-axis represents the LOD score. Two horizontal dashed lines denote genome-wide thresholds for suggestive (P = 0.63) and significant (P = 0.05) linkage.*

Table 4.1 *QTLs identified for aortic lesion areas in female F_2 mice derived from BALB-Apoe$^{-/-}$ and SM-Apoe$^{-/-}$ mice.*

Locus	Chr	LOD[a]	*p*–value[b]	Peak (cM)	Peak (Mb)	Peak marker	95% CI (cM)[c]	High allele	Mode of inheritance[d]
Athl	1	2.28	0.499	97.02	177.27	rs13476259	5.52–97.02	SMJ	Recessive
Athsql	4	2.524	0.333	71.37	119.72	rs3688968	52.32–84.32	–	Heterosis
Ath50	5	2.664	0.263	34.19	61.51	rs6354160	29.4–45.4	SMJ	Recessive
Ath42	5	2.809	0.191	69.4	108.38	rs3726547	45.4–75.4	SMJ	Recessive
Ath40	8	2.176	0.585	46.14	89.35	D8Mit50	20.14–56.14	SM	Recessive
Ath49	17	4.184	0.01	26.08	43.09	rs4231494	16.08–35.34	BALB	Dominant

[a]*LOD score threshold for suggestive QTL: > 2.108; for significant: >3.448 established by 1000 permutation tests.*

[b]*P-values represent genome-wide significance at a locus.*

[c]*95% CI (confidence interval) was obtained from whole-genome scans with J/qtl.*

[d]*Inheritance was determined based on the effect of each parental allele at the nearest genomic marker.*

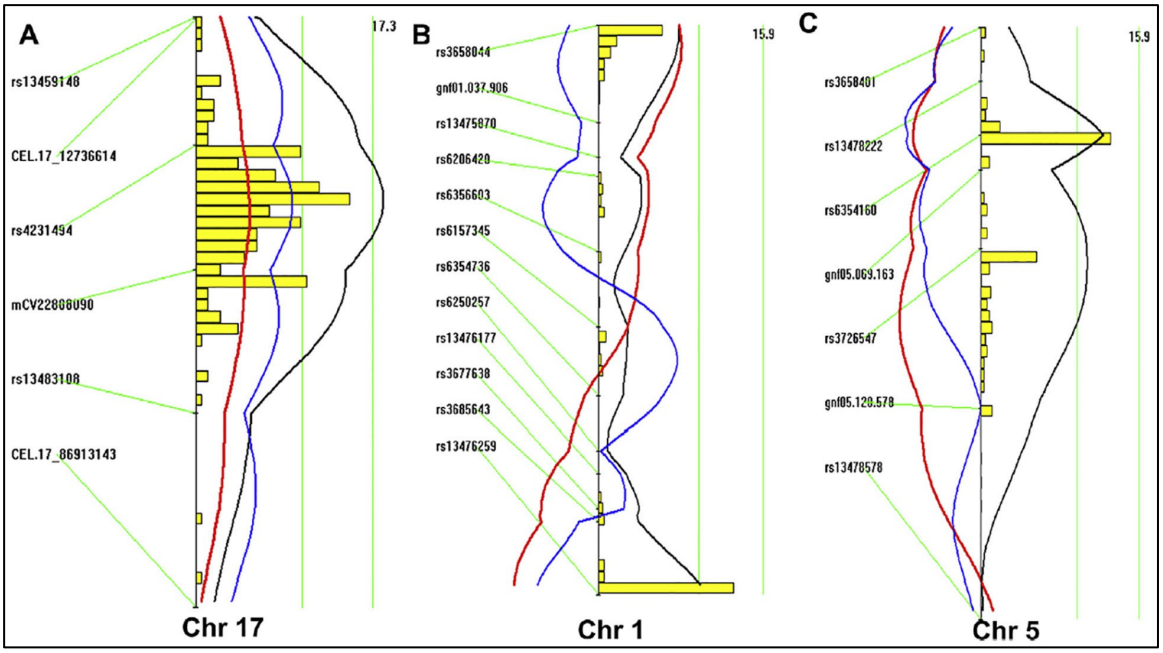

Figure 4.3. *Interval mapping graphs for atherosclerotic lesion sizes plotted by MapManager QTX. A, chromosomes 17; B, chromosome 1; and C, chromosome 5. The black line represents LOD scores calculated at a 1-cM interval, the blue plot represents the effect of BALB allele, and the red plot represents the effect of SM allele. The histogram estimates the confidence interval of a QTL. Two green vertical lines represent genome-wide LOD thresholds for suggestive (left line) and significant linkage (right line).*

Table 4.2 *Effects of BALB and SM alleles on aortic lesion area at identified QTLs in female F2s derived from BALB-Apoe$^{-/-}$ and SM-Apoe$^{-/-}$ mice.*

Locus name	Chr	Peak marker	Peak (cM)	Peak marker (Mb)	BB	BS	SS	*p*–Value
Ath1	1	rs13476259	97.02	177.27	2.0 ± 0.9914 (55)	2.097 ± 0.9426 (95)	2.561 ± 0.975 (56)	0.00565
Athsq1	4	rs3688968	71.37	119.72	2.107 ± 0.8958 (51)	2.42e ± 0.1025 (103)	1.875 ± 0.892 (52)	0.00326
Ath50	5	rs6354160	34.19	61.51	2.065 ± 0.7986 (55)	2.046 ± 0.9883 (91)	2.578 ± 0.1057 (10)	0.414
Ath42	5	rs3726547	69.4	108.38	2.037 ± 0.8441 (53)	2.079 ± 0.9479 (89)	2.536 ± 0.108 (63)	0.00599
Ath40	8	D8Mit50	46.14	89.35	1.957 ± 9.706 (60)	2.106 ± 0.9422 (86)	2.525 ± 0.9261 (54)	0.00471
Ath49	17	rs4231494	26.08	43.09	2.259 ± 9.691 (55)	2.415 ± 0.9221 (96)	1.785 ± 0.9925 (55)	0.000577

Measurements for aortic lesion areas are expressed as means ± SD. The unit for these measurements is: mm^2 × 10^6. BB, homozygous BALB allele; BS, heterozygous allele; SS, homozygous SMJ allele. Anova was performed on the data to determine the significance (P–value) of the differences among the BB, BS, and SS genotypes.

4.3.2 Candidate genes for Ath49

The QTL for atherosclerosis on chromosome 17 has also been mapped in an AKR-*Apoe*$^{-/-}$ × DBA/2-*Apoe*$^{-/-}$ intercross with the DBA/2 allele conferring susceptibility and the AKR allele conferring resistance to atherosclerosis[80] (**Figure 4.4**). The high allele strains BALB and DBA/2 share the same H2d haplotype, which falls within the *Ath49* interval, whereas the low allele strains SM and AKR have either H2v or H2k haplotype.

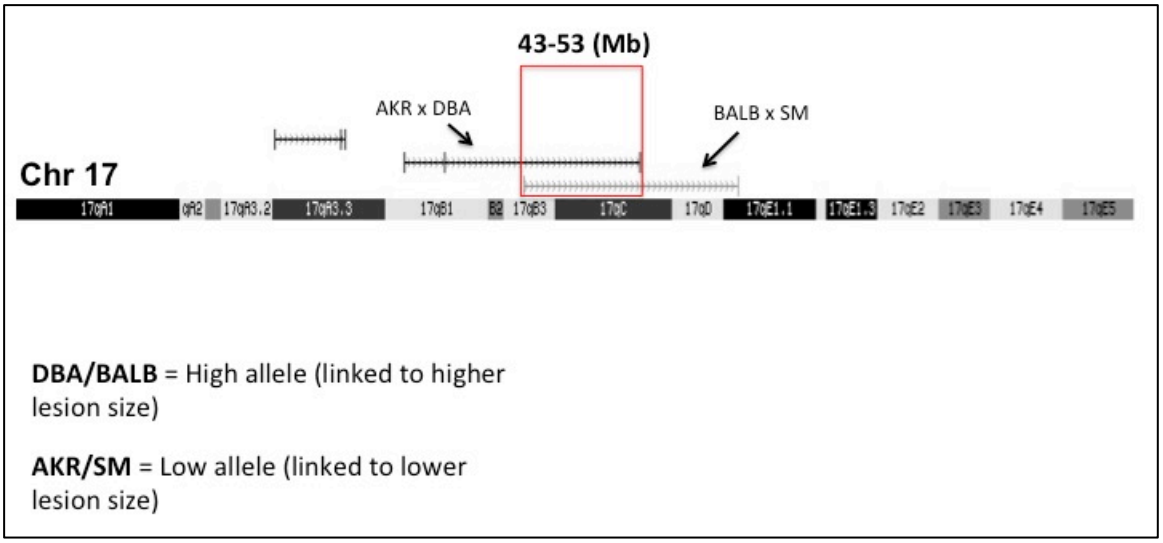

Figure 4.4 *QTL locations of Ath26 (AKR x DBA) and Ath49 (BALB x SM) on mouse chromosome 17. QTL were visualized through a custom track imported into UCSC Genome Browser. The overlapping confidence interval is highlighted in red*

To find candidate genes for the QTL, we employed a dual-armed strategy to investigate SNPs in *Ath49* that are affecting gene expression or protein structure and function (**Figure 4.5**). For gene expression, three filters were applied. Genes were initially filtered to those in which a significant change in gene expression is observed between the high allele strains (BALB, DBA/2) and low allele strains (AKR/SM) in either the aorta or the liver. This data was taken from the expression profiling in the Hybrid Diversity Mouse Panel (HMDP)(GEO accession: GSE66570)[91]. The second filter included solely those genes whose gene expression was significantly correlated with aortic lesion size. This data was obtained from correlations calculated between aortic lesion size and gene expression in the HDMP (GEO accession: GSE66570)[91]. The final filter included only those genes that had a significant expression QTL (eQTL) within the *Ath49* locus. This data was obtained for the

HDMP mice from the Systems Genetics Resource provided by the lab of Jake Lusis (https://systems.genetics.ucla.edu/data/hmdp). In this way, we are investigating only those genes in the aorta and liver that are different between the high and low alleles, have an effect on atherosclerosis, and whose expression is influenced by our QTL locus. From this strategy, three genes remained from the aorta and two from the liver. However, the genotypes of the high and low alleles were identical at all of these eQTL loci (**Tables 4.3 and 4.4**).

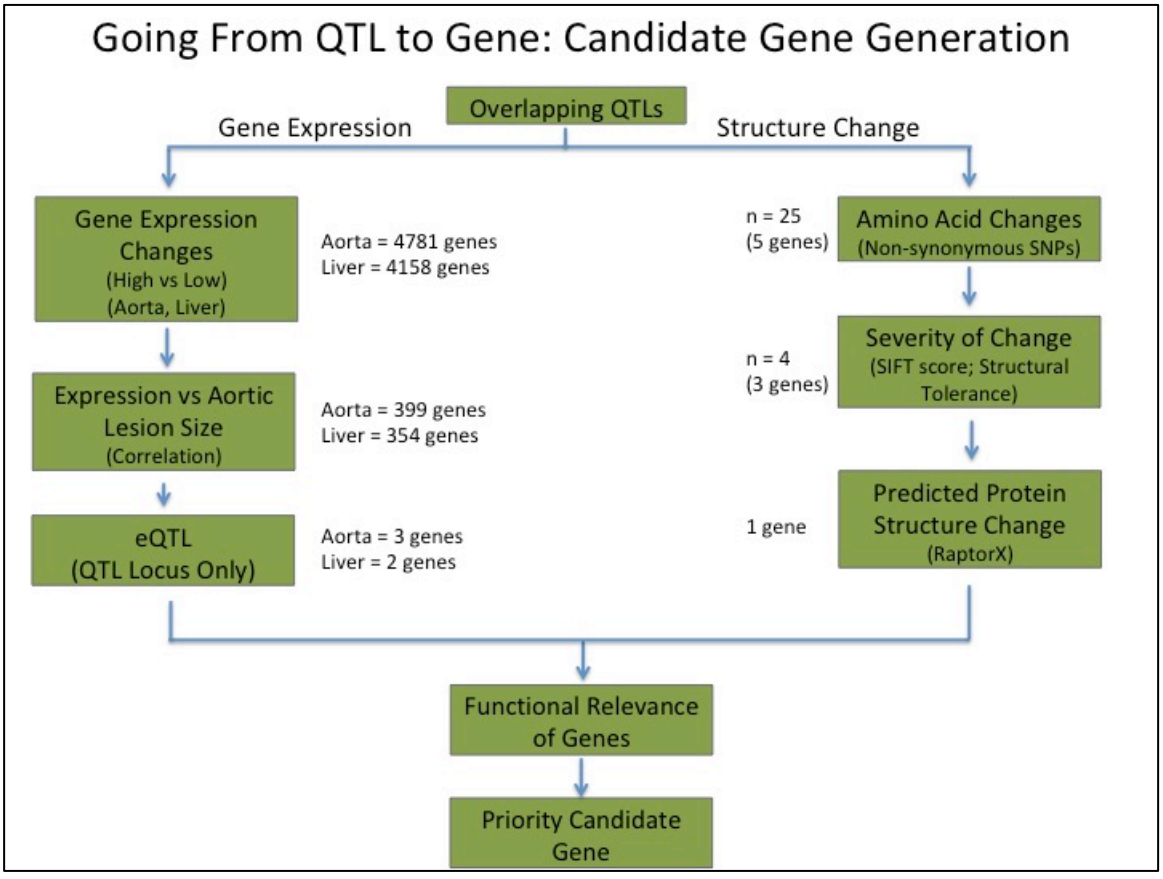

Figure 4.5 *Bioinformatics-based strategy for identifying candidate genes. Candidate genes were identified using independent datasets for looking at either gene expression*

(left) or protein structure change (right). The number of candidate genes is indicated at each step in the filtering process.

Table 4.3 *Investigation of candidate genes using aortic gene expression data*

							Low	Allele	High	Allele
Gene	logFC (Low vs High)	P-Value	Correlation	Correlation pvalue	*Ath49* eQTL	eQTL -log(p)	AKR	SM	BALB	DBA
Nras	**0.414**	0.00127	**-0.292**	0.011411092	rs29520755	6.68	T	T	T	T
Pcnp	**0.197**	0.0159	**-0.299**	0.009635889	rs33048535	8.22	G	G	G	G
					rs33699221	8.48	T	C	T	T
Zfr	**0.258**	0.0298	**-0.320**	0.005420569	rs33315472	6.08	?			
					rs33686613	6.00	T	T	T	T
					rs33401024	6.00	A	A	A	A
					rs33242678	6.00	T	T	T	T
					rs33675238	6.00	?			

Results of the candidate gene generation strategy using aortic gene expression data (n=4/5 mice per strain).

logFC = log fold change in gene expression of the low alleles compared to the high alleles.

P-Value = significance of logFC with a Benjamini & Hochberg FDR of 0.05 .

Correlation = Biweight midcorrelation of gene expression with aortic lesion size.

Ath49 eQTL = SNP with a significant eQTL linking Ath49 and gene expression.

eQTL -log(p) = Significance of eQTL linkage; -log(p) > 5.4 is considered significant.

Table 4.4 *Investigation of candidate genes using liver gene expression data*

Gene	logFC (Low vs High)	P-value	Correlation	Correlation pvalue	*Ath49* eQTL	eQTL -log(p)	Low Allele AKR	Low Allele SM	High Allele BALB	High Allele DBA
1810011O10Rik	**1.07**	0.00283	**-0.361997101**	3.13E-04	rs33271544	6.15	A	A	A	A
					rs29542098	8.68	?			
					rs29506165	5.92	A	A	A	A
					rs33048535	5.76	G	G	G	G
Slc10a2	**0.604**	0.0225	**-0.286969419**	0.004808876	rs33678407	10.50	G	G	G	G
					rs33182081	10.50	T	T	T	T
					rs33588112	10.50	?			
					rs33503146	8.87	G	G	G	G
					rs33162141	8.87	C	C	C	C
					rs33132950	8.87	T	T	T	T
					rs33409296	6.69	G	G	G	G
					rs29601524	6.52	G	A	G	G
					rs6377363	6.52	T	C	T	T

Results of the candidate gene generation strategy using liver gene expression data (n=4/5 mice per strain).

logFC = log fold change in gene expression of the low alleles compared to the high alleles.

P-Value = significance of logFC with a Benjamini & Hochberg FDR of 0.05 .

Correlation = Biweight midcorrelation of gene expression with aortic lesion size.

Ath49 eQTL = SNP with a significant eQTL linking Ath49 and gene expression.

eQTL -log(p) = Significance of eQTL linkage; -log(p) > 5.4 is considered significant.

For protein structure and function, we searched the Sanger SNP database for genes containing non-synonymous SNP(s) or SNP(s) in upstream regulatory regions that are shared by the high allele strains (BALB, DBA/2) but are different from ones carried by the low allele strain (AKR) in the confidence interval. SNPs for the SM were included afterword

as its genomic sequences for the region were not present in the Sanger SNP database. For these genes, we further searched other publicly available databases, including the Mouse Phenome Database (http:// phenome.jax.org/db/q?rtnsnp/home) and Ensembl (http://www.ensembl.org/index.html), and compiled the results in **Table 4.5**. *Tnfrsf21*, *Adgrf1*, *Adgrf5*, *Mep1a*, and *Tdrd6* are located precisely underneath the linkage peak and contain one or more non-synonymous SNPs leading to amino acid substitutions. SIFT (Sorting Intolerant From Tolerant) score, a parameter for estimating the impact of an amino acid substitution on protein function, suggest that the function of protein products encoded by *Adgrf1*, *Adgrf5*, and *Mep1α* is highly likely different between the high and low allele strains (**Table 4.5**).

Further investigation of these SNPs discovered that the non-synonymous variant in *Adgrf1* (Rs47784933; S/F, TCC/TTC; mouse population frequency: C = 83%, T = 17%) resides in no known motif and is not highly conserved (**Figure 4.6**). The variant in *Adgrf5* (Rs51535437; D/V, GAT/GTT; mouse population frequency: A = 78%, T = 22%) is in no known motif, yet is highly conserved (**Figure 4.7**). The two variants in *Mep1α* (rs215987489; F/Y; TTT/TAT; mouse population frequency: A = 72%, T = 28%)(rs218712173; N/S, AAC/AGC; mouse population frequency: T = 78%, C = 22%) are both highly conserved (**Figures 4.8 and 4.9**). These two variants in addition to the other non-synonymous SNPs all reside in the metalloprotease domain (**Figure 4.10**). We further used online software RaptorX to predict the impact of multiple amino substitutions on the 3D structure of *Mep1α* protein product (http://raptorx.uchicago.edu/StructurePrediction/). The predicted 3D structure of the *Mep1α* protein was noticeably different between the high allele strain BALB and the low allele strain AKR (**Figure 4.11**).

Table 4.5 *Haplotype analysis for Ath49 on chromosome 17 (43-54Mb)*

				Low allele		High allele					
Chr	Position	Gene	dbSNP	AKR	SMJ	BALB	DBA	Csq	Amino Acid Change	SIFT Score	Tolerated?
17	43040105	Tnfrsf21	rs48909028	A	A	G	G	missense_variant	Cn 387A/T	0.41	+
17	**43305361**	**Adgrf1**	**rs47784933**	**C**	**C**	**T**	**T**	**missense_variant**	**Cn 297 S/F**	**0.01**	**-**
17	**43441946**	**Adgrf5**	**rs51535437**	**T**	**T**	**A**	**A**	**missense_variant**	**Ch 593 D/V**	**0.02**	**-**
17	43449807	Adgrf5	rs46753646	G	G	A	A	missense_variant	Cn 798 I/V	1	+
17	43486313	Mep1a	rs50539426	T	T	C	C	missense_variant	Cn 288 V/I	1	+
17	43487139	Mep1a	rs214680904	A	A	G	G	missense_variant	Cn 255 P/L	1	+
17	**43487163**	**Mep1a**	**rs215987489**	**T**	**T**	**A**	**A**	**missense_variant**	**Cn 247 F/Y**	**0.01**	**-**
17	43487200	Mep1a	rs252705192	T	T	C	T	missense_variant	Cn 235 E/K	0.18	+
17	**43487208**	**Mep1a**	**rs218712173**	**C**	**C**	**T**	**T**	**missense_variant**	**Cn 232 N/S**	**0.02**	**-**
17	43487220	Mep1a	rs229991911	T	T	G	G	missense_variant	Cn 228 P/Q	0.13	+
17	43487302	Mep1a	rs250429860	A	A	G	G	missense_variant	Cn 201 H/Y	0.06	+
17	43487308	Mep1a	rs221971200	G	G	A	A	missense_variant	Cn 199 Y/H	0.44	+
17	43491606	Mep1a	rs46011116	G	G	T	T	missense_variant	Cn 189 N/T	0.9	+
17	43491690	Mep1a	rs50252056	G	G	A	A	missense_variant	Cn 161 F/S	0.21	+
17	43491700	Mep1a	rs255741402	T	T	C	C	missense_variant	Cn 158 G/R	0.41	+
17	43491733	Mep1a	rs107621975	C	C	G	G	missense_variant	Cn 147 Q/E	0.28	+
17	43491736	Mep1a	rs108717471	T	T	C	C	missense_variant	Cn 146 D/N	0.77	+
17	43624232	Tdrd6	rs29514745	C	C	T	T	missense_variant	Cn 1975 G/D	0.39	+
17	43624499	Tdrd6	rs221878489	T	T	C	C	missense_variant	Cn 1886 C/Y	0.36	+
17	43624829	Tdrd6	rs240899227	T	T	C	C	missense_variant	Cn 1776 R/H	0.12	+
17	43625292	Tdrd6	rs220739854	C	C	T	T	missense_variant	Cn 1622 S/G	0.14	+
17	43625324	Tdrd6	rs260352790	A	A	G	G	missense_variant	Cn 1611 A/V	1	+
17	43625634	Tdrd6	rs51245534	C	C	T	T	missense_variant	Cn 1508 A/T	0.58	+
17	43626615	Tdrd6	rs49793949	A	A	G	G	missense_variant	Cn 1181 Y/H	0.19	+
17	43630049	Tdrd6	rs33049652	A	A	G	G	missense_variant	Cn 36 V/A	0.28	+

Analysis was performed using a combination of SNP sources, including the Mouse Genomes Project and Mouse Phenome Database. Functional candidate genes are outlined in red. SIFT (Sorting Intolerant From Tolerant): an algorithm for estimating the effect of a nonsynonymous variant on protein function. A smaller SIFT score denotes a higher likelihood of protein function change.

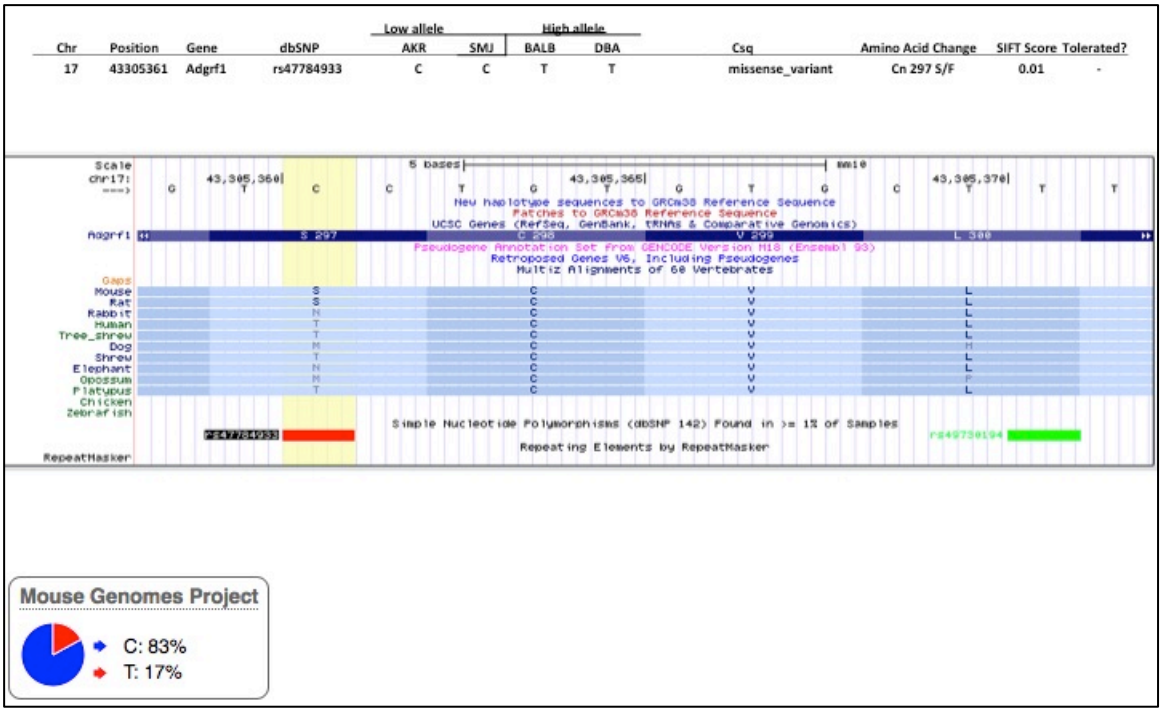

Figure 4.6 *Analysis of prioritized SNP in Adgrf1 from haplotype analysis of Ath49. SNP is highlighted in yellow and amino acid conservation was determined using UCSC genome browser (middle). Allele frequency was found using the Mouse Genomes Project data on Ensembl.*

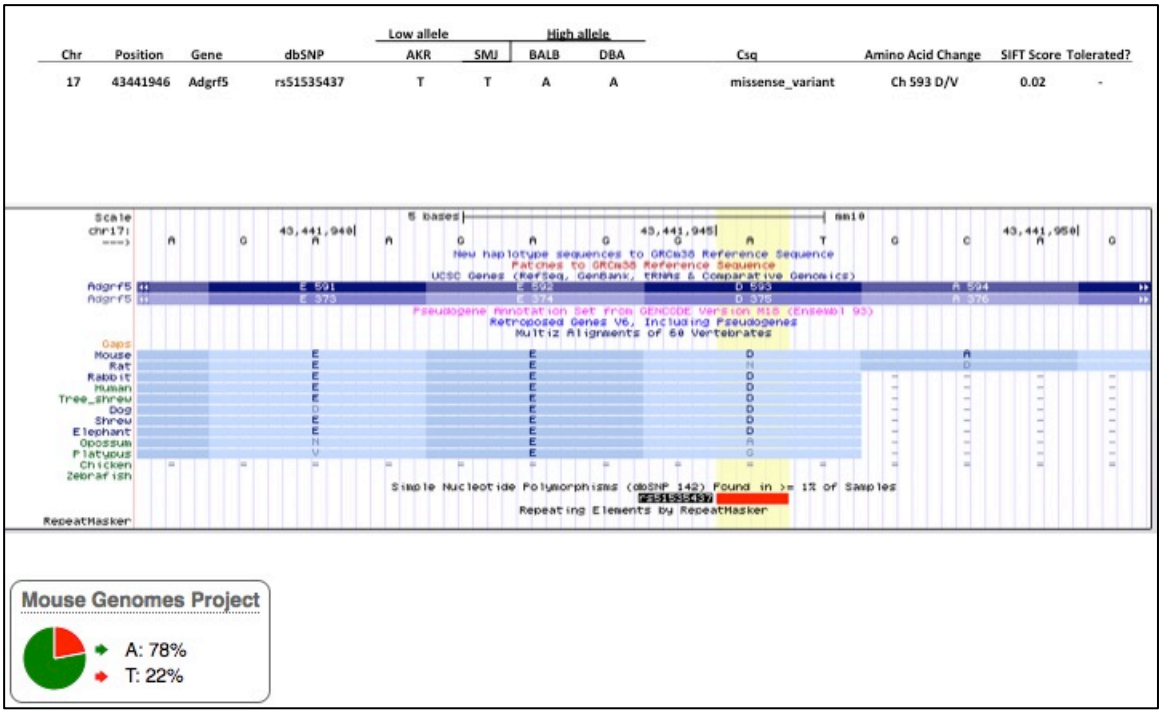

Figure 4.7 *Analysis of prioritized SNP in Adgrf5 from haplotype analysis of Ath49. SNP is highlighted in yellow and amino acid conservation was determined using UCSC genome browser (middle). Allele frequency was found using the Mouse Genomes Project data on Ensembl.*

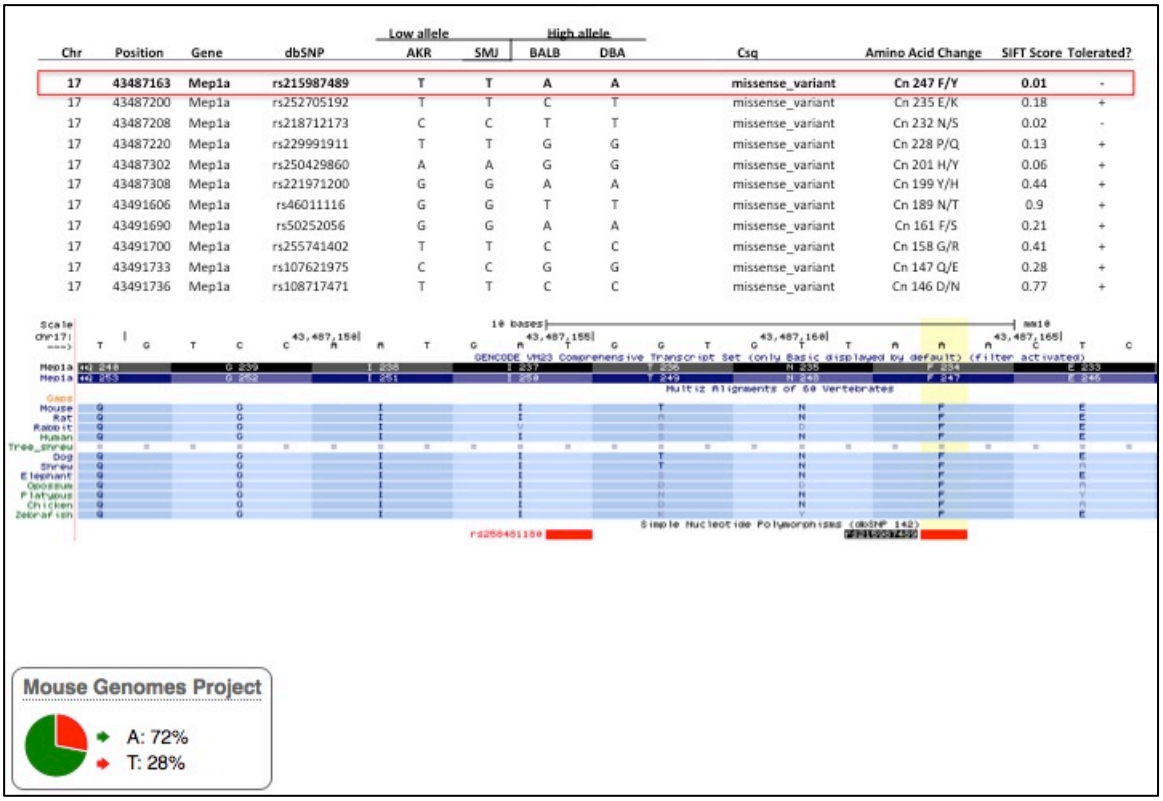

Figure 4.8 *Analysis of first prioritized SNP in Mep1α from haplotype analysis of Ath49. SNP being analyzed is highlighted in red (top). SNP is highlighted in yellow and amino acid conservation was determined using UCSC genome browser (middle). Allele frequency was found using the Mouse Genomes Project data on Ensembl.*

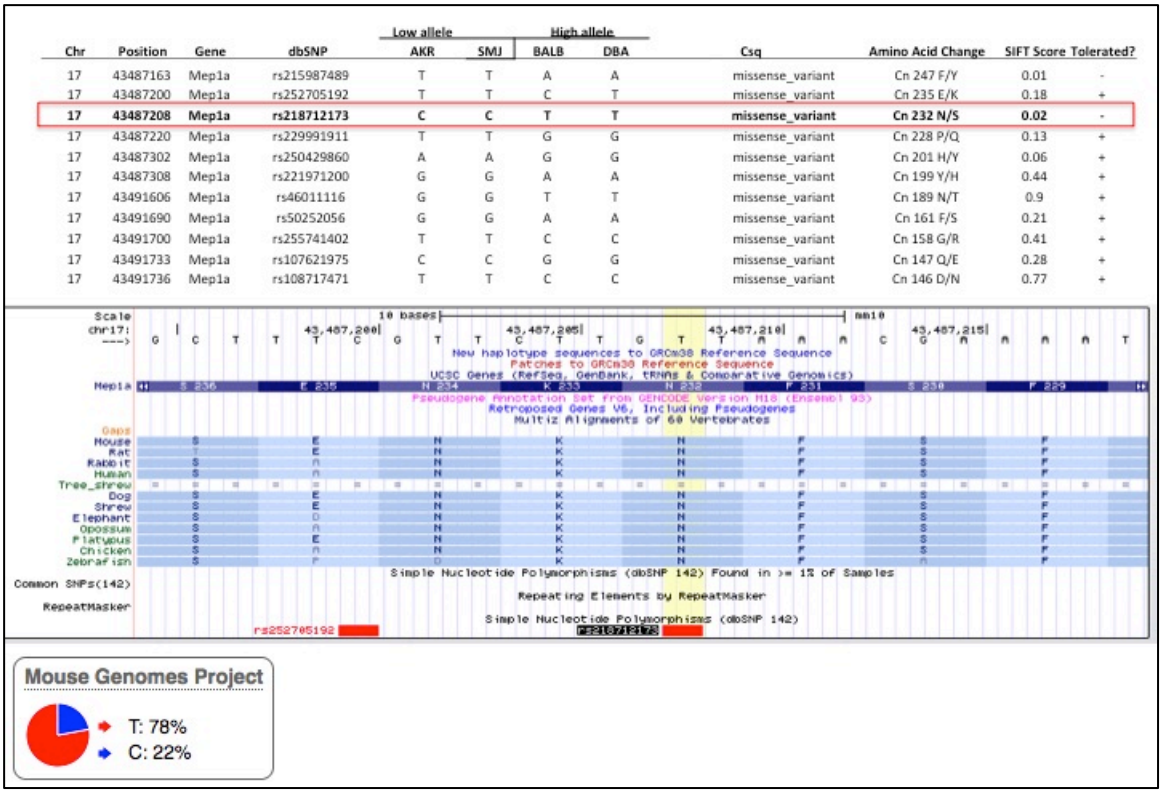

Figure 4.9 *Analysis of the second prioritized SNP in Mep1α from haplotype analysis of Ath49. SNP being analyzed is highlighted in red (top). SNP is highlighted in yellow and amino acid conservation was determined using UCSC genome browser (middle). Allele frequency was found using the Mouse Genomes Project data on Ensembl.*

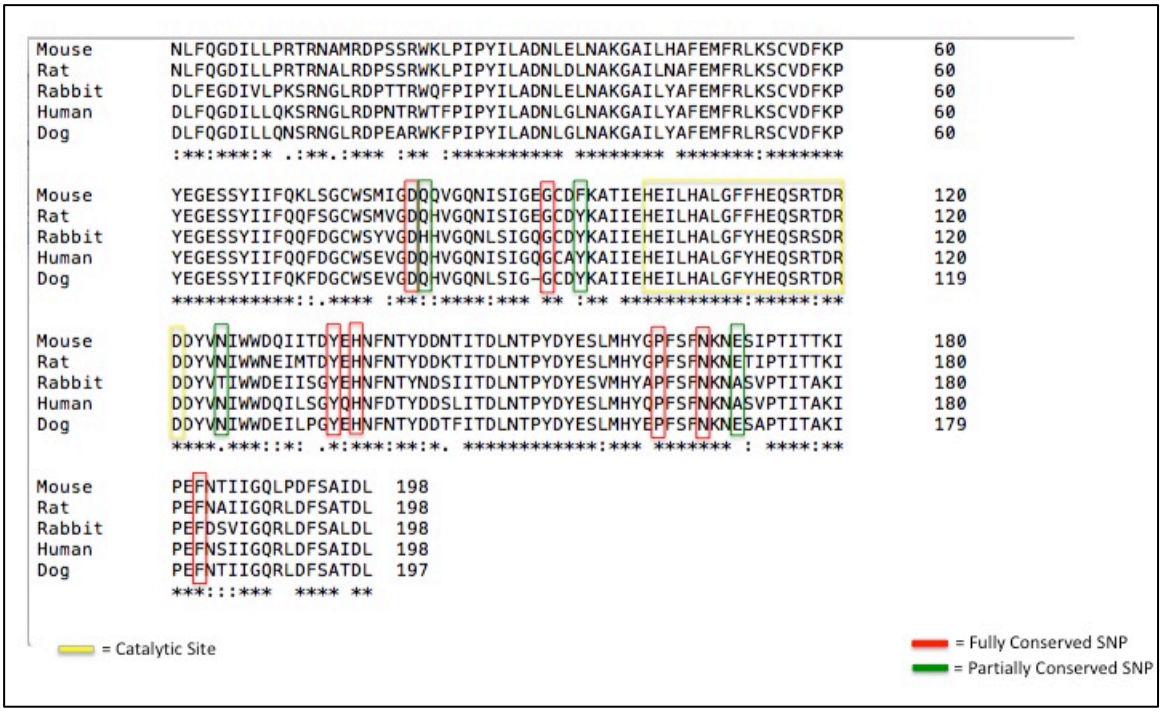

Figure 4.10 *Conservation of Mep1α metalloprotease domain. Fully conserved SNPs identified in the hapotype analysis are highlighted in red. Partially conserved SNPs are highlighted in green. The catalytic site of Mep1α is highlighted in yellow. Protein sequences were alligned using Clustal Omega.*

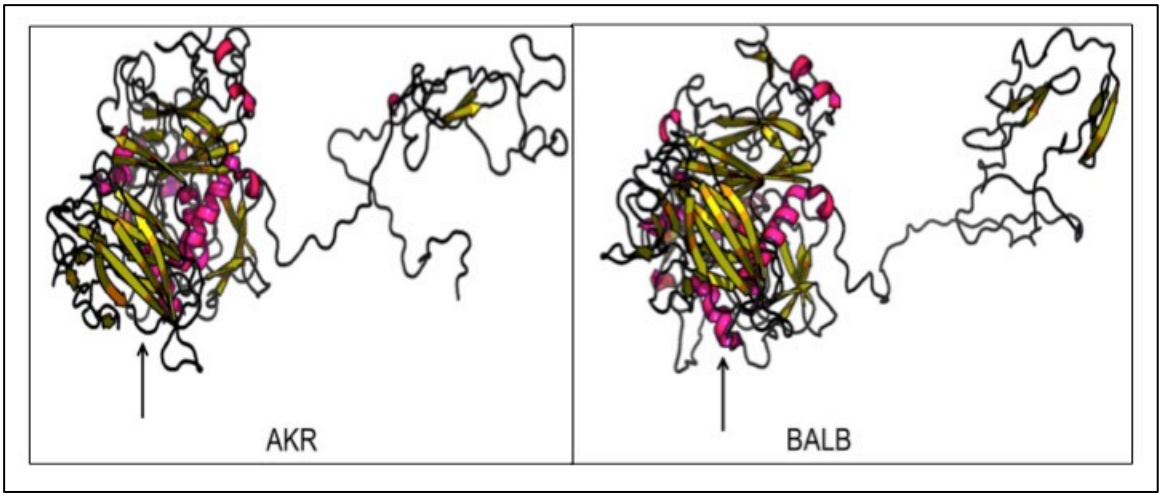

Figure 4.11 *Predicted 3D structure of Mep1a protein plotted by RaptorX software. Left panel: AKR mouse, and right panel: BALB mouse. The predicted 3D structure is noticeably different between the two mouse strains, specifically for the region pointed by the arrow.*

4.3.3 Correlations of atherosclerotic lesion sizes with plasma lipid and glucose levels

The associations of atherosclerotic lesion sizes with fasting plasma lipid and glucose levels were evaluated using the F2 population (**Figure 4.12**). A significantly inverse correlation with HDL cholesterol levels was observed (r = −0.188; p = 0.0068). Those F2 mice with higher HDL cholesterol levels tended to develop smaller atherosclerotic lesions than those with lower levels. Nevertheless, HDL only accounted for 3.5% of the variance in atherosclerotic lesion sizes of the F2 population, as denoted by the $r2$ value, a measure of the correlation between the two variables. No significant correlations were found with plasma levels of non- HDL cholesterol (r = 0.022; p = 0.757), triglyceride (R = −0.093; p = 0.184), or glucose (r = 0.00021; p = 0.998).

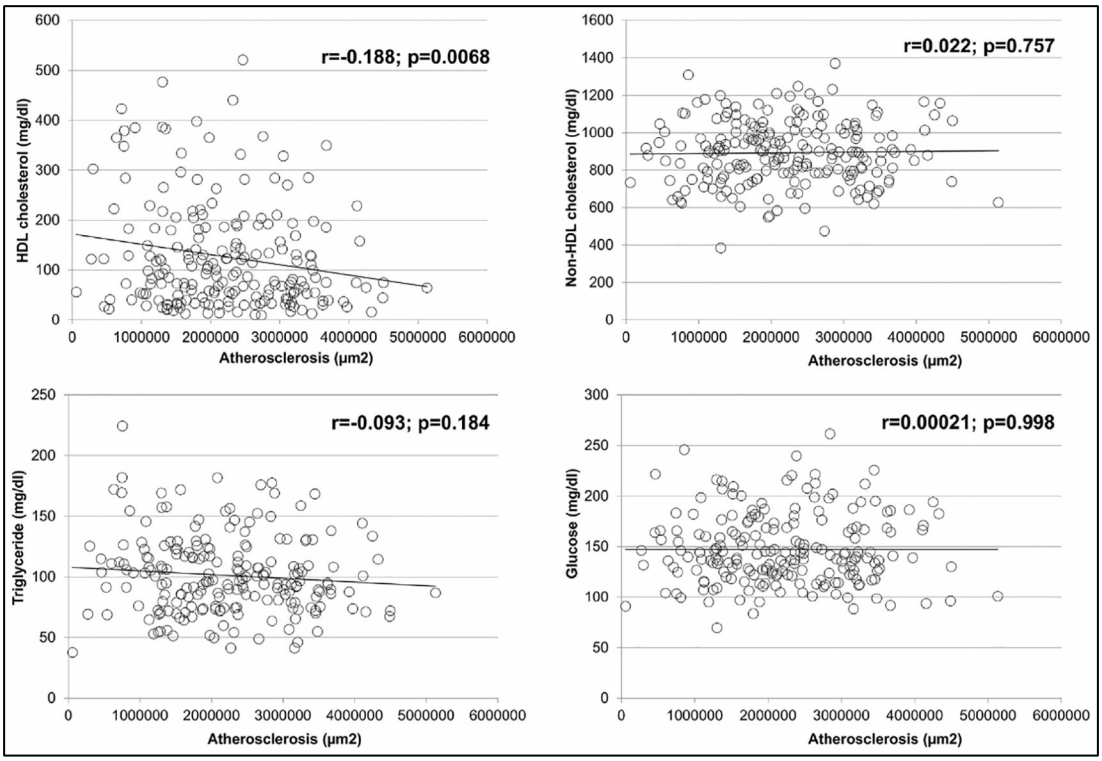

Figure 4.12 *Correlations of atherosclerotic lesion sizes with plasma levels of HDL, non-HDL cholesterol, triglyceride and glucose in the F2 population. Each point represents values of an individual F2 mouse. The correlation coefficient (r) and significance (p) are shown.*

4.4 Discussion

In this study, we performed QTL analysis using a female cohort derived from an intercross between BALB and SM *Apoe*$^{-/-}$ mouse strains and identified one significant QTL (*Ath49*) on chromosome 17 and six suggestive QTLs on chromosomes 1, 4, 5, and 8 for

atherosclerotic lesion sizes in the aorta. We then used bioinformatics-based tools to reduce the number of likely candidate genes for *Ath49* down to one. Moreover, plasma HDL cholesterol was found to explain 3.5% of the variance in lesion sizes of the F2 population.

Ath26 is an atherosclerosis susceptibility QTL on chromosome 17, initially mapped in an intercross between AKR-*Apoe*$^{-/-}$ and DBA/2-*Apoe*$^{-/-}$ mice[80]. In that cross, the QTL peaked at 36 or 47 Mb, depending on whether or not the lesion data were log-transferred. The confidence interval of Ath26 overlaps with the H2 complex (mouse major histocompatibility complex). The present study also mapped an atherosclerosis susceptibility QTL to the same region, but because it was mapped with a cross derived from distinct parental strains, the QTL was given a new designation, *Ath49*, according to the guideline established by the International Committee on Standardized Genetic Nomenclature for Mice (http://www.informatics.jax.org/mgihome/nomen/gene.shtml). Direct evidence supporting an importance role for the H2 complex in regulation of atherosclerosis susceptibility comes from our study of C3.SW-*H2bApoe*$^{-/-}$ mice[113]. C3H/HeJ mice are extremely resistant to atherosclerosis even on the Apoe null background. However, after their *H2k* allele was replaced with *H2b* allele and bone marrow was reconstituted with syngeneic or allogenic marrow, these mice exhibited a 21-fold increase in atherosclerotic lesion size[113].

By examining gene expression and protein structure and function influenced by SNPs that were shared among the high allele strains (BALB, DBA/2) but different from those carried by the low allele strain (AKR), we were able to reduce the number of candidates down to a single gene in the *Ath49* interval. This analysis was based on the observation that genetic variations between mouse strains are primarily inherited from their progenitors M. m. musculus and M. m. domesticus[168]. Mouse strains carrying an allele that results in a high value for a given trait often share the same ancestral allele harboring

the underlying causal gene, while strains carrying an allele that produces a low value for the trait share a different ancestral allele[169]. A limitation of this analysis was that the causal gene would be missed if the QTL arose from a more recent mutation.

Because a QTL arises from changes in the function or the quantity of a gene product, we took a two-armed approach to investigate these two possibilities separately. For influences on gene expression, by focusing those genes in the aorta and liver that are different between the high and low alleles, have an effect on atherosclerosis, and whose expression is influenced by our QTL locus we were able to dramatically reduce the number of candidate genes to three in the aorta and two in the liver. The genotypes of the high and low allele strains did not differ for any of these variants, so none of these candidate genes are responsible for *Ath49*. However, this strategy we have developed is applicable to any overlapping QTL locus and can provide valuable insight for QTLs whose causal genes remain undiscovered.

For influences on protein structure and function, we focused on genes that carry a non-synonymous coding SNP or a SNP in the upstream regulatory region segregating between the high allele and low allele strains of the QTL crosses. *Tnfrsf21*, *Adgrf1*, *Adgrf5*, *Mep1α*, and *Tdrd6* are located underneath the linkage peak of *Ath49*, and contain one or more non-synonymous SNPs or a SNP in the upstream regulatory region segregating between the high allele and low allele parental strains. *Adgrf1*, *Adgrf5*, and *Mep1α* contain amino acid changes predicted to influence the protein's structure, however only those in *Mep1α* were shown to be potentially affecting protein structure and function. From this, we believe *Mep1α* should be prioritized for future studies on its influence in atherosclerosis.

Association of the MHC complex with coronary artery disease has been observed in recent human GWAS studies[61,170]. The present analysis of human GWAS meta-analysis dataset also revealed multiple variants in the MHC region associated with coronary artery

disease. KCNK5 has been suggested to be an underlying causal gene contributing to coronary artery disease[61]. However, as KCNK5 is located on mouse chromosome 14 but not 17, obviously it is not the causal genes of *Ath49*. Furthermore, the chromosomal region harboring KCNK5 has only shown a suggestive linkage with atherosclerotic lesions[73].

In this study, we identified multiple suggestive QTLs for atherosclerotic lesion sizes, including the ones on chromosome 5. The LOD score plots for chromosome 5 revealed two distinct peaks, suggesting the existence of two QTLs for atherosclerosis. The distal QTL was overlapping with *Ath42* mapped in a B6 × BALB *Apoe*$^{-/-}$ intercross[76]. The proximal QTL replicated a suggestive QTL mapped in a B6 × C3H Apoe$^{-/-}$ intercross[83] and was named *Ath50*. Naming a suggestive locus is considered appropriate if it is repeatedly observed[66]. Promising candidate genes for this QTL include *Lrrc66*, *Pdgfra*, *Epha5*, *Cenpc1*, *Uba6*, and *Tmprss11g*. These genes are located underneath the linkage peak, and contain one or more SNPs segregating between the high allele and low allele parental strains of the QTL crosses (data not shown). Moreover, the human syntenic region harboring these genes has been shown to be associated with coronary heart disease[61].

A slight but statistically significant correlation was observed between HDL cholesterol levels and aortic atherosclerotic lesion sizes in the F2 cross. Similar findings have been observed with other mouse crosses[77,79,128]. These findings are consistent with the observation made from Mendelian randomization studies of HDL for role in human coronary heart disease[171,172]. Although Mendelian randomization studies of blood lipids have suggested a role for triglyceride and LDL cholesterol in human coronary heart disease, no significant correlations with atherosclerotic lesions were observed in the present cross or previous crosses [88,122].

In summary, we have identified a major QTL for atherosclerosis in the mouse MHC region using a segregating F2 population, and further applied bioinformatics-based tools to

define underlying candidate genes. As the MHC region harbors genes that play an important role in both innate and adaptive immunity, our findings highlight the significance of inflammation in atherogenesis and its potential for developing anti-atherosclerotic therapy.

4.5 Methods

Mice

BALB-Apoe$^{-/-}$ and SM-*Apoe*$^{-/-}$ mice were made in our laboratory using the congenic breeding method as previously reported[114]. The creation of a female F2 cohort from the two Apoe$^{-/-}$ strains was recently described[159] (**Chapter 2**). The animals were weaned at 3 weeks of age, and at 6 weeks of age switched onto a Western diet. After 12 weeks of Western diet, mice were euthanized for assessment of atherosclerotic lesion formation in the aorta.

Quantifiation of aortic atherosclerosis

Atherosclerotic lesion areas in the aortic root of mice were measured as previously reported[88]. Lesion areas were measured on oil red O stained sections using Zeiss AxioVision 4.8 software. The eight largest lesion areas were added up for each mouse and this sum was used for statistical analysis.

Measurements of plasma glucose and lipid levels

Plasma total cholesterol, HDL cholesterol, triglyceride and glucose levels were measured using commercial kits as reported[173]. Non-HDL cholesterol was calculated as the difference between total and HDL cholesterol.

Genotyping

F2 mice were genotyped using the Illumina LD linkage panel, as reported[159] (**Chapter 2**). Microsatellite markers were typed by PCR for regions of chromosome 8 that were not covered by informative SNP markers. A total of 149 markers were included in QTL analysis.

Statistical analysis

QTL analysis was performed using J/qtl and MapManager QTX software, as reported[159] (**Chapter 2**). LOD threshold values were determined from 1000 permutations of the observed data.

Prioritization of candidate genes

Bioinformatics-based tools were used to prioritize candidate genes for significant atherosclerosis QTL that was mapped in two or more crosses derived from different parental strains. Two separate analyses were performed, one for SNPs influencing gene expression and one for SNPs influencing protein structure and function. For the gene expression analysis, likely candidate genes were those who showed a significant change in gene expression between the high and low alleles in the aorta and the liver, were significantly correlated with aortic lesion size, and had a significant eQTL linking the gene expression to our locus of interest. Data used for this analysis was obtained from publicly available sources (aorta and liver gene expression and correlation with aortic lesion size: GEO accession = GSE66570)[91] (aorta and liver eQTL data (https://systems.genetics.ucla.edu/data/hmdp).

For the protein structure and function analysis, likely candidate genes were defined as those containing a non-synonymous SNP in a coding region, and this SNP was shared by the parental strains carrying the high allele but different from the one shared by the parental strains carrying the low allele of a QTL, as reported[79]. Analysis was performed

using a combination of SNP sources, including the Sanger Mouse Genomes Project, Mouse Phenome Database, and Ensembl. SNP frequency was determined from information provided by the WTSI Mouse Genomes Project in the Ensembl browser. Amino acid conservation was determined from information provided by the UCSC Genome Browser. Web-based software RaptorX was used to predict the potential impact of an amino acid substitution on the 3D structure of protein product (http://raptorx.uchicago.edu/ Structure_ Prediction/predict).

Chapter 5

Mep1α Affects Atherosclerosis Through Large-Scale Changes In Plaque Composition

Andrew T. Grainger and Weibin Shi

Manuscript in Preparation

5.1 Abstract

Atherosclerosis is a complex disease that is the primary cause of heart attack, ischemic stroke and peripheral arterial disease. We have recently discovered a novel genetic locus influencing aortic lesion size at the major histocompatibility complex (MHC) on mouse Chromosome 17, named *Ath49*. Through bioinformatics, we identified *Mep1α* as a gene at this locus likely influencing atherosclerosis. In a *Mep1α*$^{-/-}$.*Apoe*$^{-/-}$ double knockout mouse model, aortic lesion size was significantly reduced when mice were fed either chow or Western diet. Investigation of lesion stability showed a decrease in necrosis and fibrous cap thickness in *Mep1α*$^{-/-}$ mice and increase in overall trichrome stain area. Immunofluorescent staining uncovered a marked increase in macrophage and neutrophil content in *Mep1α*$^{-/-}$ mice. Investigation of plasma MCP-1, CXCL5, and MDA levels showed *Mep1α*$^{-/-}$ mice have a decrease in CXCL5 and MDA, suggesting a reduction in systemic inflammation and oxidative stress. In this study, we have discovered that *Mep1α* is a novel gene negatively influencing atherosclerosis through widespread influence on plaque composition.

5.2 Introduction

Atherosclerosis is the primary cause of heart attack, ischemic stroke and peripheral arterial disease. It is considered a complex disease resulting from interactions between environmental and genetic factors[146]. Numerous studies have highlighted the importance of genetic factors in atherosclerosis' pathogenesis, including prospective studies of twins, families and cohorts[46,47,48] and genome-wide association studies (GWAS)[62,61,59,60]. Based on the results of these studies, it is apparent that the majority of the genome's influence on heart disease is complex, probably involving many genes of small effect. The 304

independent variants associated at 5% FDR identify 243 loci that implicate pathways in blood vessel morphogenesis as well as lipid metabolism, nitric oxide signaling and inflammation attest to this fact[62]. However, all the loci identified account for only 21.2% of the genetic heritability of coronary heart disease[62], suggesting that many more loci remain to be discovered.

A complementary approach to human studies in the finding of novel genes and pathways involved in atherosclerosis is to study animal models. Animal models have the advantage of allowing strict control over environmental influence and accurate phenotypic characterization of atherosclerotic lesions. Apolipoprotein E-null (*Apoe*$^{-/-}$) and LDL receptor-null (*Lldlr*$^{-/-}$) mouse models reproduce all phases of atherosclerotic lesions seen in humans[71,72]. Over a dozen intercrosses or backcrosses have been generated from atherosclerosis-susceptible and atherosclerosis-resistant inbred strains carrying the *Apoe*$^{-/-}$ or *Lldlr*$^{-/-}$ gene, leading to identification of over 50 unique atherosclerosis susceptibility loci (http://www.informatics.jax.org/ allele).

We have recently performed one such study using an intercross between BALB/cJ.*Apoe*$^{-/-}$ and SM/J.*Apoe*$^{-/-}$ and discovered a novel genetic locus influencing aortic lesion size at the major histocompatibility complex (MHC) on mouse Chromosome 17, named *Ath49*[174] (**Chapter 4**). Bioinformatics analysis identified *Mep1α* as the primary candidate gene. *Mep1α* showed a large number of amino acid changes in its main metalloprotease domain and was the only one that showed predicted changes in its protein structure. For these reasons, *Mep1α* was prioritized for further investigation.

Mep1α is a metalloprotease involved in a large number of processes affecting inflammation and fibrosis[175,176,177,178,179,180,181,182], many of which have a direct effect on atherosclerosis[37]. Moreover inhibition of metalloproteases, including *Mep1α*, through administration of actinonin decreased atherosclerotic plaque formation in mice[183]. This

strongly suggests *Mep1α* is playing a direct role in atherosclerosis. Here, we investigate Mep1α's influence in atherosclerosis through the characterization of a *Mep1α*$^{-/-}$.*Apoe*$^{-/-}$ double knockout mouse model.

5.3 Results

5.3.1 Mep1α$^{-/-}$ mice have normal lipid and glucose levels but decreased aortic lesion size

The normal mouse model for atherosclerosis consists of *Apoe*$^{-/-}$ mice fed a Western diet for advanced lesion formation and a chow diet for early lesion formation. Therefore to asses the effects of *Mep1α* on atherosclerosis, *Mep1α*$^{-/-}$.*Apoe*$^{-/-}$ double knockout mice were generated and fed either a chow or Western diet for 12 weeks. After 12 weeks, plasma lipid and glucose levels were measured from both C57BL/6J (B6) *Mep1α*$^{-/-}$.*Apoe*$^{-/-}$ and normal B6.*Apoe*$^{-/-}$ mice. No difference in HDL, Non-HDL, or fasting glucose was observed in either chow or Western diet mice, while triglycerides were lower in double knockout mice on chow diet (**Figure 5.1A-D**). Aortic lesion size was markedly smaller in *Mep1α*$^{-/-}$ mice compared to the controls (**Figure 5.1E-F**). On chow diet, B6 mice had an average total lesion size of $2.95x10^5 \pm 6.09x10^4$ while *Mep1α*$^{-/-}$ mice had an average total lesion size of $9.78x10^4 \pm 1.89x10^4$, a 66.9% reduction (B6 n = 28, *Mep1α*$^{-/-}$ n = 13; p = 0.0042; Male + Female). On Western diet, the average of the top 5 lesion areas was used to assess aortic lesion changes. For females, B6 mice showed an average lesion size of $5.82x10^5 \pm 2.02x10^4$ while *Mep1α*$^{-/-}$ mice had a significantly smaller average lesion size of $4.12x10^5 \pm 3.62x10^4$, a 25.5% reduction (B6 n = 39, *Mep1α*$^{-/-}$ n = 15; p = 0.0015). For males, B6 mice showed an average lesion size of $3.83x10^5 \pm 3.32x10^4$ while *Mep1α*$^{-/-}$ mice had a smaller average lesion

size of $3.83x10^5 \pm 4.22x10^4$ (B6 n = 23, *Mep1α*$^{-/-}$ n = 11; p = 0.59). However this decrease in lesion area was not significant.

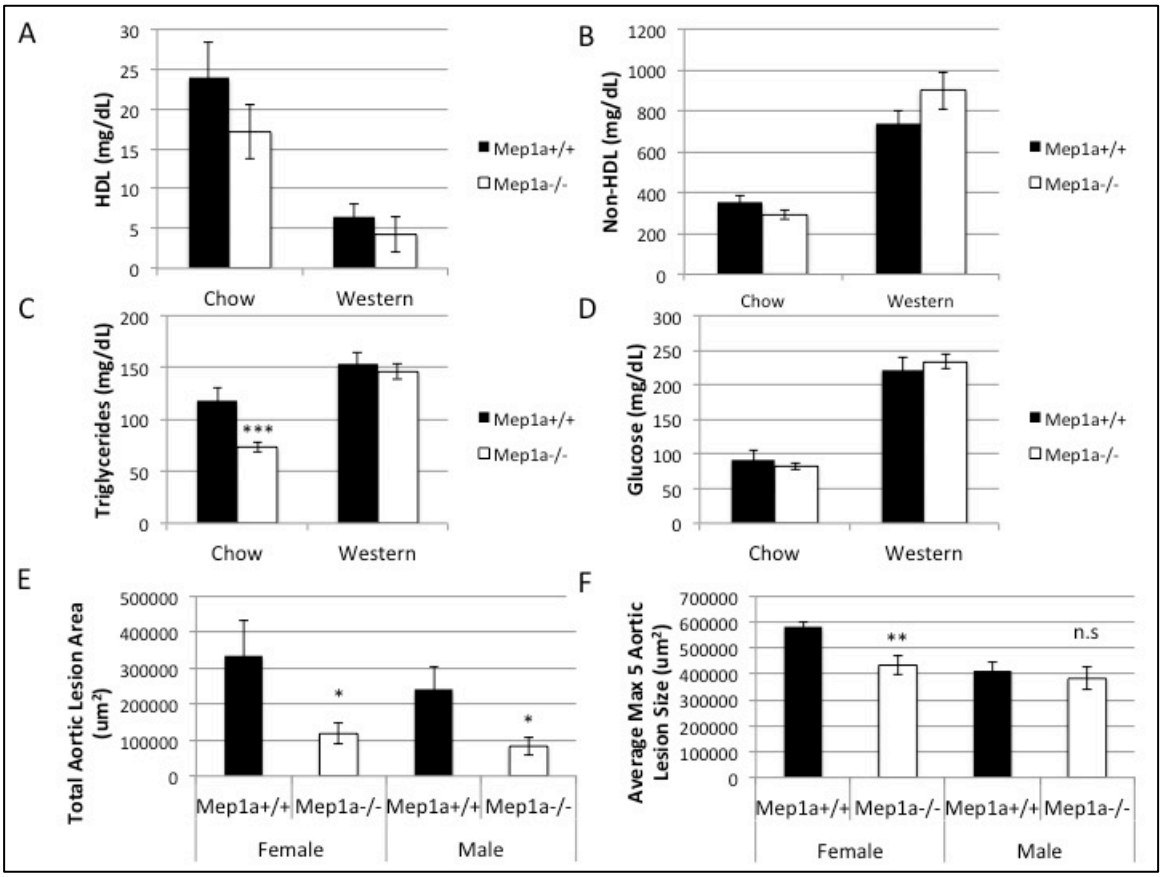

Figure 5.1 *Mep1α*$^{-/-}$ *mice have normal lipid and glucose levels but decreased aortic lesion size. A) Average HDL levels from Mep1α*$^{+/+}$ *and Mep1α*$^{-/-}$ *mice fed chow (n = 10 per group; p = 0.27) and western diet (n = 10 per group; p = 0.45) B) Average non-HDL levels from Mep1α*$^{+/+}$ *and Mep1α*$^{-/-}$ *mice fed chow (n = 10 per group; p = 0.14) and western diet (n = 10 per group; p = 0.16) C) Average triglycerides levels from Mep1α*$^{+/+}$ *and Mep1α*$^{-/-}$ *mice fed chow (n = 10 per group; p = 0.0041) and western diet (n = 10 per group; p = 0.58) D) Average glucose levels from Mep1α*$^{+/+}$ *and Mep1α*$^{-/-}$ *mice fed*

chow (n = 10 per group; p = 0.53) and western diet (n = 10 per group; p = 0.58) E) Total aortic lesion area quantification for mice fed chow diet, either male Mep1α$^{+/+}$ n = 12, Mep1α$^{-/-}$ n = 7 mice; p = 0.032) or female (Mep1α$^{+/+}$ n = 16, Mep1α$^{-/-}$ n = 5 mice; p = 0.046) . F) Average lesion area of the top five lesion sizes for mice fed western diet, either Female (Mep1α$^{+/+}$ n = 39, Mep1α$^{-/-}$ n = 15 mice; p = 0.0015) or male (Mep1α$^{+/+}$ n = 23, Mep1α$^{-/-}$ n = 11 mice; p = 0.59).

5.3.2 Lesion Stability Assessment in Mep1α$^{-/-}$ mice

Mep1α$^{-/-}$ mice have decreased collagen content in the dermis, and we wanted to assess whether this persists in aortic lesions as changes in collagen content have the potential to affect overall lesion stability[178]. The standard hallmarks of plaque stability were assessed using hematoxylin and eosin (H&E) and Masson's trichrome staining on aortic lesions from mice fed western diet (**Figure 5.2**). H&E staining appeared similar between the *Mep1α$^{-/-}$* mice and normal B6 mice (**Figure 5.2A-B**), however Masson's trichrome staining appeared markedly different (**Figure 5.2C-D**). The percentage of necrotic lesion area was significantly decreased in *Mep1α$^{-/-}$* mice (B6 = 25.5% ± 2.2%, *Mep1α$^{-/-}$* = 9.3% ± 1.2%; $p < 0.0001$), whereas the overall unstained area in H&E stained lesions remained the same (B6 = 33.6% ± 1.2%, *Mep1α$^{-/-}$* = 35.45% ± 1.2%; p = 0.30) (**Figure 5.2E-F**). Interestingly, this unstained area significantly decreased in the *Mep1α$^{-/-}$* mice in trichrome-stained lesions, while the unstained area of the B6 mice remained the same (B6 = 33.2% ± 2.5%, *Mep1α$^{-/-}$* = 17.11% ± 1.3%; $p < 0.0001$) (**Figure 5.2G**). The color present in these previously unstained locations within *Mep1α$^{-/-}$* lesions was a deep purple, suggesting a mixture of blue collagen and red muscle fiber stains (**Figure 5.2C-D**).

Quantification of average fibrous cap thickness showed an overall decrease in *Mep1α*$^{-/-}$ mice (B6 = 30.6 ± 1.3 um, *Mep1α*$^{-/-}$ = 21.6 ± 0.8 um; p <0.0001). The minimum (B6 = 9.0 ± 0.5 um, *Mep1α*$^{-/-}$ = 8.9 ± 0.5 um; p = 0.93) and maximum (B6 = 42.1 ± 1.8 um, *Mep1α*$^{-/-}$ = 43.5 ± 2.3 um; p = 0.62) fibrous cap thicknesses remain unchanged (**Figure 5.2H-J**).

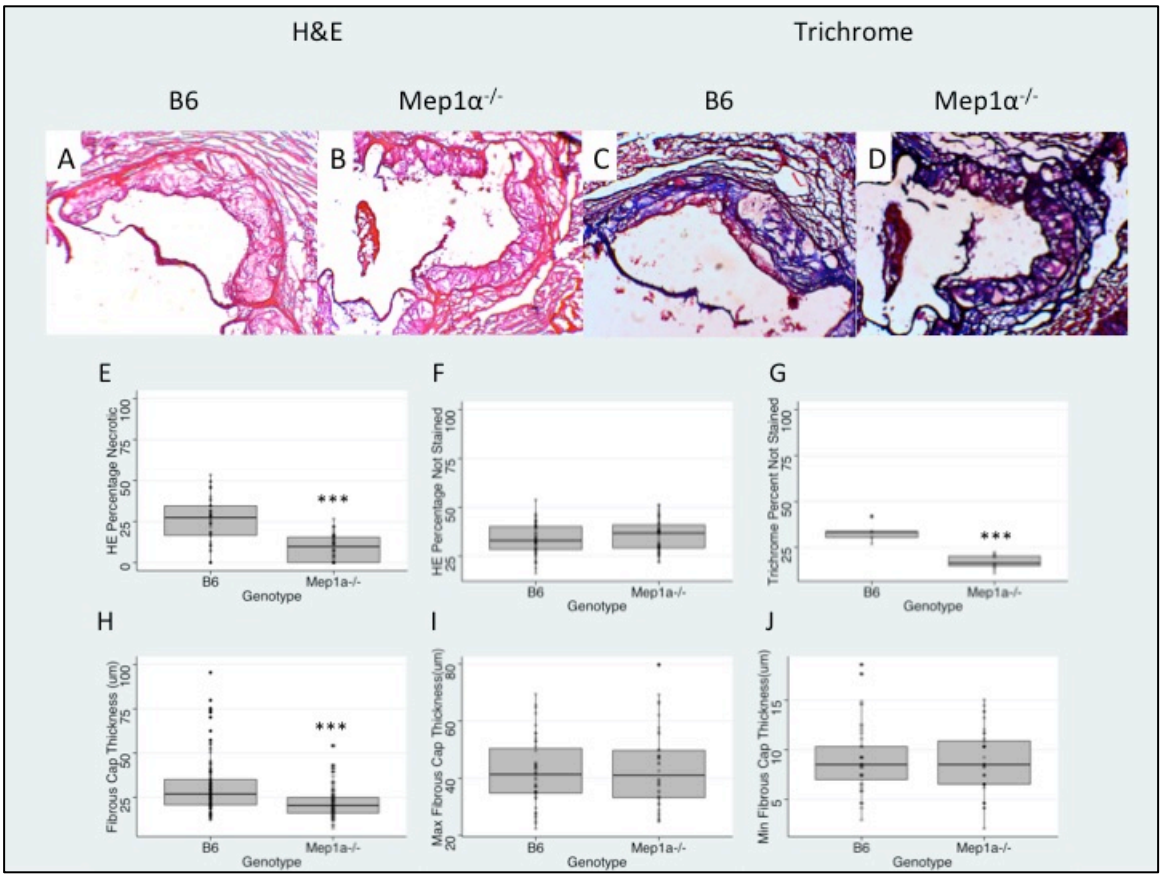

Figure 5.2 *Mep1α*$^{-/-}$ *mice have changes plaque stability. A/B) Representative images of HE staining for B6 (A) and Mep1α*$^{-/-}$ *(B) mouse aorta fed western diet. C/D) Representative images of trichrome staining for B6 (C) and Mep1α*$^{-/-}$ *mice (D). Quantification of classic plaque stability metrics were performed using both HE and trichrome-stained sections. Metrics pertaining to necrosis included percentage of*

lesion area in HE stained slides that was necrotic (E) (B6 n=43, Mep1α$^{-/-}$ n=44 images, n=4 mice per group; p < 0.0001), percentage unstained (F)(B6 n=43, Mep1α$^{-/-}$ n=44 images, n=4 mice per group; p = 0.2954), and percentage of lesion area in trichrome stained slides that was necrotic (G) (B6 n=5, Mep1α$^{-/-}$ n=9 images, n = 2 B6 mice, n = 4 Mep1α$^{-/-}$ mice; p , 0.0001). Metrics pertaining to fibrous cap formation included average fibrous cap thickness in HE stained sections (H) (B6 n= 129, Mep1α$^{-/-}$ n = 108 measurements, two outer and one middle measurement per image)(n = 4 mice per group)(p < 0.0001), maximum fibrous cap thickness (I) (B6 n = 43, Mep1α$^{-/-}$ n = 36 images, n = 4 mice per group; p = 0.6155), and minimum fibrous cap thickness (J) (B6 n = 43, Mep1α$^{-/-}$ n = 36 images, n = 4 mice per group; p = 0.9321).

5.3.3 Mep1α$^{-/-}$ lesions have increased macrophage content

Mep1α is located proximal to the MHC and can affect cytokine production and macrophage transmigration[176]. Therefore we wanted to investigate the role *Mep1α* plays in the aortic lesions' composition, particularly in respect to immune cell content. Initial immunofluorescent staining showed a large amount of *Mep1α* in the chow diet lesions from normal B6 mice as well as significant staining in Western diet lesions (**Figure 5.3A, C**). Both the chow and Western diet lesions of *Mep1α*$^{-/-}$ show an increase in macrophage content (**Figure 5.3B, D**). The cells within Western diet *Mep1α*$^{-/-}$ lesions are on average 76.1% ± 6.0% macrophages while normal B6 lesions are on average 20.2% ± 3.6% macrophages, a 376% increase (P < 0.0001) (**Figure 5.4A-C**). No change in smooth muscle cell content was observed between normal B6 and *Mep1α*$^{-/-}$ lesions (B6 = 5.7% ± 4.6%, *Mep1α*$^{-/-}$ = 12.4% ± 2.9%; p = 0.14) (**Figure 5.4D**). Smooth muscle cells in the adjacent medial layer of the aorta

showed a trend towards an increase in *Mep1α*$^{-/-}$ mice (B6 = 13.6% ± 4.9%, *Mep1α*$^{-/-}$ = 33.1% ± 11.7%; p = 0.056)(**Figure 5.4E**).

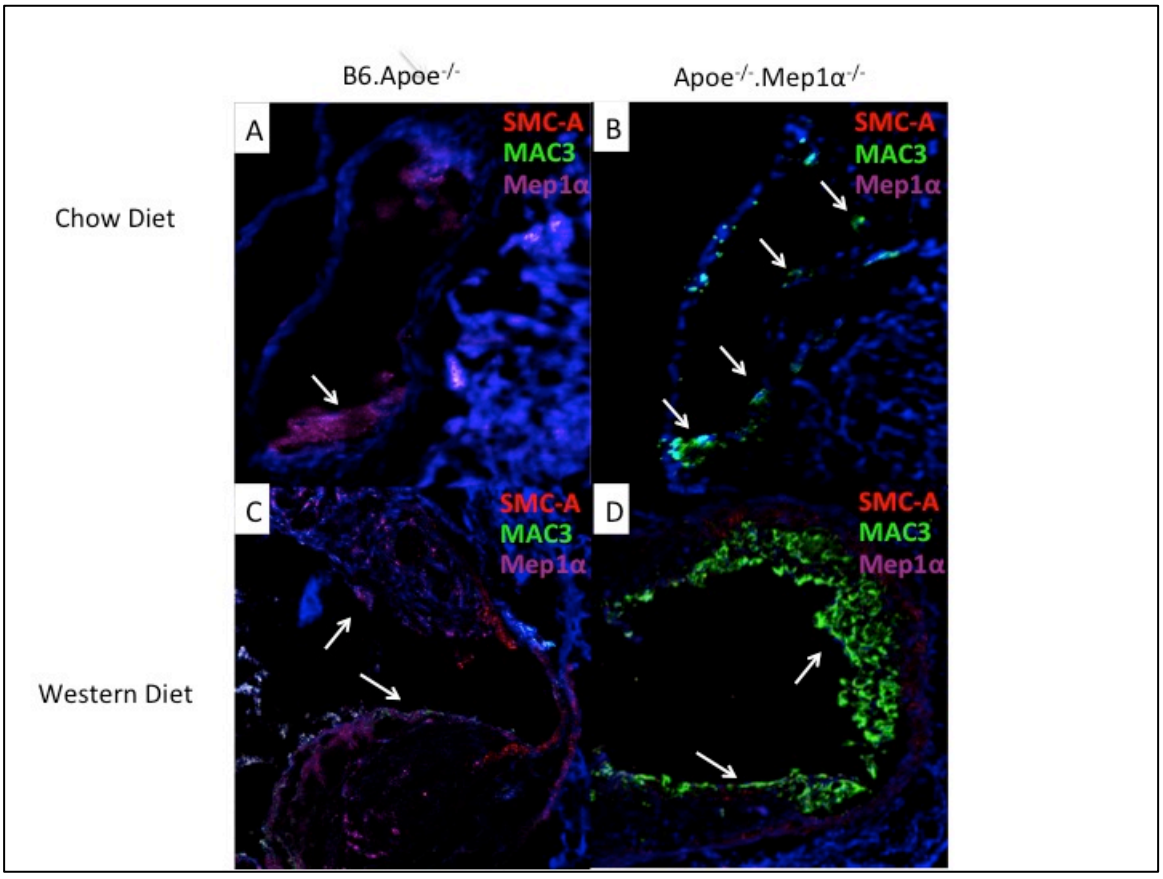

Figure 5.3 *Lesion morphology differences of B6 and Mep1α*$^{-/-}$ *mice. Changes in lesion cell types from B6 mice (A) and Mep1α*$^{-/-}$ *(B) fed chow diet and from B6 mice (C) and Mep1α*$^{-/-}$ *(D) fed western diet. Green = macrophage, red = smooth muscle cells, magenta = Mep1α. White arrows indicate presence of aortic lesions*

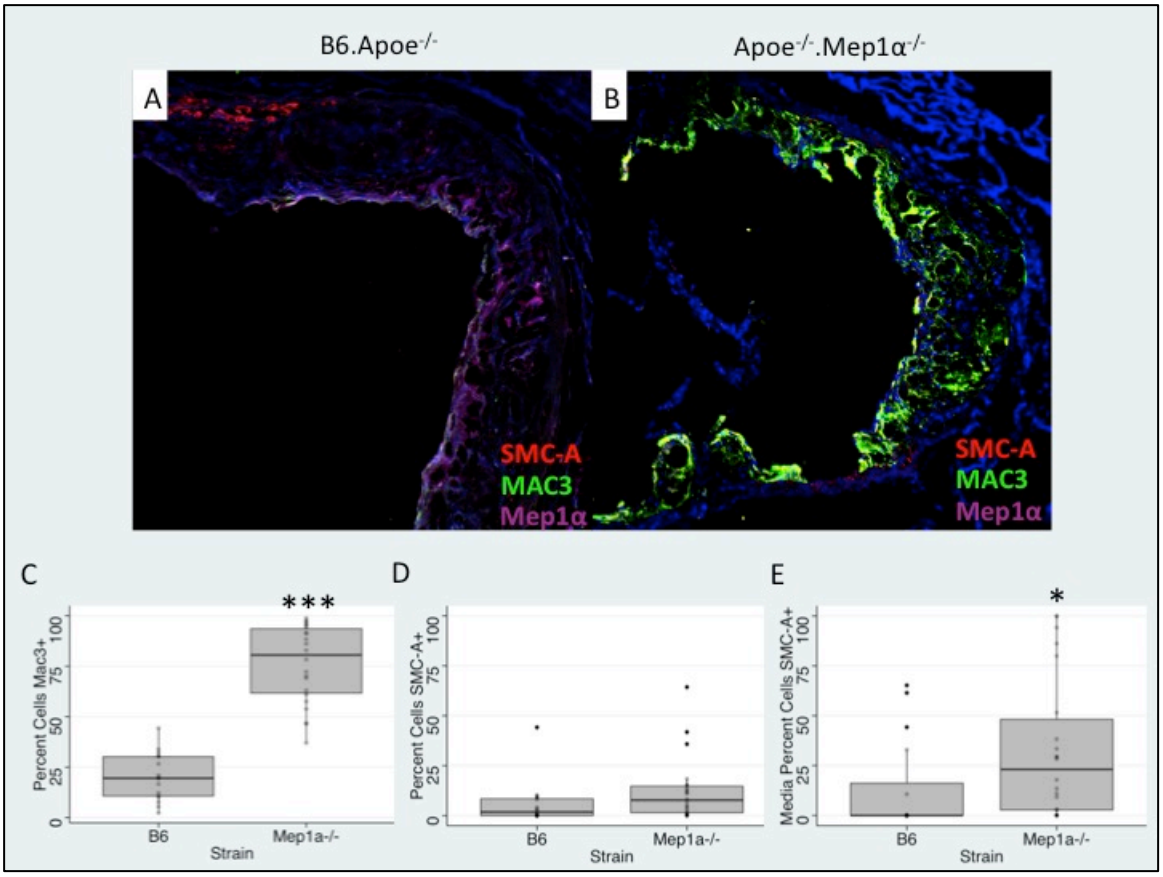

Figure 5.4 *Mep1α$^{-/-}$ lesions have increased macrophage content. Representative immunofluorescence images of a normal B6 mouse (A) and a Mep1α$^{-/-}$ mouse (B) (Red = SMC-A, Green = Macrophage, Magenta = Mep1α, Blue = DAPI). Quantification of cellular components from macrophage (C) (n = 9 B6 mice with n = 16 images; n = 9 Mep1α$^{-/-}$ mice with n = 22 images; p <0.0001) and smooth muscle cells (D) (p = 0.14) in aortic lesions and smooth muscle cells in the medial layer (E) (p = 0.048).*

5.3.4 Mep1α$^{-/-}$ lesions have increased neutrophil content

With a three-fold increase in macrophage content within the *Mep1α$^{-/-}$* lesions, we investigated whether other immune-related cell types were altered. Neutrophil content was

significantly increased in *Mep1α*$^{-/-}$ lesions, with 5.4% ± 1.3% of the total lesion area being positive for neutrophil marker Ly6g compared to 0.4% ± 0.2% of the total lesion area for normal B6 mice, a 12-fold increase (p = 0.001)(**Figure 5.5A-G**). Staining for the neutrophil recruitment cytokine CXCL5 showed significant co-localization with Ly6g staining, however quantification showed the total amount of CXCL5 present in aortic lesions remained unchanged (B6 = 17.8% ± 5.7%, *Mep1α*$^{-/-}$ = 13.2% ± 2.7%; p = 0.48)(**Figure 5.5H**).

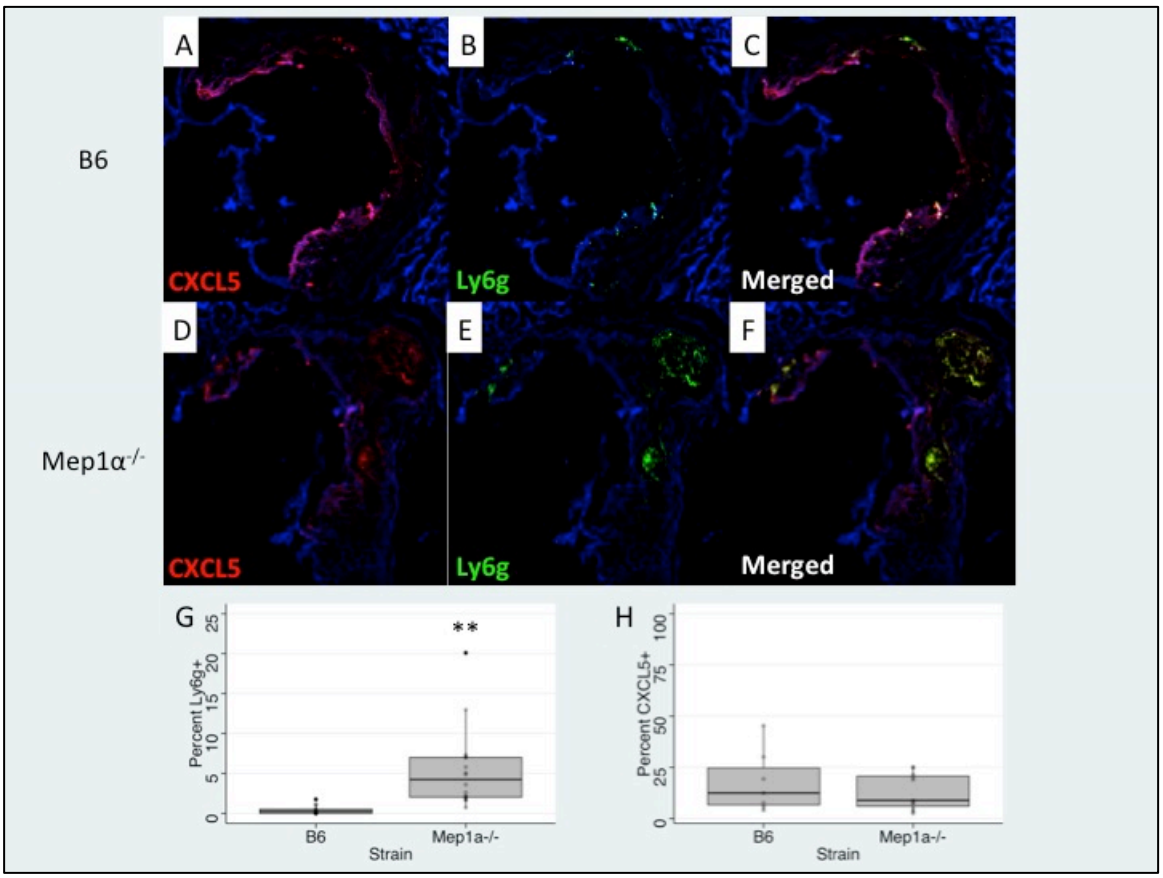

Figure 5.5 *Mep1α*$^{-/-}$ *lesions have increased neutrophil content. CXCL5 (red) and ly6g (green) presence and co-localization in B6 (A-C) and Mep1a*$^{-/-}$ *(D-F) mice. Quantification of percentage lesion area of Ly6g (n = 6 B6 mice with n = 13 images; n =*

7 Mep1α$^{-/-}$ mice with n = 16 images; p = 0.0012) (G) and CXCL5 (n = 4 B6 mice with n = 7 images; n = 5 Mep1α$^{-/-}$ mice with n = 11 images; p = 0.47) (H) staining.

5.3.5 Mep1α$^{-/-}$ mice have decreased plasma CXCL5 and oxidative stress

Because of the increase in macrophages and neutrophils in aortic lesions, we performed ELISA assays on chow and Western diet plasma to test for MCP-1, the primary macrophage recruitment cytokine, and CXCL5, the neutrophil recruitment cytokine. Surprisingly, there was no change in plasma MCP-1 levels on either chow (B6 = 14.1 ± 8.7 mg/dl, *Mep1α$^{-/-}$* = 20.4 ± 11.5 mg/dl; n = 10 mice per group; p = 0.22) or Western diet (B6 = 75.7 ± 45.4 mg/dl, *Mep1α$^{-/-}$* = 72.0 ± 6.0 mg/dl; n = 10 mice per group; p = 0.82) (**Figure 5.6A**). *Mep1α$^{-/-}$* mice showed a 75% decrease in plasma CXCL5 on chow diet (B6 = 1855.0 ± 1292.43 mg/dl, *Mep1α$^{-/-}$* = 451.0 ± 250.8 mg/dl; n = 10 mice per group; p = 0.008) and 57% decrease on Western diet (B6 = 1385.6 ± 649.3 mg/dl, *Mep1α$^{-/-}$* = 595.6 ± 582.4 mg/dl; n = 10 mice per group; p = 0.01) (**Figure 5.6B**). MDA, a marker for oxidative stress, showed a significant decrease in *Mep1α$^{-/-}$* mice for both chow diet (B6 = 39.6 ± 2.6 mg/dl, *Mep1α$^{-/-}$* = 33.2 ± 0.7 mg/dl; n = 10 per group; p = 0.04) and Western diet (B6 = 76.7 ± 6.2 mg/dl, *Mep1α$^{-/-}$* = 53.4 ± 3.1 mg/dl; n = 10 per group; p = 0.006) (**Figure 5.7**).

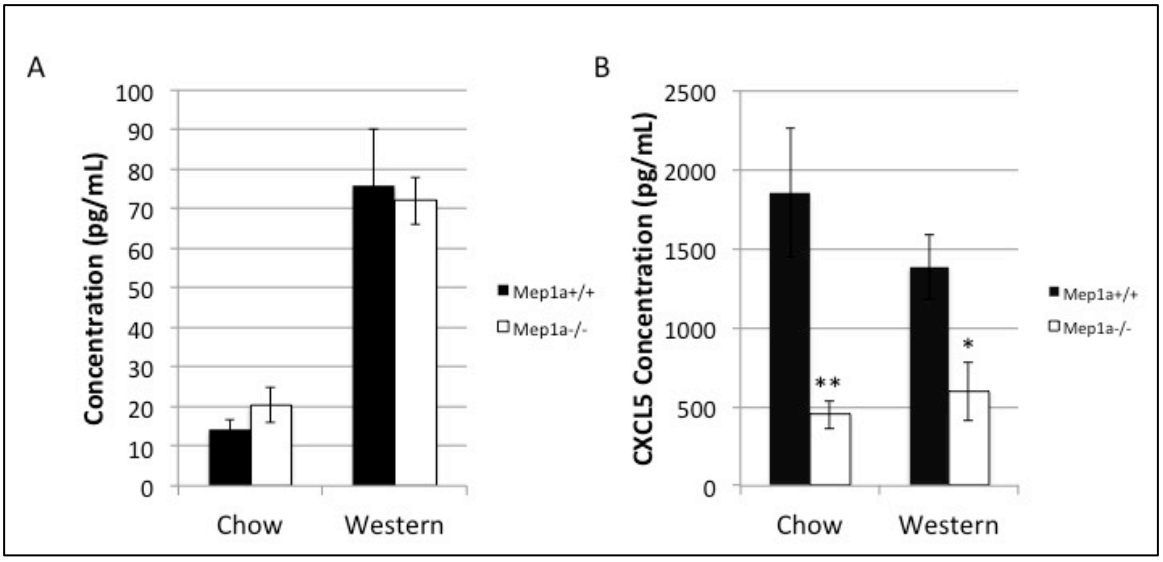

Figure 5.6 *Mep1α$^{-/-}$ mice show decreased circulating CXCL5. Plasma cytokine levels for main recruiting factor of macrophage (MCP-1) (A) (n = 10 per group; chow: p = 0.22, western: p = 0.82) and neutrophils (CXCL5) (B) from mice fed either chow or Western diet (n = 10 per group; chow: p = 0.008, western: p = 0.01).*

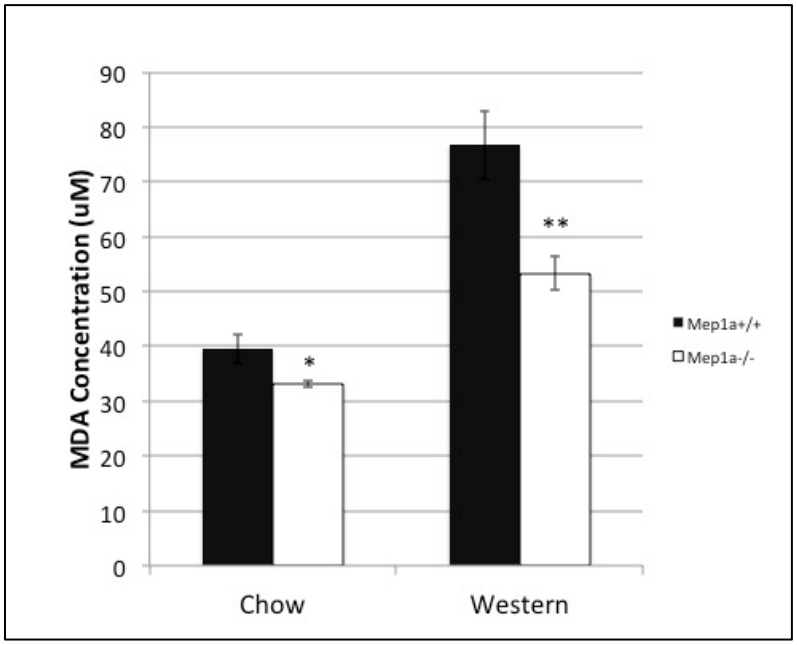

Figure 5.7 *Mep1α$^{-/-}$ show decreased oxidative stress. Plasma MDA levels measured from chow and Western diet mice (n = 10 per group; chow: p = 0.04, western: p = 0.006).*

5.4 Discussion

In this study, we used a *Mep1α$^{-/-}$.Apoe$^{-/-}$* double knockout mouse to investigate *Mep1α*'s involvement in both early and late-stage lesion formation. Measurement of plasma lipid and glucose levels showed no changes between normal B6. *Apoe$^{-/-}$* mice and double knockout mice, suggesting that *Mep1α* has no influence over lipid or glucose production. Despite normal lipid and glucose levels, double knockout mice fed chow diet and female mice on Western diet exhibited significantly smaller lesions. This indicates that *Mep1α* is involved in lesion formation, and that we have discovered a novel influencer of atherosclerosis.

H&E staining of Western diet-fed B6. *Apoe*$^{-/-}$ and double knockout mice showed that *Mep1α*$^{-/-}$ mice had decreased necrosis and fibrotic cap thickness. Normal stable plaques have less necrosis with a thicker fibrous cap[44], so this data is conflicting in determining overall lesion stability. To add to the complexity, we observed an increase in overall positive trichrome stain area in the *Mep1α*$^{-/-}$ lesions. This increased stain area was primarily occupied by a purple-colored stain, which indicates a mixture of the blue collagen stain and red muscle fiber stain. This stain was unique to the *Mep1α*$^{-/-}$ mice. *Mep1α* is predicted to be capable of degrading a large number of extracellular matrix (ECM)-related proteins[184], so we hypothesize that this stain is contains ECM-related proteins that are either being produced or not being degraded in the absence of *Mep1α*.

Immunoflourescent staining for *Mep1α* in B6. *Apoe*$^{-/-}$ mice showed a large amount present in chow diet, or early-stage, lesions. This suggests *Mep1α* plays an integral role in early lesion formation. The *Mep1α* stain persisted in the Western diet, or late-stage, lesions, indicating that *Mep1α* plays potential roles throughout the progression of atherosclerosis.

We observed a marked increase in macrophage and neutrophil content within the *Mep1α*$^{-/-}$ plaques. Because the neutrophil stain resided exclusively in areas of necrosis, we hypothesize that the necrosis we are observing in *Mep1α*$^{-/-}$ plaques is due to dead neutrophils and not dead macrophages, as is normally the case.

We observed no change in local CXCL5 abundance in the aortic lesions, however the location of the stain differed between B6 and *Mep1α*$^{-/-}$ mice. The CXCL5 localizes proximal to the endothelial cells in B6 lesions, while it exclusively co-localizes with Ly6g stain in *Mep1α*$^{-/-}$ mice. Despite a similar amount of CXCL5 in aortic lesions, the systemic plasma CXCL5 was significantly decreased. Plasma MCP-1 was unchanged, while the local macrophage content of the *Mep1α* lesions was significantly higher. A marked decrease in oxidative stress was also observed in the *Mep1α*$^{-/-}$ mice. These findings suggest that systemic inflammation is

decreased. We hypothesize that this decrease is likely due to the decrease in expression of specific cytokines. These findings indicate that *Mep1α*'s involvement in the lesion microenvironment is a local phenomenon and not indicative of its influences on systemic inflammation.

Because *Mep1α* is predicted to have over 150 potential substrates[184], determining what the substrates are in an atherosclerotic setting will be difficult. *Mep1α* is cleaved and can be used both intracellularly or as an extracellular aggregate, thus substrates in early and late-staged lesions are potentially numerous[175]. Examples of possible substrates include those that both directly affect inflammation and have been proven to influence atherosclerosis. Notable candidates that fit these criteria include *IL-1β* and *IL-6R* [181,37]. Additional candidates consist of proteins that are predicted to be degraded by *Mep1α*, are relevant to ECM production, and influence atherosclerosis. These candidates include *Nidogen-1*[185], *Cadherin-2*[186], *Tenascin-C*[187], *Syndecan-1*[188,189], Syndecan-4[190,191], TGF-β1[37], Collagen-*α2 (Collagen (I))*[192,193], Collagen-*α1 (Collagen (XVIII))*[192,193], *CTGF*[194], *MMP1*[44], and *ADAM9*[195,196]. These candidates are expressed in the three cell types with the largest impact on atherosclerotic lesion formation: the vascular smooth muscle cells, endothelial cells, and macrophages[27]. We hypothesize that part of *Mep1α*'s influence on atherosclerosis is likely through degradation of a subset of these proteins.

In summary, we discovered that *Mep1α* is a novel gene that is negatively affecting atherosclerosis. We have found that the plaque environment has significantly changed through differences in the hallmarks of lesion stability and in the aortic lesion's cellular composition. Currently, the specific cell types and pathways through which *Mep1α* is acting remains unknown. Despite this, *Mep1α* poses a promising potential target for therapeutics for prevention of atherosclerosis.

5.5 Methods

Ethics statement. All procedures were in accordance with current National Institutes of Health guidelines (https://grants.nih.gov/grants/olaw/Guide-for-the-Care-and-use-of-laboratory-animals.pdf) and approved by the institutional Animal Care and Use Committee. Blood was drawn from the retro-orbital plexus of overnight fasted mice with the animals under isoflurane anesthesia.

Animals, experimental design and procedures. *Mep1α*$^{-/-}$*Apoe*$^{-/-}$ double knockout mice were created through the crossing of C57BL/6J *Mep1α*$^{-/-}$ mice to C57BL/6J *Apoe*$^{-/-}$ mice. C57BL/6J *Mep1α*$^{-/-}$ mice were provided by Christoph Becker-Pauly at the University of Kiel. Quantitative PCR and western blotting were performed to validate double knockout mice did not express *Mep1α* (**Figure 5.8**). Mice were weaned at 3 weeks of age onto a rodent chow diet. At 6 weeks of age, the Western diet mouse group were started on a Western diet containing 21% fat, 34.1% sucrose, 0.15% cholesterol, and 19.5% casein *by weight* (Harlan Laboratories, TD 88137) and maintained on the diet for 12 weeks. At 6 weeks of age, the chow diet group remained on the rodent chow diet for 12 weeks. Mice were bled twice: once before initiation of the Western diet and once at the end of the 12-week feeding period. Overnight fasted mice were bled into tubes containing 8 μL of 0.5 mol/L ethylenediaminetetraacetic acid. Plasma was prepared and stored at -80 ^{o}C before use.

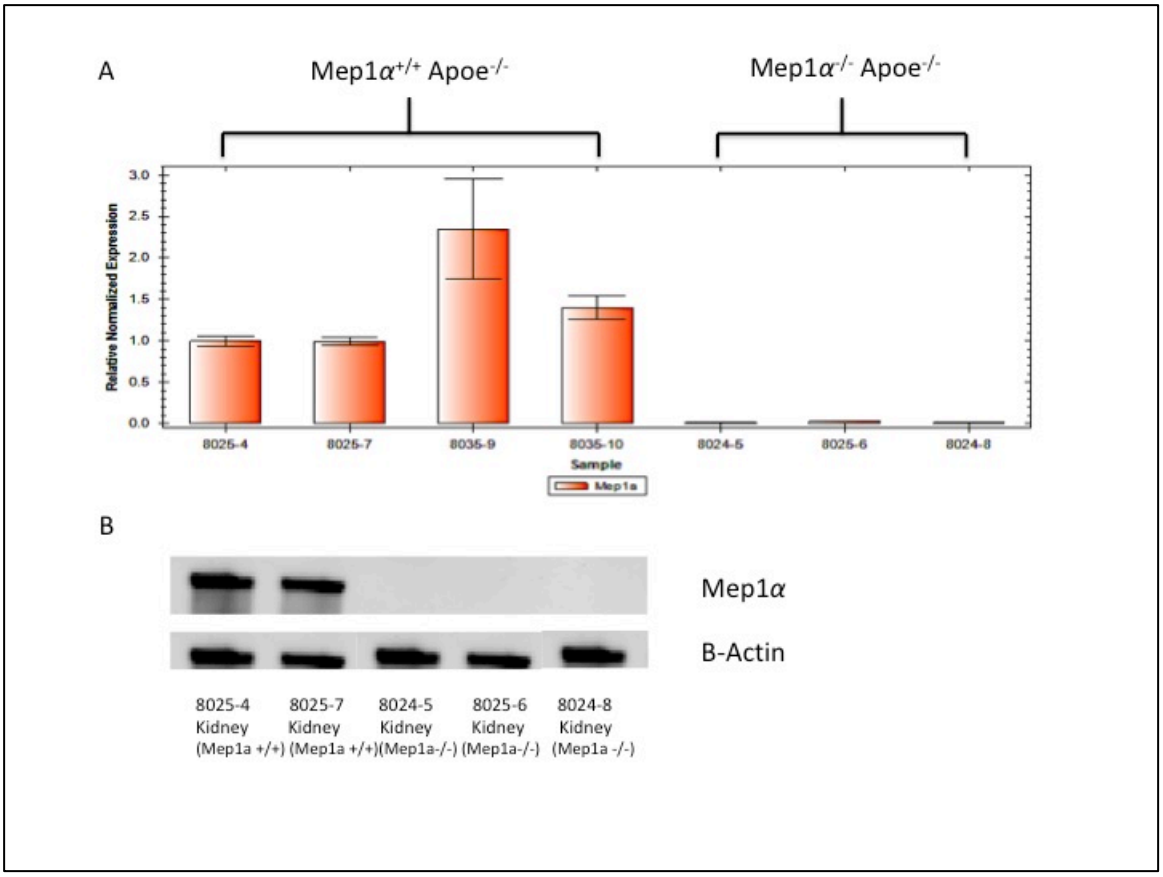

Figure 5.8 *Verification of Mep1α knockout in Apoe$^{-/-}$.Mep1α$^{-/-}$ double knockout mice. A) Quantitative PCR for Mep1α in the kidney of B6 (left) and Mep1α$^{-/-}$ (right) mice. Expression is normalized to β-actin. B) Western blot for Mep1α in the kidney of B6 (left) and Mep1α$^{-/-}$ (right) mice.*

Housing and husbandry. Breeding pairs were housed in a cage of 1 adult male and 2 females, and litters were weaned at 3 weeks of age onto a rodent chow diet in a cage of 5 or less. All mice were housed under a 12-h light/dark cycle at an ambient temperature of 23 °C and allowed free access to water and drinking food. Mice were fasted overnight before blood samples were collected.

Measurements of plasma glucose and lipid levels. Plasma glucose was measured with a Sigma glucose (HK) assay kit, as reported[159]. Briefly, 6 µl of plasma samples were incubated with 150 µl of assay reagent in a 96-well plate for 30 min at 30 °C. The absorbance at 340 nm was read on a Molecular Devices (Menlo Park, CA) plate reader. The measurements of total cholesterol, HDL cholesterol, and triglyceride were performed as reported previously[112]. Non-HDL cholesterol was calculated as the difference between total and HDL cholesterol.

Quantification of aortic atherosclerosis. Atherosclerotic lesion sizes in the aortic root of mice were measured as previously reported[88]. Briefly, the vasculature of mice was perfusion-fixed with 4% PFA (paraformaldehyde) through the left ventricle of the heart. The aortic root and adjacent heart were harvested, embedded in optimal cutting temperature compound and cross-sectioned in 10-µm thickness. Sections were stained with oil red O and hematoxylin and counterstained with fast green. Atherosclerotic lesion areas were measured using Zeiss AxioVision 4.8 software. Atherosclerotic lesion areas were measured using Zeiss AxioVision 4.8 software. For chow diet mice, the aortic lesion area from all sections was measured and the summation of all values measured was used as the value for statistical analysis. For mice fed Western diet, the top 5 largest aortic lesion areas were averaged for each mouse and this value was used for statistical analysis. *Mep1α*$^{+/+}$*Apoe*$^{-/-}$ (B6) and *Mep1α*$^{-/-}$*Apoe*$^{-/-}$ littermates were initially used for lesion comparisons. Because the lesion sizes of the B6 littermates were comparable to those we had previously quantified, non-littermate B6 mice were included for lesion comparison to increase the number of mice used and increase its accuracy.

Lesion Stability Assesment. Metrics pertaining to aortic lesion stability were measured

using aortic root sections from Western diet mice and were stained using either hematoxylin and eosin (H&E) or Masson's trichrome. The amount of necrosis and fibrous cap thickness were assessed using H&E. Quantification of necrotic area and overall lesion area was quantified by manual outlining using ImageJ. Quantification of unstained area was done using the Versatile Wand Tool plugin for ImageJ to select the area corresponding to the color of unstained tissue. The amount of collagen, bone, and muscle fibers were assessed using Masson's trichrome. Quantification of necrotic area and overall lesion area was quantified by manual outlining using ImageJ. Quantification of staining was done using the Versatile Wand Tool plugin for ImageJ to select the area corresponding to the color of the stain in Masson's trichrome.

Immunofluorescent Staining. Fluorescent immunostaining was performed on frozen aortic root sections from Western diet mice. OCT (Tissue Tek) was removed through incubation in acetone, then slides were blocked with 10% normal donkey serum in 10% Bovine Serum Albumin (BSA) in PBS. Slides were incubated with the following primary antibodies: rat anti-mouse CD107b (Mac3)(1:100 dilution; BD Pharmigen 550292), rat anti-mouse Ly6g (Gr-1) (1:200 dilution; clone RB6-8C5; eBioscience 14-5931-82), goat anti-mouse α-smooth muscle cell actin (SMC-A) (1:100 dilution; Ango Biosciences ARG63621), rabbit anti-mouse Mep1α (1:1000 dilution; provided by Christoph Becker-Pauly, University of Kiel), goat anti-mouse LIX (CXCL5) (undiluted; R&D Systems LIX ELISA Kit polyclonal LIX Conjugate P129149). Slides were incubated with the following fluorescent secondary antibodies: Goat anti-rat IgG (H+L) (1:250 dilution; Alexa Fluor 488; Cell Signaling 4416S), donkey anti-goat IgG (1:250 dilution; NL557; R&D Systems NL001), donkey anti-goat IgG (H+L) (1:250 dilution; Alexa Fluor 633/Cy3; Invitrogen A21082). After staining, slides were

mounted using Fluoromount-G with DAPI (Southern Biotech). Images were taken using a Zeiss LSM 800 confocal microscope.

Cell Component Analysis. Immunofluorescent images were analyzed using the Fiji package for Image J[197]. Cells positive for Mac3 or SMC-A were counted and normalized to the total number of DAPI positive cells in the aortic lesion to get the percent positive cells. Ly6g and CXCL5 stain was not confined to cells, so the area of the stain was calculated and normalized to the aortic lesion area to get the percentage positive area.

ELISA. ELISA assays for MCP-1 (R&D), CXCL5 (R&D), and MDA (TBARS assay, Cayman Chemical) were performed on blood plasma extracted from mice fed Western diet at time of sacrifice.

Statistical Analysis. For comparison of two groups of continuous variables with normal distribution and equal variances, two-tailed unpaired Student's *t*-tests were performed with a confidence level of 95%. All data are presented as the mean ± SEM.

Chapter 6

Deep learning-based quantification of abdominal fat on magnetic resonance images

Andrew T. Grainger, Nicholas J. Tustison, Kun Qing, Rene Roy, Stuart S. Berr, and Weibin Shi

Published in: A.T. Grainger, N.J. Tustison, K. Qing, R. Roy, S.S. Berr, W. Shi, Deep learning-based quantification of abdominal fat on magnetic resonance images. PLoS One 13 (2018), e0204071.

6.1 Abstract

Obesity is increasingly prevalent and associated with increased risk of developing type 2 diabetes, cardiovascular diseases, and cancer. Magnetic resonance imaging (MRI) is an accurate method for determination of body fat volume and distribution. However, quantifying body fat from numerous MRI slices is tedious and time-consuming. Here we developed a deep learning-based method for measuring visceral and subcutaneous fat in the abdominal region of mice. Congenic mice only differ from C57BL/6 (B6) *Apoe* knockout (*Apoe*$^{-/-}$) mice in chromosome 9 that is replaced by C3H/HeJ genome. Male congenic mice had lighter body weight than B6-*Apoe*$^{-/-}$ mice after being fed 14 weeks of Western diet. Axial and coronal T1-weighted sequencing at 1-mm-thickness and 1-mm-gap was acquired with a 7T Bruker ClinScan scanner. A deep learning approach was developed for segmenting visceral and subcutaneous fat based on the U-net architecture made publicly available through the open-source ANTsRNet library a growing repository of well-known neural networks. The volumes of subcutaneous and visceral fat measured through our approach were highly comparable with those from manual measurements. The Dice score, root-mean-square error (RMSE), and correlation analysis demonstrated the similarity between two methods in quantifying visceral and subcutaneous fat. Analysis with the automated method showed significant reductions in volumes of visceral and subcutaneous fat but not non-fat tissues in congenic mice compared to B6 mice. These results demonstrate the accuracy of deep learning in quantification of abdominal fat and its significance in determining body weight.

6.2 Introduction

Obesity is excessive fat accumulation in the body to the extent that an individual has increased risk for an array of chronic diseases, including type 2 diabetes (T2D), cardiovascular diseases, and cancer [198]. It is a growing epidemic in the US and globally. In 2014, 37.7% of adults and 17% of youth aged 2–19 years were obese in the US[199], and 650 million adults aged 18 years or older were obese worldwide (http://www.who.int/mediacentre/factsheets/fs311/en).

The diagnosis of obesity has primarily relied on a few anthropometric measurements, such as body mass index (BMI), waist circumference, or waist-to-hip ratio. A BMI of ≥30 kg/m^2 is considered obese, and a BMI of 25 to <30 kg/m^2 is defined as overweight. However, these indirect measurements neither allows for distinguishing fat from skeletal muscle nor distinguishing between visceral and subcutaneous fat. Excessive body fat rather than skeletal muscle is related to both increased plasma levels of free fatty acids and pro-inflammatory cytokines as well as endoplasmic reticulum stress, all of which contribute to development of insulin resistance, type 2 diabetes, and atherosclerosis[4]. Central or abdominal obesity has been shown to be more closely associated with risk of coronary artery disease and type 2 diabetes[5,6].

Thus, there is a medical demand for accurately measuring the amount and distribution of body fat to better understand its impact on health and disease. Such imaging modalities as magnetic resonance imaging (MRI) and computed tomography (CT) can clearly distinguish fat from other tissues and thus allow for accurate measurement of fat and non-fat tissue amounts[7]. Compared to CT, MRI involves no ionizing radiation and is more efficient in differentiating fat from non-fat tissues[200,201]. Quantification of body fat volume using MRI or CT involves analysis of many cross-sectional or longitudinal slices across the region of interest. Thus, manual measurement of fat volume with MRI or CT

images is a tedious and time-consuming task. To save time and also reduce subjective influences from observers, several semi-automated algorithms have been developed for quantifying body fat[8–11]. However, nearly all of the algorithms are still dependent on expert knowledge for tuning the features of images and their accuracy and reliability are often low.

Due to recent successes, deep learning with convolutional neural nets has gained popularity in the literature for tackling problems in the areas of image recognition, classification and segmentation[202]. Development of deep learning algorithms relies on neural networks, a computational architecture that can be trained with a large number of annotated images or data, identify features from them, and make predictions of other data. ANTsRNet is a collection of deep learning architecture of neural networks ported to the R language and built on the Keras neural network[203]. Here, we applied these architectures and other open-source software packages developed by the Advanced Normalization Tools team (which includes one of the co-authors) to provide a comprehensive protocol for automatically segmenting abdominal visceral and subcutaneous fat of mice on MR images. The trained networks provided publicly can be directly applied to these MR data for conducting objective and expedited measurements of fat volume in mice.

6.3 Results

6.3.1 Phenotypic difference in body weight

Chromosome 9 congenic mice were genetically identical to B6-*Apoe*$^{-/-}$ mice except for the chromosome 9 region (15.6–115.6 Mb), which was replaced with the C3H genome[95]. After being fed 14 weeks of Western diet, male congenic mice displayed significantly lighter body weight than B6-*Apoe*$^{-/-}$ mice (30.67 ± 1.05 vs. 38.81 ± 2.66 g; $P = 0.0021$; n = 7 and 17, respectively) (**Figure 6.1**). This result confirmed the existence of *W10q13*, a locus for body

weight on mouse chromosome 9 we previously mapped in an intercross derived from B6-*Apoe*$^{-/-}$ and BALB/cJ-*Apoe*$^{-/-}$ mice[76].

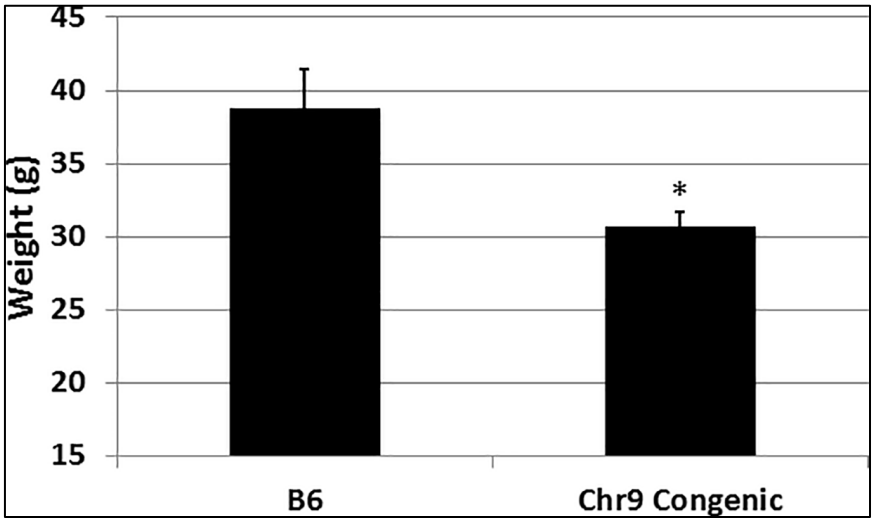

Figure 6.1 *Body weight (g) of male congenic and B6-Apoe*$^{-/-}$ *mice fed a Western diet. Results are means ± SE of 7 B6 and 17 congenic mice after 14 weeks of Western diet. * P < 0.05 vs. B6 mice.*

6.3.2 Manual Measurement of body fat

MRI scans were performed within one week before mice were euthanized. Noticeable differences in subcutaneous and visceral fat amounts between congenic and B6-*Apoe*$^{-/-}$ mice could be seen on MR images and gross anatomic examination (**Figure 6.2**). To quantify the differences, the volumes of subcutaneous and visceral fat were measured on 4 representative slices (1 slice at the level of liver, 2 at kidney, 1 at pelvic) spanning the abdominal region for congenic and B6-*Apoe*$^{-/-}$ mice (4 mice per group). Congenic mice had a

subcutaneous fat volume of 284.0 ± 72.4 mm^3/slice, significantly smaller than the volume of 627.5 ± 29.9 mm^3/slice in

B6-*Apoe*$^{-/-}$ mice (p = 0.0046) (**Figure 6.3**). The visceral fat volume was also significantly smaller in congenic mice (554.6 ± 70.3 mm^3/slice) than in B6 mice (1045.6 ± 136.4 mm^3/slice) (p = 0.019). In contrast, non-fat tissue volume was comparable between congenic and B6-*Apoe*$^{-/-}$ mice (2603.8 ± 66.6 vs. 2597.7 ± 189.5 mm^3/slice).

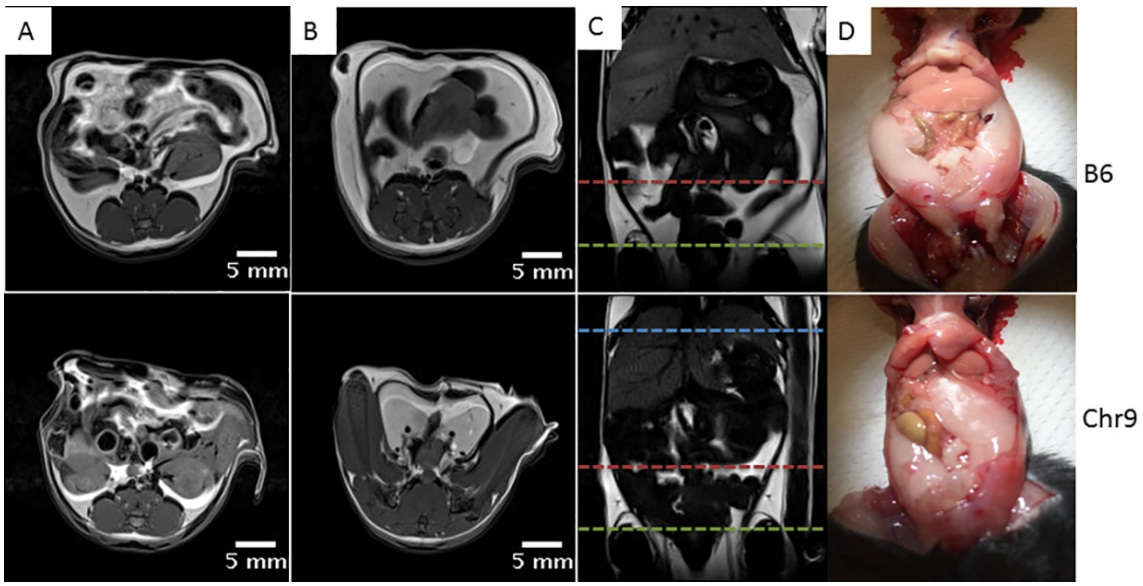

Figure 6.2 *Representative MR images of congenic and B6-Apoe*$^{-/-}$ *mice fed a Western diet. Axial MR slices at the levels denoted by the red (A) and green lines (B) across the coronal slice (C). D) Gross examination of abdominal fat. Top row: B6-Apoe*$^{-/-}$ *mice; bottom row: congenic mice.*

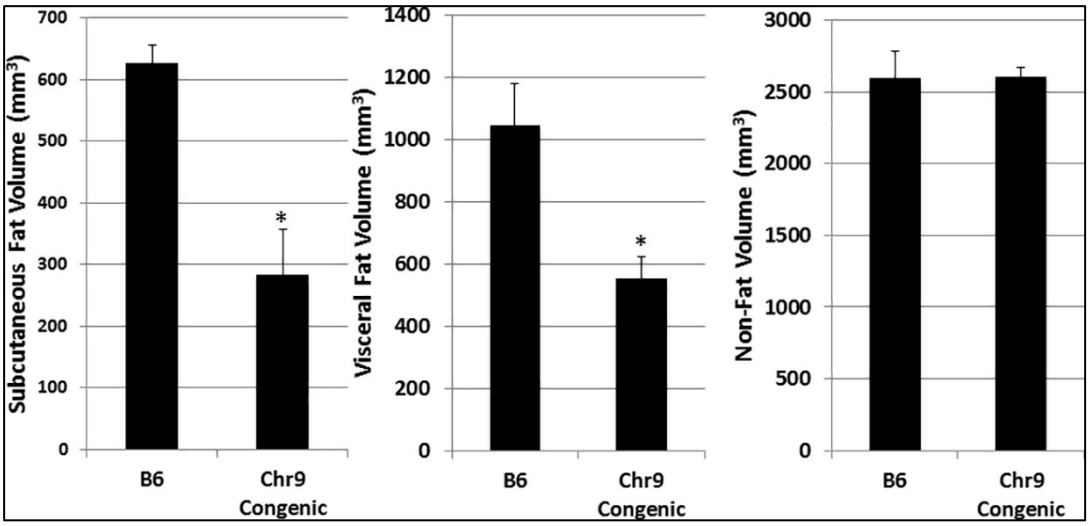

Figure 6.3 *Fat and non-fat volumes of congenic and B6-Apoe$^{-/-}$ mice measured manually using axial MR slices. Results are means ± SE of 4 mice per group. * P = 0.012*

6.3.3 Automatic measurement of visceral and subcutaneous fat

A modified deep learning approach with a novel template-based data augmentation strategy was employed. The U-net model was trained with axial MR images spanning the entire abdominal region of mice. We then applied the learned U-net model to analysis of 60 water-filtered evaluation MR images from three B6 mice that were previously quantified manually to generate corresponding probability image sets designating subcutaneous and visceral regions. Automatic segmentations of visceral and subcutaneous fat for either water-filtered or unfiltered slices generated from corresponding probability images were highly consistent with the manually generated segmentations of the same input images (**Figures 6.4 and 6.5**). These segmentation images were then used to generate the regions of interest on the original MR slices for quantifying both visceral and subcutaneous fat volumes.

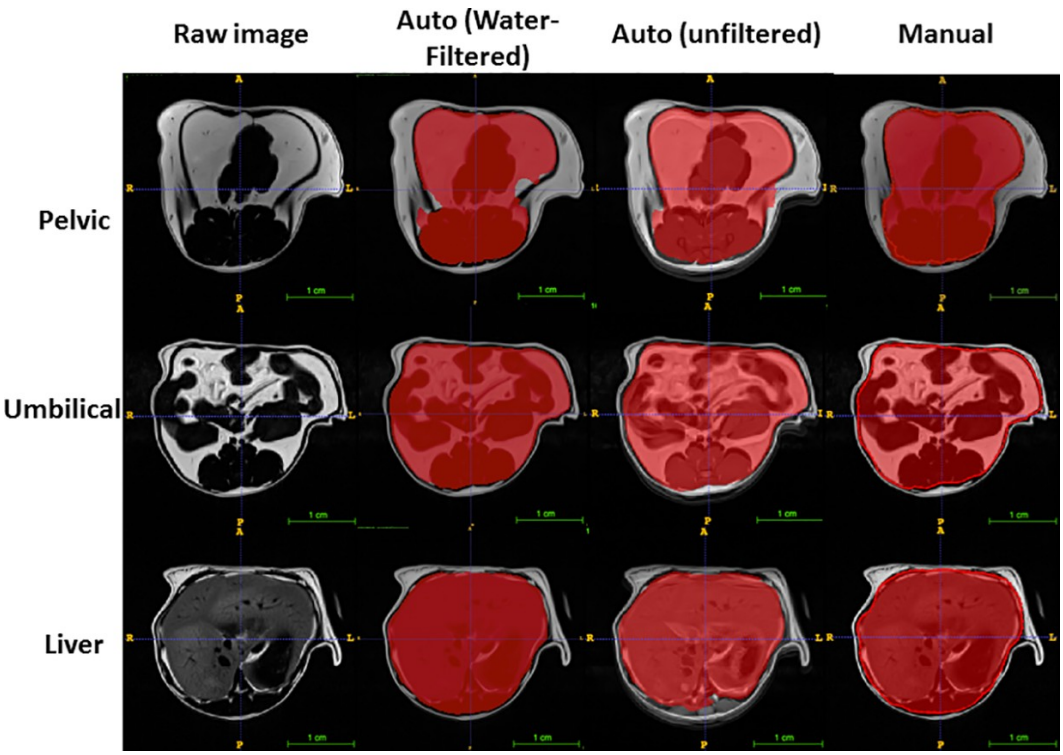

Figure 6.4 *The accuracy of deep learning in deriving the area containing visceral fat at three different levels: pelvic, kidney and liver on MR images. Prediction of the segmentation in the red area by deep learning is highly consistent with the input data. The red area is where visceral fat is included. Auto: prediction made by deep learning; manual: the red line is drawn with ImageJ and the area within the red line is used as input for visceral fat.*

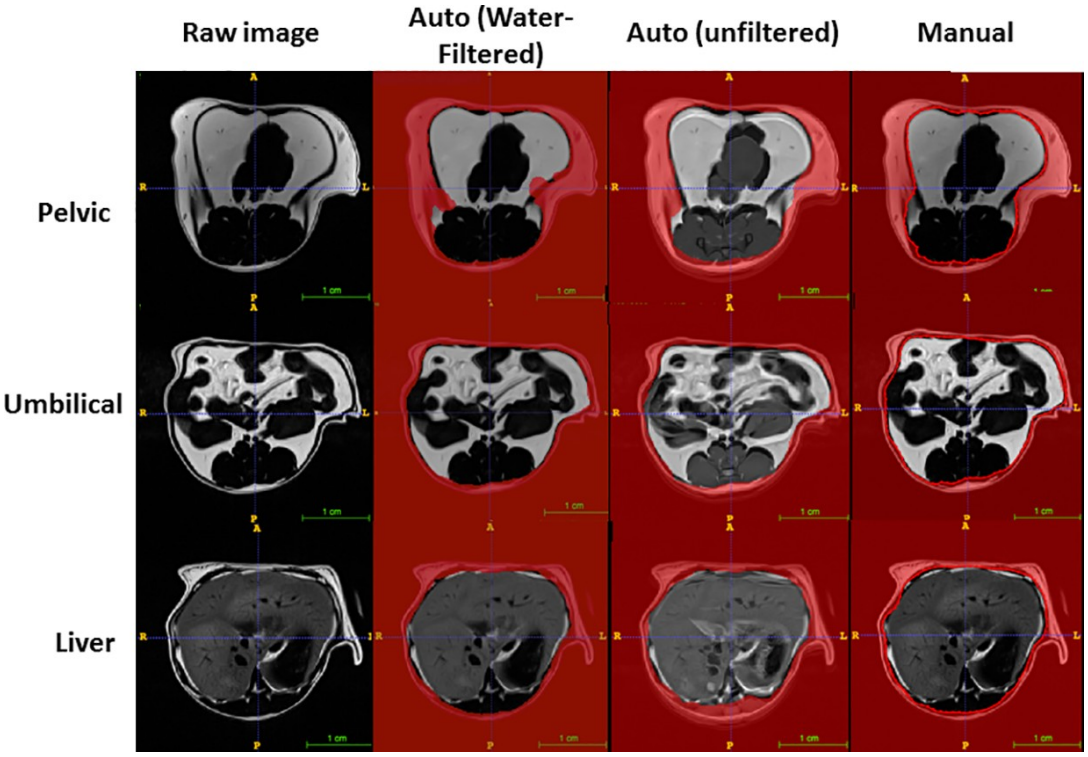

Figure 6.5 *The accuracy of deep learning in deriving the area containing subcutaneous fat at three different levels: Kidney and liver on MR images. The red area is where subcutaneous fat is located. Auto: prediction made by deep learning; manual: the red line is drawn with ImageJ and the area outside the red line is used as input for subcutaneous fat.*

Comparative analyses were performed to verify the accuracy of the automated method through multiple metrics, including Dice's similarity coefficient, RMSE, correlation coefficient, and real measurement results. We analyzed 43 slices that had not been included in the original training set and calculated Dice coefficient values measuring the level of overlap between the manual and automatic segmentation results. The average Dice coefficient value was 0.968 ± 0.00267 (Min = 0.919, Max = 0.987) (**S6.1**).

The volumes of visceral fat measured from sequential water-filtered slices were comparable between the two measurements (**Figure 6.6A**). The total visceral fat volume measured by the automated method was also comparable to the volume achieved by the manual method (5463 ± 625 vs. 5331 ± 642 mm^3; p = 0.890) (**Figure 6.6B**). Correlation analysis showed an extremely high agreement in measurements made by the two methods (R^2 = 0.99; p = 9.8E-64) (**Figure 6.6C**). The RMSE values supported this close correlation (**S6.1**). The volumes of subcutaneous fat measured on sequential water-filtered slices were also comparable between the two measurements (**Figure 6.7A**). The total subcutaneous fat volume from auto-segmentation was also similar to the volume obtained from the manual method (3571 ± 141 vs. 3617 ± 122 mm^3; *P* = 0.533) (**Figure 6.7B**). Correlation analysis shows a high agreement in measurement results achieved from the two methods (R^2 = 0.96) (**Figure 6.7C**).

The automated method also applied to measurement of fat volumes on unfiltered MR slices, though they had not been used to train the algorism. The Dice coefficient for the corresponding unfiltered images still shows a high degree of overlap with manual segmentations with a value of 0.962 ± .00467 (Min = 0.772, Max = 0.990), however the overlap is slightly worse than with water-filtered images. The total visceral fat volume was slightly higher than the volume measured manually on water-filtered slices (6232 ± 834 vs. 5331 ± 642 mm^3; p = 0.565) (**Figure 6.6**). Correlation analysis showed a high degree of correlation between the results achieved from the two measurements (R^2 = 0.95; p = 8.5E-3). The total subcutaneous fat volume was 3606 ± 117 mm^3, comparable to the volume of 3571 ± 141 mm^3 measured manually on water filtered slices (*P* = 0.816). A significant correlation between the manual measurement made on water-filtered images and the automated measurement made on unfiltered images was observed (R^2 = 0.8278; p = 4.9E-

24). A small mismatch in the measurement results was observed in slices acquired at a higher location where the liver was located, though the difference was not significant.

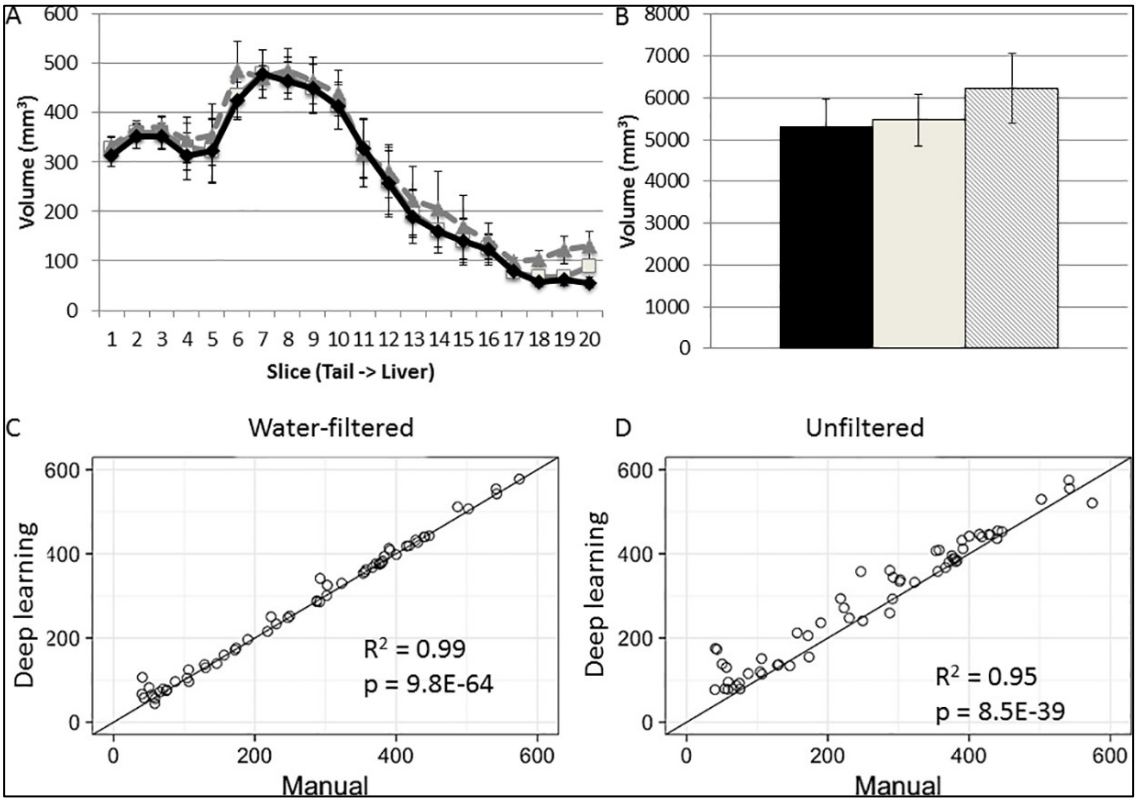

Figure 6.6 *Comparison between the automated and manual measurements in quantification of visceral fat. A) The volumes of visceral fat on 20 sequential axial slices from the tail root (slice 1) to the diaphragm (slice 20) of B6-Apoe$^{-/-}$ mice measured by manual (black) and deep learning (water filtered = solid; unfiltered = dashed). B) The total fat volume of B6-Apoe$^{-/-}$ mice measured by manual (black) and deep learning (grey; water filtered = solid, unfiltered = dashed). C) Correlation analysis of visceral fat volumes on water-filtered MR slices measured with the two methods. Each dot represents on axial MR slice. R^2 and P values are shown in the figure. D)*

Correlation analysis of visceral fat volumes on unfiltered MR slices measured with the two methods.

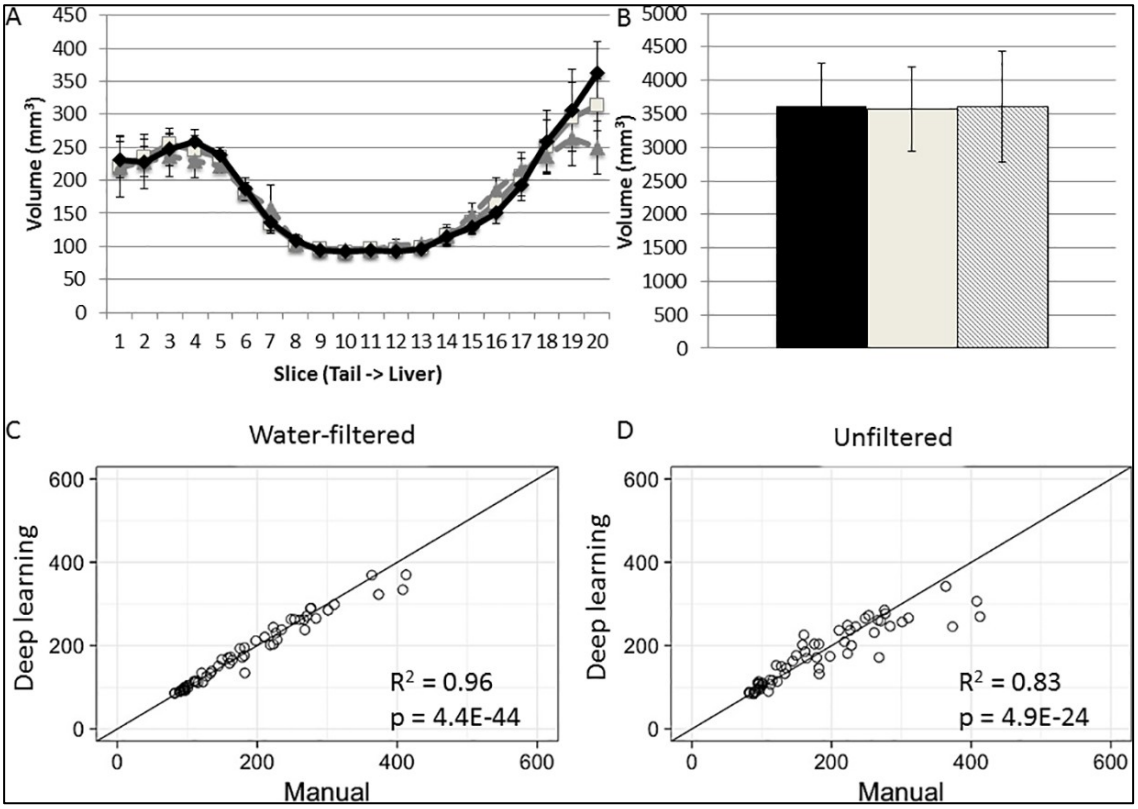

Figure 6.7 *Comparison between the automated and manual measurements in quantification of subcutaneous fat. A) The volumes of subcutaneous fat on 20 sequential axial slices from the tail root (slice 1) to the diaphragm (slice 20) of B6-Apoe$^{-/-}$ mice measured by a manual (black) or automated method (water filtered = solid; unfiltered = dashed). B) The total subcutaneous fat volume of B6-Apoe$^{-/-}$ mice measured by a manual (black) or automatic method (water filtered = solid; unfiltered = dashed). C) Correlation analysis of subcutaneous fat volumes on water-filtered MR slices measured with the two methods. Each dot represents a slice. R^2 and P values are*

shown in the figure. D) Correlation analysis of subcutaneous fat volumes on unfiltered MR slices measured with the two methods.

6.3.4 Differences between congenic and B6-*Apoe*$^{-/-}$ mice in subcutaneous and visceral fat

The automated method was used to measure fat volumes on axial water-filtered slices acquired from congenic and B6-*Apoe*$^{-/-}$ mice (n = 4 per group). Congenic mice had less subcutaneous and visceral fat than B6- *Apoe*$^{-/-}$ mice in most slices scanned (**Figure 6.8**). The differences were more significant in the lower half of the abdominal region for visceral fat. The total visceral fat volume of congenic mice was reduced by 48.8% relative to the volume of B6 mice (2652 ± 465 vs. 5183 ± 523 mm^3; p = 0.011). The subcutaneous fat volume of congenic mice was reduced by 37.1% (2184 ± 363 vs. 3471 ± 141 mm^3; p = 0.016).

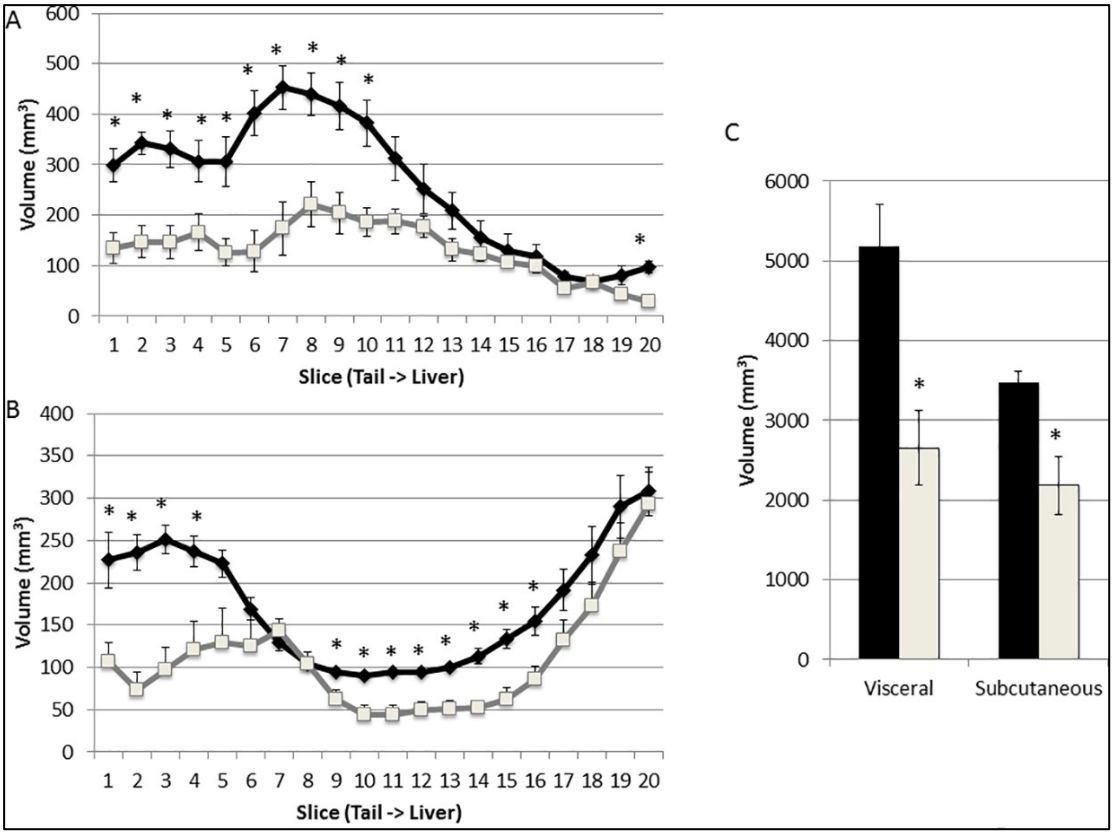

Figure 6.8 *Comparison between congenic and B6-Apoe$^{-/-}$ mice in visceral (A) and subcutaneous fat volumes (B) measured by the automated method on water filtered MR slices. MR slices span from the pelvic cavity (slice 1) to the top of the liver (slice 20). Results are means ± SE of 4 mice per group on each slice. C) The total volume of visceral and subcutaneous fat in the abdominal region of congenic and B6-Apoe$^{-/-}$ mice. * P < 0.05.*

6.4 Discussion

An increasing prevalence of central obesity and its close association with cardiovascular disease and type 2 diabetes demands a reliable technique for abdominal fat quantification. MRI is the most effective imaging modality currently used for detecting the fat tissue in vivo, but quantifying fat volume from numerous slices remains a challenge. In

this study, we developed a deep learning-based method for measurement of abdominal fat using MR images obtained from two mouse strains that differed markedly in body weight on a high fat diet. Volumes of visceral and subcutaneous fat and non-fat tissues measured by the algorithm have shown a high degree of consistency with those achieved by a manual method. Moreover, we have found that body fat rather than non-fat tissues accounts for the difference between the two strains in body weight.

In this study, we trained the modified U-net network with 37 original MR images spanning the abdominal region of mice. Despite the limited number, the algorithm has shown a better performance than the prior best automatic method for quantification of visceral and subcutaneous fat using MR images[11]. We calculated the Dice coefficient value to determine the overlap between the new algorithm and the manual method in the measurement of visceral fat with the same set of MRI slices. Our results have shown a high similarity between the two methods in quantifying visceral fat, by achieving an average Dice value of 0.968. Pearson's correlation analysis was performed to evaluate the consistency of subcutaneous and visceral fat volumes measured by the two methods. The correlation coefficient of the values obtained from the two methods was 0.99 for visceral fat quantification and 0.96 for subcutaneous fat quantification made on water-filtered MR images. Correlation may not be a perfect measure of similarity because a predictor that consistently predicts half of the actual weight, would yield a correlation coefficient similar to the one when the predictor exactly predicts the actual weight. In contrast, the root mean square error (RMSE) does not suffer from this caveat. The small differences between values from the two methods demonstrate the effectiveness of deep learning in quantifying abdominal fat (S6.1). To evaluate performances of the algorithm at different segmentation locations, we compared the volume values of visceral and subcutaneous fat on each of the sequential MRI slices as well as the total fat volumes obtained from the two methods. The

automatically obtained fat volumes have shown a high consistency with those obtained manually.

We also applied the algorithm to the quantification of visceral and subcutaneous fat volumes on unfiltered MR images acquired from mice. Though the network had not been trained with unfiltered MR images, the results achieved still showed a high level of consistency with those achieved manually for both visceral and subcutaneous fat. Automated segmented images had an average Dice coefficient of 0.962 compared to manually-generated counterparts. The correlation coefficient of the fat volume results on individual slices was 0.95 for visceral fat and
0.83 for subcutaneous fat. However, we did find that the total visceral fat volume had a slightly larger value than the manual measurement and the correlation coefficients of the fat volume results with respect to the comparison between the two measurement methods were smaller relative to those achieved from water filtered MR images. A possible explanation for the gap in fat volume quantification is that the algorithm had not been trained with unfiltered MR images, which have a less clear contrast between the fat and non-fat tissue compared to water filtered MR images. It's also noteworthy that the automated and manual measurement results were not made from the same MR images but the images from two separate scans at comparable locations.

We observed that adipose tissue rather than non-fat tissues underlies the difference between congenic and B6-*Apoe*$^{-/-}$ mice in body weight on the Western diet. Indeed, congenic mice had a ~50% reduction in visceral fat volume and ~40% reduction in subcutaneous fat compared to the B6 mice but there was no reduction in non-fat tissue volume. The congenic mice were constructed to verify a locus initially mapped in an intercross between B6 and C3H *Apoe*$^{-/-}$ mice and then replicated in at least four independent crosses[75,79,79,83]. The congenic strain was highly resistant to atherosclerosis[95]. As obesity is a major risk for

atherosclerosis, the reduction in abdominal fat could contribute to the resistance of congenic mice to atherosclerosis. Because congenic mice had reductions in both visceral and subcutaneous fat, this result does not allow for judging which type of fat might contribute to their resistance to atherosclerosis.

In addition, we have found that the pattern of fat distribution differed in the two strains: visceral fat mostly accumulated in the lower half of the abdomen in B6-*Apoe*$^{-/-}$ mice whereas the visceral fat of congenic mice more evenly distributed in longitudinal direction. B6-*Apoe*$^{-/-}$ mice showed much more subcutaneous fat accumulation in the distal end of the abdomen compared to congenic mice.

In summary, the use of deep learning can accurately quantify visceral and subcutaneous fat of mice on MR images. The next logical step is to determine whether this new approach can be applied to humans. Nevertheless, because of the anatomical and biochemical similarities between the mouse and the human, we are confident that this algorithm should be applicable to quantification of abdominal fat in humans.

6.5 Methods

Ethics statement

All procedures were performed in accordance with the current National Institutes of Health guidelines (https://grants.nih.gov/grants/olaw/Guide-for-the-Care-and-use-of-laboratory- animals.pdf) and approved by the Institutional Animal Care and Use Committee of the University of Virginia (Assurance #A3245-01, Animal Protocol #3109).

Animals

Apoe knockout (*Apoe*$^{-/-}$) mice with a C57BL/6 (B6) genetic background were purchased from the Jackson Laboratory. Congenic mice, which only differ from B6- *Apoe*$^{-/-}$ mice in chromosome 9 region from 15.6 to 115.6 Mb, were constructed using the classical congenic breeding method as previously described[95]. The donor chromosome 9 region was derived from C3H/HeJ (C3H) mice. Mice were fed 14 weeks of Western diet (TD88137, Envigo), stating at 6 weeks of age.

Magnetic resonance imaging (MRI)

MRI for the abdominal region, which extended from the top of the diaphragm to the bottom of the pelvic cavity, was performed on a 7T Clinscan system (Bruker, Ettlingen, Germany). Mice were anesthetized under isoflurane inhalation during imaging and respiratory motion monitored with an MR-compatible gating system for mice (SA Instruments, Inc., Stony Brook, NY). Axial and coronal T1-weighted sequencing (20 axial and 18 coronal slices; axial voxel size = 0.156×0.156×1.000 mm; coronal voxel size = 0.430×0.430×1.000 mm) at 1-mm- thickness and 1-mm-gap was acquired. Axial 2D images were acquired once with and once without water filtration, and coronal imaging was performed only under the water filtration mode.

Manual measurement

The volumes of visceral and subcutaneous fat as well as non-fat tissues were quantified on water-filtered axial slices using the auto-thresholding function of the Fiji package for ImageJ[197]. Briefly, MR images (all 35 x 35 mm; Pixel width/height = 0.182291; Voxel depth = 2.0) were converted to binary ones with fat being white and non-fat black, and the "Analyze Particles" function was then used to determine the total white area (mm^2) on the binary images. The volume of fat (mm^3) was obtained by multiplying each slice's

total area (mm^2) by 2 (1-mm slice thickness +1-mm interslice gap), and the total fat volume was the sum of the fat volumes measured from all consecutive slices. The manual measurement results were used as the gold standard for comparisons with the automated measurement results.

Automatic measurement

The automated method for body fat quantification consists of steps, including training data preparation, template-based data augmentation, and fat quantitation. The complete flowchart of the process is shown in **Figure 6.9**.

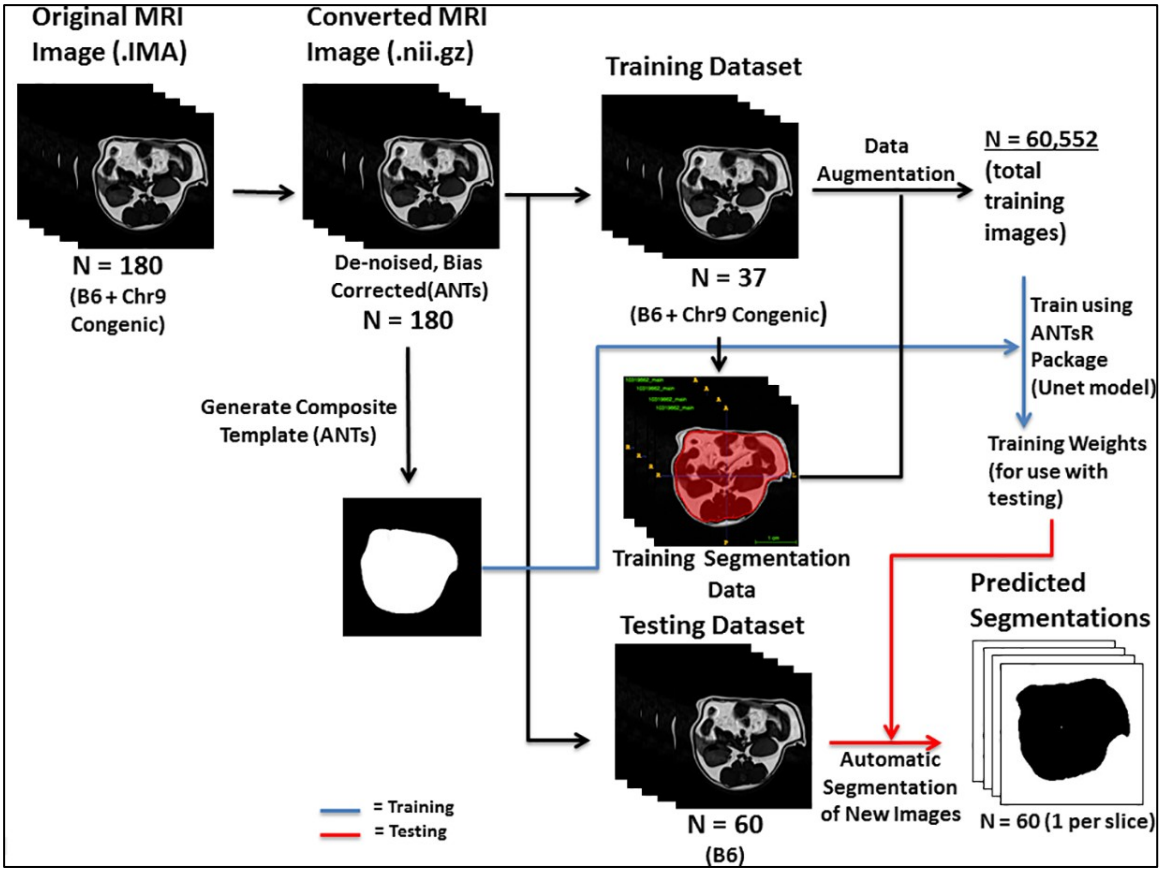

Figure 6.9 *Flowchart illustrating the major steps employed in segmentation and quantification of subcutaneous and visceral fat in MR images. MR images were first converted to a format compatible with the ANTs software packages (NifTI; .nii.gz), and then processed for denoising and bias correction. 37 representative original MR images and their corresponding segmentation images from B6 mice and congenic mice were used for establishing the core training dataset. A novel data augmentation strategy was used to increase the number of training images from 37 to 60,552, which were used to train a U-net-based architecture. A testing dataset consisting of 60 MR images from 3 B6 mice was run through the model to test for accuracy of the method.*

Training data preparation

Original MR images were converted to the Nifti format and underwent denoising[204] and bias correction[205] as preprocessing steps to improve image quality. The fat tissue compartment was segmented into subcutaneous and visceral fat based on the peritoneal wall on each image. Visceral fat is the fat within the contour of the abdominal wall, and subcutaneous fat is the fat within the abdominal wall covering the abdominal muscles. Manual labeling was performed using the open source segmentation tool ITK-SNAP[206]. 37 water-filtered 2D MR images from 5 B6 and 3 congenic mice were each manually segmented with fully annotated area for visceral fat and area for subcutaneous fat and these images formed the core training data set. Of them, 26 images were obtained from B6 mice and 11 from congenic mice. These 2D images had a dimension of 35 x 35 x 1 mm with a 1mm gap between consecutive images. They were selected as a diverse and representative set of images spanning the entire abdominal region and varying in intensity and fat amount.

A composite binary mask template was then generated from the segmentation maps of the core training set using the ANTs toolkit (https://github.com/ANTsX/ANTs)[207]. The use of the manually edited binary masks over the original MR images is due to the lack of correspondence of internal structures within the template cohort and our interest in capturing the global shape variation between the visceral and subcutaneous fat regions for the purposes of data augmentation (explained in the next sub-section).

Template-based data augmentation

The need for large training data sets is a well-known limitation associated with deep learning algorithms[208]. Whereas the architectures developed for such tasks as the ImageNet competition have access to millions of annotated images, such data access is not always available and such is typically the case in medical imaging. In order to achieve data set sizes

necessary for learning functional models, various data augmentation strategies have been employed, including application of intensity transformations, such as brightening and enhanced contrast, and simple spatial transformations, such as arbitrary rotations and translations. Regarding the latter, such transformations are not ideal as they might not reflect what is typically seen in medical images and might not sufficiently sample the shape-space of the population currently being studied for generalizability.

In this work, we used a template-based approach whereby image data sampled from the population was used to construct a representative template that was optimal in terms of both shape and intensity[207]. In addition to the representative template, this template-building process yields the transformations to/from each individual image to the template space (**Figure 6.10**). This permits a propagation of the training data to the space of each individual image. In the simplest case, the training data was used to construct the template and then each individual training data was propagated to the space of every other individual training data. In this way, a training data set of size N can be expanded to a data set of size N^2. A more complicated use case could build a template from M data sets (where one would expect $M > N$). Transformations between the training data and the template could then be used to propagate the training data to the spaces of the individual members of the template-generating data for an augmented data set size of $M \times N$.

We used 37 MR images thus permitting 372 = 1,369 possible deformable shapes which comprise the first level of augmented data that can be further augmented by more conventional strategies (e.g., brightness transformations, translations, etc.). The second level of augmented data created two horizontal flip images per level one image as well as 36 rotation images per flip image. In this way we augmented the 1,369 first-level images to 98,568 second-level images (1,369 * 2 * 36).

We implemented tailored batch generators to implement this data augmentation strategy and to generate the data batches per epoch on the fly. Although slower, this avoids the problem of loading all training data in memory during learning. Specific parameters from training can be found in the scripts we wrote for the project available in our bitbucket.org repository [https://bitbucket.org/atg3qz/unet_fat_mri/src].

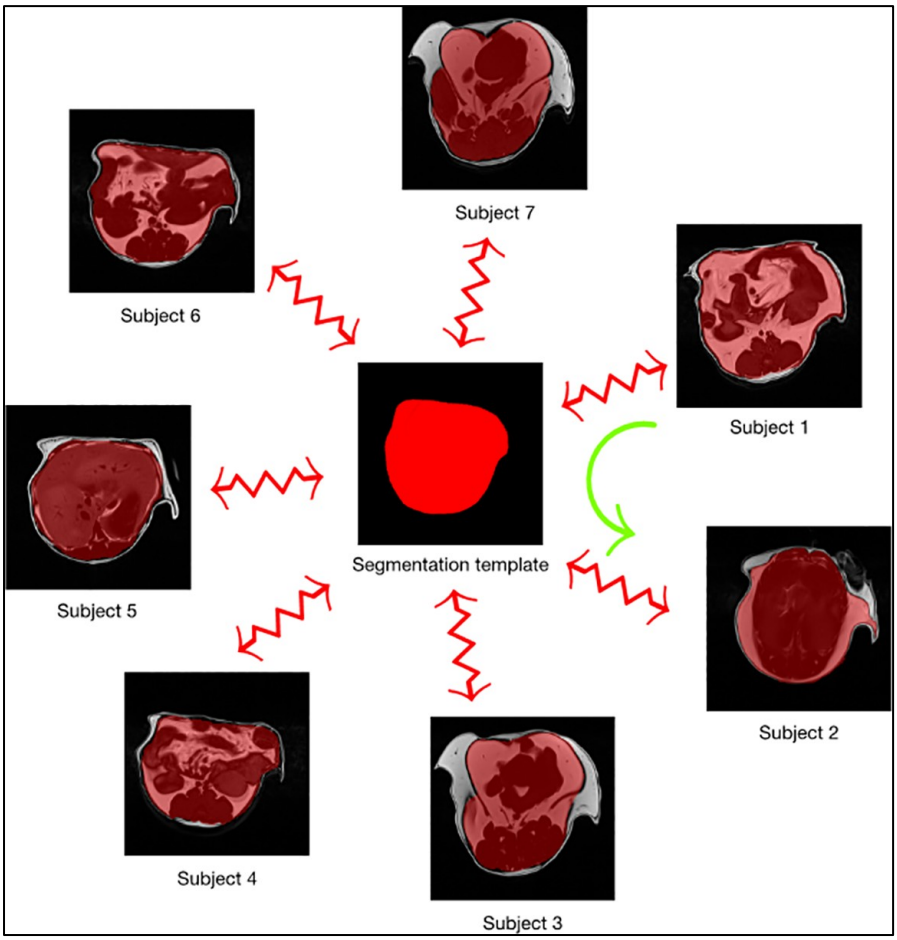

Figure 6.10 *Example MR images showing data augmentation strategy. We introduced a novel data augmentation strategy for training through ANTs-based template*

construction. A template was created from the training data segmentation images where the red area includes visceral fat and the foreground designates regions of subcutaneous fat. 37 images were used to create such a template that permitting 372 = 1369 possible deformable shapes which are further augmented by random horizontal flipping and randomized rotation.

U-net for segmentation of abdominal fat in MRI

U-net is a well-known convolutional neural network architecture for voxelwise classification labeling [https://arxiv.org/abs/1505.04597]. It has been employed in various segmentation applications such as MRI of the knee[209], brain tumors in PET imaging[210], and histology images[211]. We combined U-net with our template-based data augmentation scheme described in the previous section for segmenting abdominal fat and subsequent quantification.

ANTsRNet, an open-source repository for deep learning architectures

We also introduced ANTsRNet [https://github.com/ntustison/ANTsRNet], a collection of well-known deep learning architectures ported to the R language. ANTsRNet is built using the Keras neural network library[203] (available through R) and is highly integrated with the ANTsR package, the R interface of the ANTs toolkit. Consistent with other ANTs-based software offerings, ongoing development is currently carried out on GitHub using a well-commented coding style, thorough documentation, and self-contained working examples. One such architecture is the well-known U-net architecture

[https://arxiv.org/abs/1505.04597] where we have replaced the cross-entropy loss function with a multi-label Dice function based on previous work [http://hdl.handle.net/10380/3141].

Input and testing nifti images are all of size 192 × 192, as is the template, such that no resampling is required. I/O from the disk to the ANTsRNet software is handled by antsReadImage/antsWriteImage functions available as part of ANTsR. During data augmentation, the antsApplyTransforms program was used to transform a randomly chosen image/segmentation mask pair to a randomly chosen target in the training cohort. During data augmentation, a digital "coin toss" was used to randomly flip the image/segmentation pair in a left-right direction followed by a randomly chosen rotation angle between 0 and 359 degrees. These latter operations were handled by the magick package in R. Specific parameters for the 2-D U-net architecture for both models are as follows:

- Adam optimization:
  - learning rate = 0.0001.
  - optimization function: Dice coefficient.
- Number of epochs: 40.
- Convolution layers:
  - -kernel size: 3×3.
  - -activation: rectified linear units (ReLU).
  - -number of filters: doubled at every layer starting with N = 32.
- Max pooling layers:
  - size: 2×2.
  - stride length: 2×2.
- Upsampling/transposed convolution (i.e., deconvolution) layers:

- kernel size: 2×2.
- stride length: 2×2.
- activation: rectified linear units (ReLU).

Training took approximately 1.5 hours. After model construction, prediction per image (after preprocessing) takes < 1 second per image. Both model construction and prediction utilized a NVIDIA Titan Xp GPU. Both training and prediction are handled by custom-built R scripts available as part of the Bitbucket repository associated with this work.

Fat quantification with segmented images

The novel segmentation images were evaluated for accuracy for quantitation of visceral and subcutaneous fat using the Fiji package. To measure visceral fat, the segmentation image was used as a template to generate a contour, which was then overlapped onto the original MR image to outline the region containing visceral fat (**Figure 6.11**). The outlined area was made binary with visceral fat to the black color and other components to white. The "Create Selection" tool was then used to generate a contour perfectly outlining the visceral fat component. The 'Restore Selection" tool applied to place this selection onto the original MR image, giving an exact outline of the visceral component in the image. This selected area was copied and pasted to a new image (all images 35 x 35 mm; Pixel width/height = 0.182291; Voxel depth = 2.0), and the area of the image was measured in the same manner as described above under "Manual measurement". To quantify subcutaneous fat, the "Make Inverse" tool was used to select the inverse of the visceral component and the selected area was again pasted to a new image.

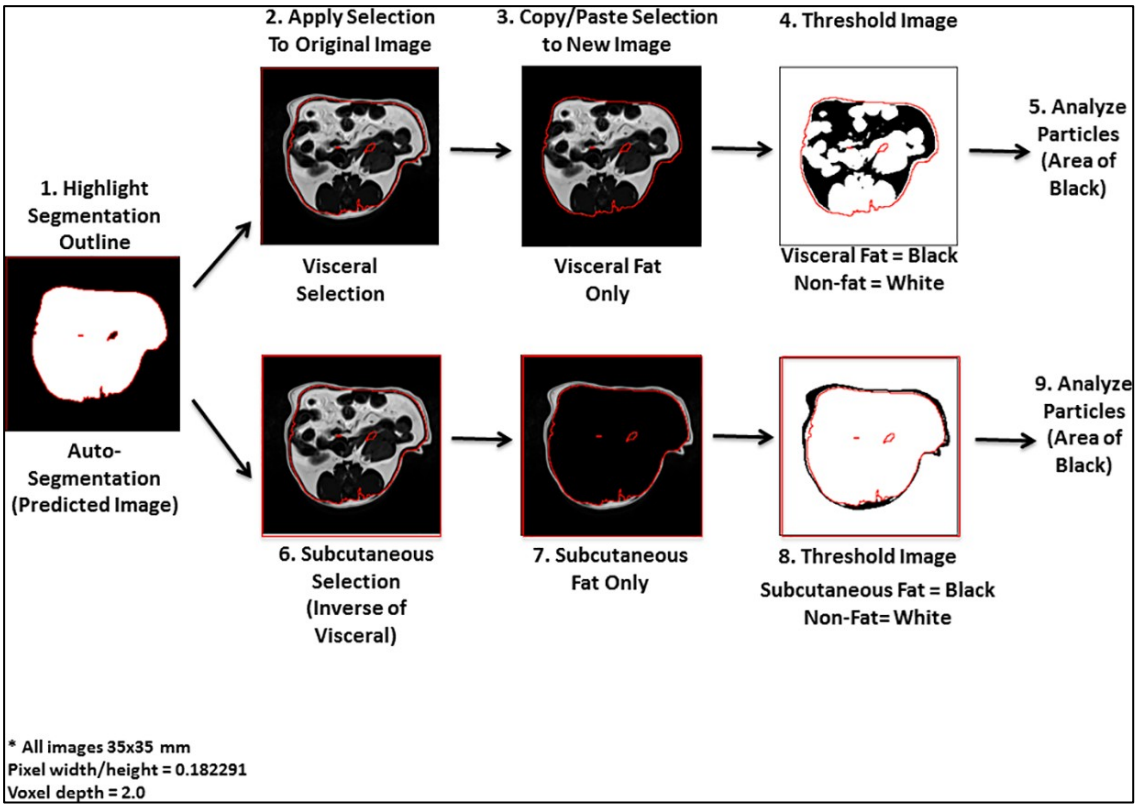

Figure 6.11 *Flowchart showing major steps in automated quantification of visceral and subcutaneous fat. Images were created using the Fiji package of ImageJ. All MR images were set to the dimensions of 35x35 mm, a pixel width and height of 0.182291, and a voxel depth of 2. 1) Make the visceral segmentation image binary (outside = black, visceral component = white), and use the "Create Selection" tool to generate a selection perfectly outlining the visceral component. 2) Use the 'Restore Selection" tool to place this selection on the original MR image, giving an exact outline of the visceral component. 3) Copy and paste this selected area to a new image, making sure it is the same dimensions as the original MR image to ensure accurate quantification (all images 35 x 35 mm; Pixel width/height = 0.182291; Voxel depth = 2.0). 4) Threshold the image to make a binary image in which the fat is one color and everything else is*

another. 5) Quantify this new image by using the "Analyze Particles" tool. 6) On the original MR image with the visceral fat component selected (from 2), use the "Make Inverse" tool to obtain the subcutaneous fat component. 7–9) Repeat the same quantification process as seen in 3–5.

Statistical analysis

Comparisons were made between the automated and manual methods in quantification of visceral and subcutaneous fat as well as between B6 and congenic mice for differences in body weight and fat volume. The Dice score, generated from the ANTs Utilizes Software package, and the root-mean-square error (RMSE) were used for comparing the similarity of the values of fat volumes measured by the manual and automatic methods in the same set of MR images. Correlation analysis was performed to determine the correlation of visceral or subcutaneous fat volumes measured by the two methods. For comparisons of fat volumes between congenic and control mice at multiple slices, two-way ANOVA was conducted. When the F value was significant ($P < 0.05$), Student's t-test was performed to determine differences between the groups. T-test was also used to determine the difference in body weight between the 2 groups of mice. Differences were considered statistically significant at $P < 0.05$.

Chapter 7

Deep learning-based quantification of abdominal subcutaneous and visceral fat volume on CT images

Andrew T. Grainger, Michael H. Quinones, Nicholas J. Tustison, Samantha Epstein, Daniela Fuller, Aakash Jha, Kevin L. Allman, Arun Krishnaraj, Weibin Shi

Manuscript Submitted

7.1 Abstract

Obesity is a growing global epidemic and associated with risk for an array of diseases, particularly metabolic syndrome and type 2 diabetes. Computed tomography (CT) is one of the most accurate methods for quantification of body fat. Here we developed a deep learning-based algorithm using the U-Net architecture to measure abdominal fat on CT images. Sequential CT images spanning the abdominal region of seven subjects were manually segmented for subcutaneous fat (SAT) and the abdominal cavity. The resulting segmentation maps were augmented using a template-based data augmentation approach to create a large pool of data for network training. Network performance was evaluated on both sequential CT slices from three subjects and randomly selected CT images from the upper, central, and lower abdominal regions of 100 subjects through comparison with human rater assessments. Both subcutaneous and visceral fat volumes achieved by the two methods were highly comparable with an overall Dice similarity coefficient of 0.95. Pearson's correlation coefficients between the two methods were 0.99 and 0.98 and the overall percent residual squared error (RSE) were 6.3% and 12.9% for subcutaneous and visceral fat, respectively. For the central abdominal region, the correlation coefficient for visceral fat reached 0.99 and percent RSE was down to 8.5%. Manual segmentation and quantification of subcutaneous and visceral fat on 271 CT slices took 22 hours while automated segmentation took ~5 min. Our data demonstrates the accuracy and efficiency of deep learning in quantifying abdominal fat, particularly in the central abdominal region containing more fat.

7.2 Introduction

Obesity, defined as excessive fat accumulation in the body, is a growing global epidemic with particular impact on the US population which has experienced a marked increase in obesity levels over the last 50 years[199,212]. It is associated with increased risk for

a variety of chronic diseases, including metabolic syndrome, type 2 diabetes, cardiovascular diseases, and cancer[198]. Anthropometric measurements such as body mass index (BMI), waist circumference, and waist-to-hip ratio are used to diagnose obesity. However these indirect measurements do not account for weight from skeletal muscle, nor do they distinguish between differential fat distribution such as visceral and subcutaneous fat. Fat accumulation in the abdominal region has been shown to be more closely associated with increased risk of coronary artery disease and type 2 diabetes[4,5(p)].

Computed tomography (CT) is an imaging modality that can easily distinguish fat from other tissues and thus allow for accurate measurement of fat and non-fat tissue amounts in the body[6]. Quantification of body fat volume using CT involves analysis of many cross-sectional slices across the region of interest so manual measurement of fat volume with this modality is a laborious task. To simultaneously expedite the process and reduce subjective influences from observers, several semi-automated algorithms have been developed for quantifying body fat[7–10]. However, nearly all of the algorithms are still dependent on expert knowledge for tuning the features of images and their accuracy and reliability are often low.

Deep learning using convolutional neural networks has gained recent popularity in the literature for tackling problems in a multitude of areas, including image recognition, classification and segmentation[202]. Development of deep learning algorithms relies on large cohorts of training data to identify important features of targets for predictions in new data. ANTsRNet is a collection of deep learning network architectures ported to the R language and built on the Keras framework[213]. We have previously applied ANTsRNet and other open-source software packages developed by the Advanced Normalization Tools team to provide a comprehensive protocol for automatically segmenting abdominal fat using mouse

MR images[214]. Here, we tested the hypothesis that deep learning could accurately quantify abdominal fat on human CT images.

7.3 Results

7.3.1 U-net-Based Deep Learning for Subcutaneous Fat Selection.

A modified deep learning approach with a template-based data augmentation strategy was employed to measure abdominal fat volume on CT images. Under this modified approach, CT images from seven patients were used to construct a representative template that allows each individual training image to be propagated to the space of every other individual training image so as to expand a training data set of size N to N^2. The U-net model was trained with 613 axial CT images from the seven full scans covering the entire abdominal region. We then validated the learned U-net model by analyzing 271 images from three separate full scans and 300 images randomly selected representing the upper, middle and lower abdominal regions of 100 subjects that were not used in the training. The U-net-based algorithm successfully generated the selections designating the subcutaneous fat and abdominal cavity regions that were highly consistent with the manually generated selections created for the same input images (**Figure 7.1**). These area selections were then used to quantify both subcutaneous and visceral fat volumes.

We calculated Dice coefficient values that measure the level of overlap between the manual and automatic segmentation images. The overall average Dice coefficient value was 0.94, suggesting a high degree of similarity in selection shape and area (min Dice = 0.80; max Dice = 0.98) (**see S7.1**). The average Dice coefficient at each individual slice also showed a high degree of similarity between manual and automatic segmentation images across the entire abdominal scan.

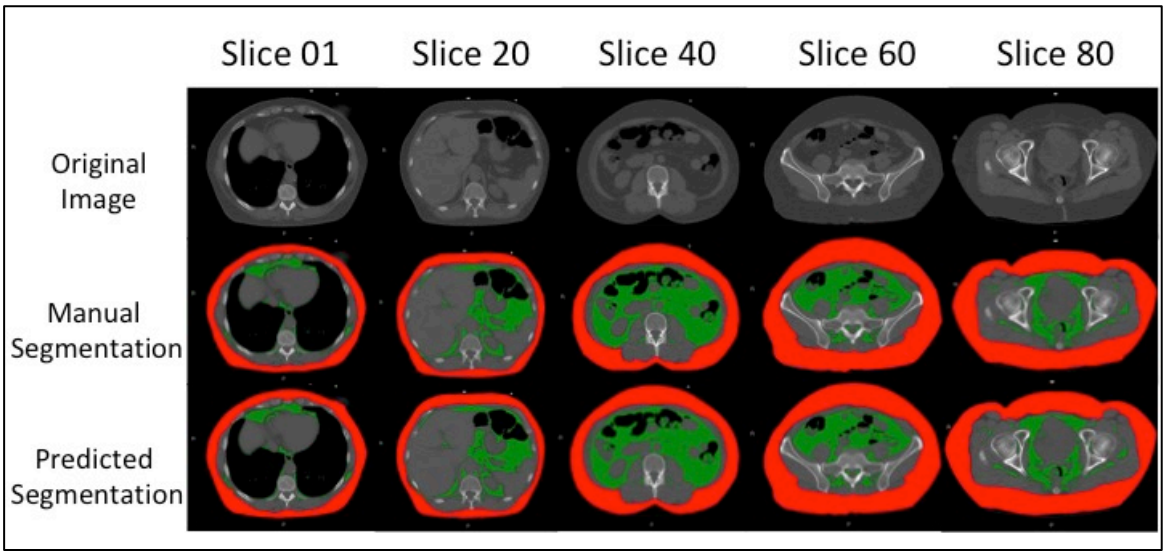

Figure 7.1 *Deep learning properly generates SAT segmentation images in testing CTs at different levels. Representative images taken every 20 slices from a testing subject show consistency between the manual and automatic methods in segmenting SAT on CT images. The red area denotes the SAT segmentation image that will be used for quantification and the green area denotes quantified VAT after image thresholding. Predicted: segmentation made by deep learning.*

7.3.2 Comparison of Fat Volume Measured from Manual and Deep Learning-Based Methods.

Initial comparison was made on measurements from consecutive CT slices from three test subjects for the measurement of SAT volumes. As shown in **Figure 7.2a**, fat

volumes measured on each sequential slice of all 3 subjects were comparable between the two methods. For each of the three individual scans used for testing the model, the total SAT volume was comparable between the two methods (**Figure 7.2b**). The difference for scan 1 was 59,155 mm^3 or 0.97% of the total volume, the difference for scan 2 was 692,520 mm^3 or 5.4%, and the difference for scans 3 was 214,197 mm^3 or 2.1% of the total volume. The average difference was 28,252 mm^3 or 3.0% of the total volume. Pearson's correlation coefficient suggests a high agreement in fat volumes measured on 265 slices by the two methods (**Figure 7.2c**). The residual, which is the difference between the manual and automated measurements, was also determined at each sequential slice for the 3 test subjects. The average percentage residuals for SAT fluctuated between 0~10% among all individual slices (**Figure 7.2d**).

The volumes of VAT (visceral fat) on sequential slices were also comparable between the two methods for all three test scans (**Figure 7.3a**). The difference between the total VAT volumes achieved from the auto-segmentation and that from the manual segmentation was also small (**Figure 7.3b**). The difference for scan 1 was 59,155 mm^3 or 2.7% of the total volume, the difference for scan 2 was 692,520 mm^3 or 9.3% of the volume, and the difference for scans 3 was 214,197 mm^3 or 5.3% of the total volume. The average difference between fat volumes achieved from the two methods was 282,521 mm^3, which accounts for 6.2% of the average visceral fat volume. Pearson's correlation analysis showed a high agreement in measurement results from the two methods (**Figure 7.3c**). The average percentage residuals between the two methods for VAT measurement fluctuated mostly lower than 20% among individual slices, although a few residuals are above 30% at the upper abdominal region (**Figure 7.3d**).

Performance of the deep learning-based algorithm was further validated with CT images from additional 100 subjects randomly selected from the University's PACS

database. One image was randomly chosen from each of the upper, middle and lower abdominal region of an individual. For 12 CT slices, quantification of adipose tissue compartments was impossible with the automated method because subcutaneous fat area was discontinuous or the muscle layer was incomplete. The image training did not include images in which the SAT was discontinuous, and therefore the algorithm assumes a continuous SAT layer. When a continuous SAT layer is not present, it artificially creates one and over-represents the area in which the SAT is predicted to reside. For the rest of the 288 slices, the mean Dice value was 0.95 (minimum Dice = 0.91, maximum Dice = 0.97) (**Table 7.1**). Correlation analysis showed a high degree of agreement between the two methods in measurements of SAT (0.99) and VAT (0.99). The random squared error was 8.0% for SAT and was 11.3% for VAT.

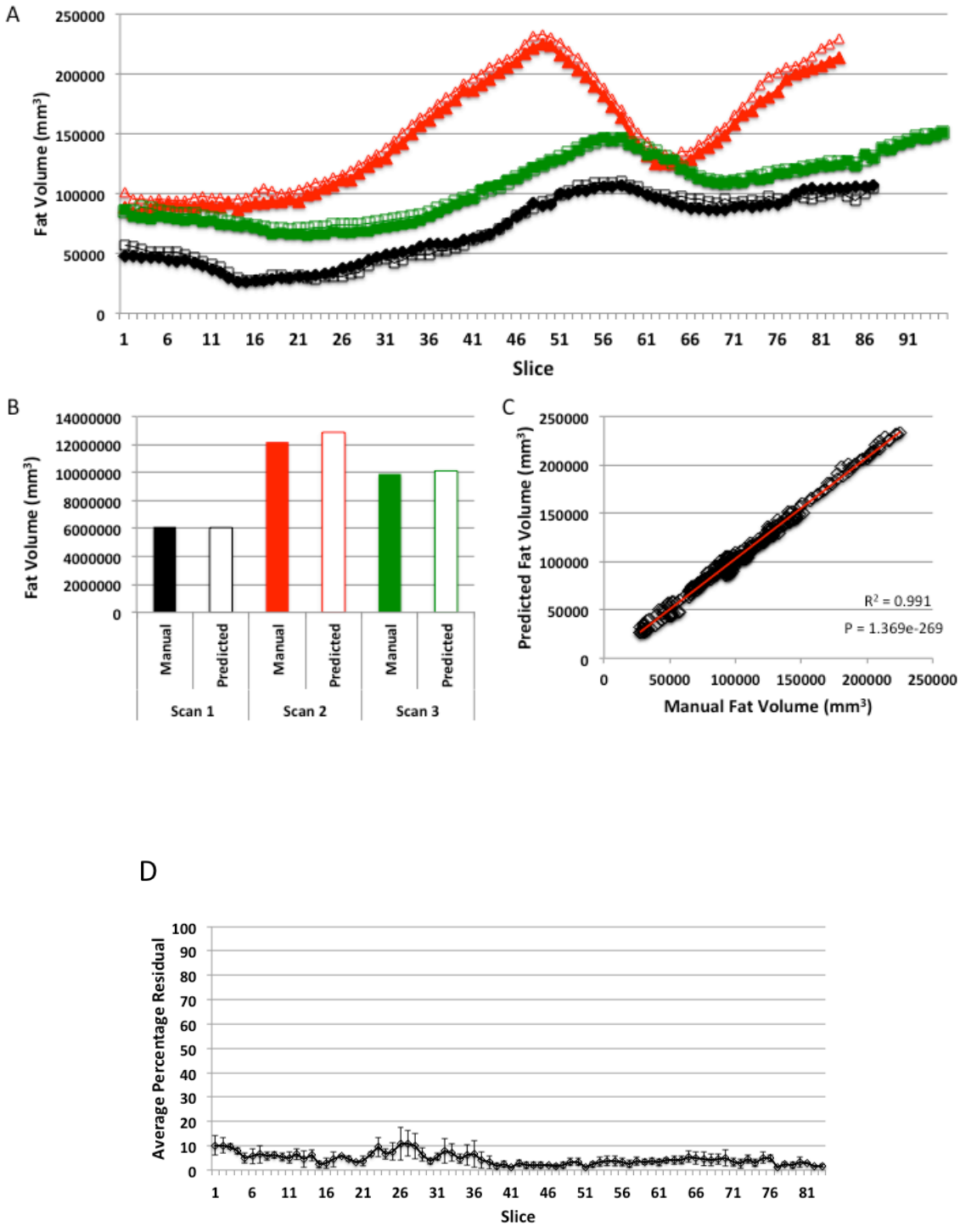

Figure 7.2 *Comparison between the predicted and manual segmentation images in quantification of SAT. Images segmented using the predicted segmentation image can adequately quantify SAT volumes. A) Comparison of SAT volumes from images using*

manual or predicted segmented images for SAT quantification, from the top (slice 1) to the bottom of the abdominal cavity in three independent CT scans (black = scan 1, red = scan 2, green = scan 3; solid = manual, hollow = predicted). B) Comparison between SAT volumes from manual or predicted segmentations (black = scan 1, red = scan 2, green = scan 3; solid = manual, hollow = predicted). C) Linear regression analysis of fat volumes from manual or predicted segmentation images (R^2 = 0.991; p = 1.369e-269). Each data point represents one CT slice. D) The percent residual at each sequential CT slice covering the abdominal region for the 3 test subjects. It reflects the difference between measurements of SAT from manual and predicted segmentations. Results are means ± SE for 271 consecutive images from 3 subjects.

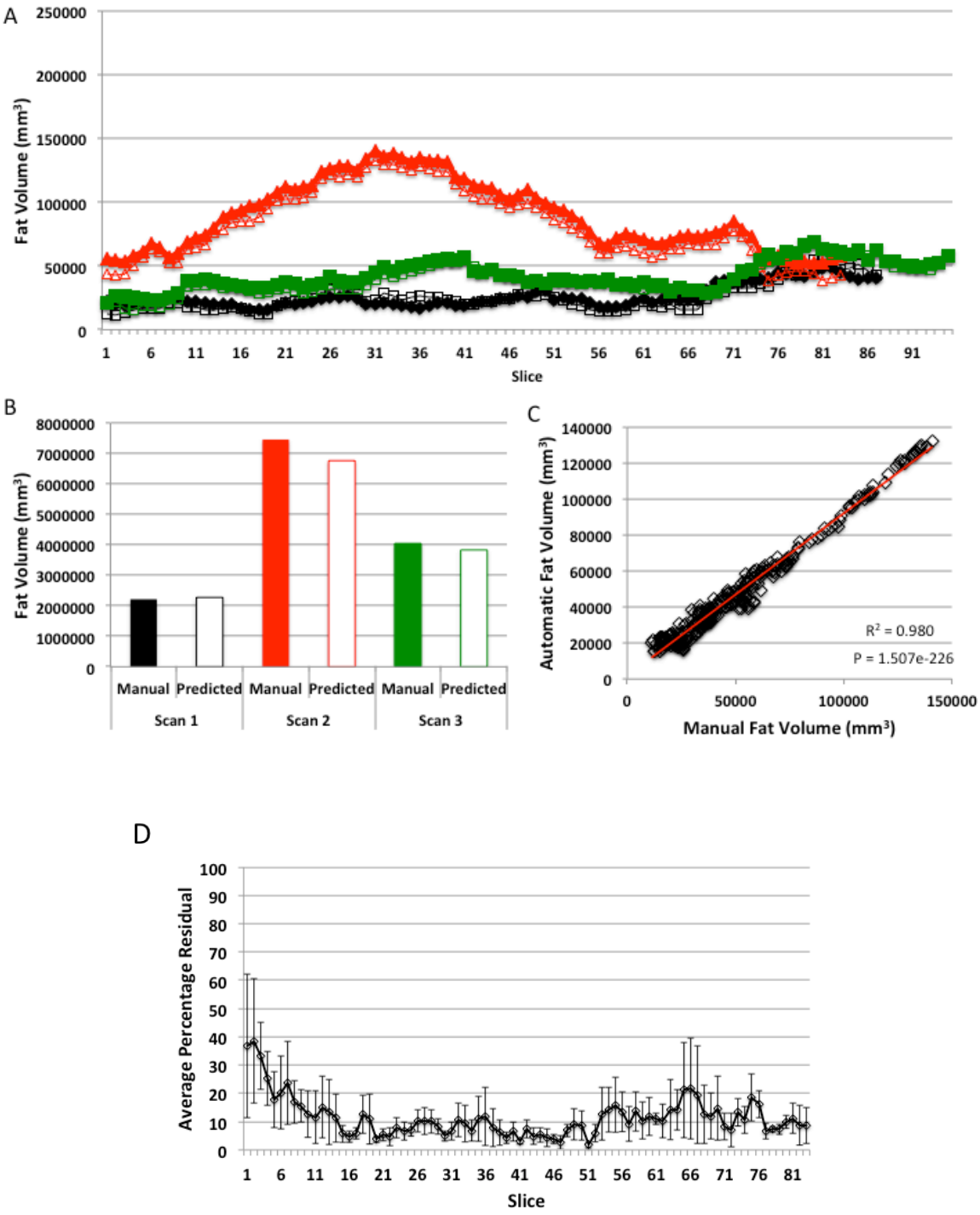

Figure 7.3 *Automated segmentation images can adequately quantify both slice-specific and total VAT volume. A) Comparison of VAT volumes calculated from manual and automated segmentation images from the lung (slice 1) to the end of the*

abdominal cavity in three independent CT scans (black = scan 1, red = scan 2, green = scan 3; solid = manual, hollow = automated). B) Comparison between total VAT volumes (black = scan 1, red = scan 2, green = scan 3; solid = manual, hollow = predicted). C) Linear regression analysis of fat volumes (R^2 = 0.980; p = 1.506e-226). Each data point represents one CT slice image. D) The percent residual at each sequential CT slice for VAT. Data is the average percentage residual ± SE for 271 slices from 3 CT scans.

Table 7.1: *Prediction Image and Fat Volume Similarity of a Validation Cohort*

	Average Dice Value	SAT Volume R^2	SAT Volume P-Value	VAT Volume R^2	VAT Volume P-Value	Percent RSE SAT Volume	Percent RSE VAT Volume
Validation Cohort Images (n = 288)	0.953 ± 0.001	0.992	1.84E-100	0.987	5.54E-92	7.965	11.30
All Validation Images (n = 559)	0.944 ± 0.002	0.994	2.49E-217	0.989	8.85E-193	5.494	8.51

Image similarity (Dice) and fat volume accuracy (R^2 and P-Value) of images generated using the neural network. The mean difference between the manually calculated fat volume and predicted fat volume (Residual Standard Error (RSE)), and the percentage of total volume that is calculated differently between the two segmentation images. Validation cohort data is calculated from a validation cohort consisting of 50 adult males and 50 adult females with a large range of age and BMIs. All validation images consisted of the combination of all images from the full abdominal scans from three individuals and the validation cohort. Values shown are from bin 2, or the central region of the abdomen spanning from below the lungs to the top of the hip.

7.3.3 Computational Time

Manual segmentation and quantification of subcutaneous and visceral fat on 271 images from 3 full CT scans took 22 hours. It took ~5 minutes to perform the same steps on the same CT images using the deep learning-based algorithm.

7.3.4 Correlations between Abdominal Fat Volumes and Body Mass Index (BMI)

Correlations of total, subcutaneous and visceral fat volumes with BMI were evaluated using data from the above validation cohort. Fat volumes in the abdominal region from the disappearance of the lung to the appearance of the hip bone were measured using the automated method. This abdominal region approximately corresponds to the range of lumbar 2-5 vertebrae where visceral fat is measured with CT[215,216]. BMI was significantly correlated with total (R^2=0.145; P<0.001) and SAT volumes (R^2=0.246; P<0.001) (**Figure 7.4A,B**). There was no correlation between BMI and VAT volume (R^2=0.0134; P=0.144; **Figure 7.4C**). Because the amounts of abdominal fat vary between individuals, fat volume was normalized to non-fat mass measured on each CT slice for all subjects to account for the influence of abdominal dimensions on individual variations in abdominal fat. After normalization with nonfat mass, total fat showed an improved association with BMI and subcutaneous fat showed a reduced association with BMI based on R^2 and P values (**Figure 7.4D,E**). No correlation was found between visceral fat volume and BMI (**Figure 7.4F**).

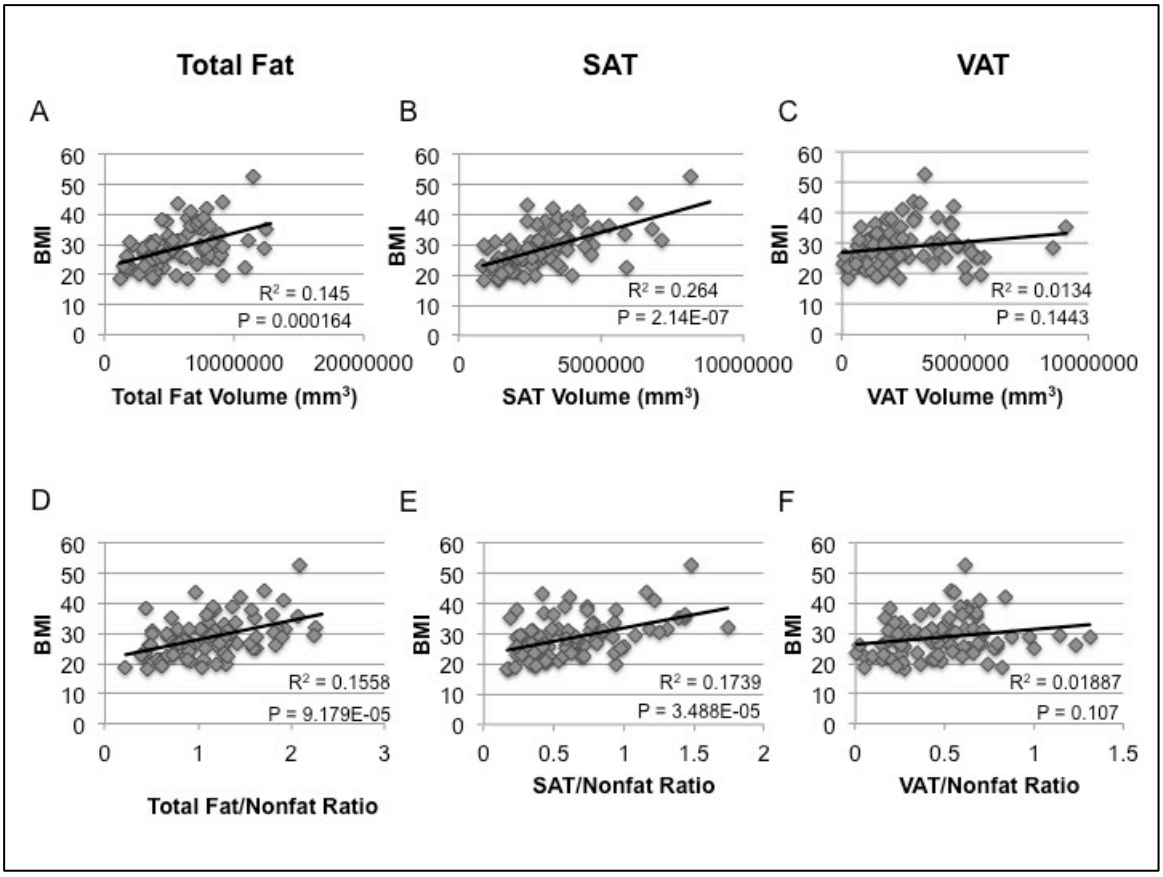

Figure 7.4 *Total fat and SAT are correlated with BMI. Correlation of BMI with fat volumes quantified from all slices within bin 2 from the individuals used for the validation cohort (n = 95 people) A-C) Total fat volume, SAT volume, and VAT volume calculated using the predicted segmentation images. D-F) Total fat volume, SAT volume, and VAT volume were normalized by dividing by the nonfat volume within the same region of interest (bin 2). This adjusted value was then used for correlation as a measure of the ratio of fat to nonfat within an individual.*

7.4 Discussion

The steady increase of central obesity as a worldwide epidemic and its close association with risk of cardiovascular disease and type 2 diabetes demands an efficient approach for abdominal fat quantification. Common anthropometric measurements such as body mass index, waist circumference, and waist-hip ratio cannot accurately discern adipose tissue distribution in the body. The noninvasive modalities MRI and CT allow accurate measurements of body fat and non-fat compositions. However, body fat is barely quantitatively assessed in clinical practice partially due to the laborious nature of the current available methods. We recently have used the U-Net architecture to accurately measure the abdominal fat volume of mice using MR images[214]. However, a major challenge with MR is that T1 images, commonly used for fat analysis, are not calibrated so the image intensity can vary from acquisition to acquisition and scanner to scanner. In contrast, CT is a calibrated image and Hounsfield units reflect tissue physical density. Here we have successfully applied the deep learning-based approach to the measurement of abdominal fat on CT scans in an efficient and quick fashion. Volumes of visceral and subcutaneous fat measured with our algorithm have shown a high degree of consistency with those measured by the manual method.

A few deep learning-based methods to measure body fat on CT scans have been reported[217–219]. Compared with the previous studies, our present study has several features: First, the inclusion of all CT slices in the abdominal region was performed as compared to one or few slices used in other studies. Previous studies have restricted training and evaluation to one or a few selected slices in the abdominal region, and this both prevents accurate evaluation of regions outside of the trained region and provides incomplete picture of phenotypes in the region. Fat distribution varies greatly in the abdominal region of obese subjects so obesity and overall fat content may not be accurately

reflected from one or few slices. Thus, our method of including all CT slices should benefit precision and improve sensitivity for detecting change over time in abdominal fat.

Second, this study employed a template-based data augmentation strategy whereby images sampled were used to construct a representative template that was optimal in terms of both shape and intensity[207]. This approach permits a substantial augmentation of the training dataset and overcomes the limitation of deep learning in most cases which require large-scale training datasets. Despite fewer CT slices used for the training, our algorithm has shown a comparable or even better performance than the previous deep learning-based methods for quantification of body fat using CT images[217–219].

Our algorithm was accurate in quantifying both subcutaneous fat and visceral fat across the abdominal region on which it was trained. There was no significant difference between performance at the upper abdominal region versus performance at the central or lower abdominal region for subcutaneous fat. However, visceral fat showed a larger deviation than subcutaneous fat based on the correlation coefficient and residuals values yielded from comparison between the two measurements. Inconsistency is particularly obvious in the upper abdominal region where low adiposity results in a larger percentage residual. The partial volume effect of gas and watery contents within the gastrointestinal tract whereby visceral fat is attached might also contribute to the inconsistency in visceral fat measurement[220]. The finding that subcutaneous fat in the upper abdominal region did not show such a deviation supports this speculation.

It's extremely challenging to manually segment visceral fat on numerous CT slices due to its irregular shape and extensive distribution within the abdominal cavity. However, it's easier to separate fat from nonfat elements (air, background, waterish tissue, and bone) and isolate visceral fat on CT slices through thresholding. Thus, we chose to directly segment subcutaneous fat and use a segmentation corresponding to the abdominal cavity

for visceral fat measurement. The chance of overestimating visceral fat volume from other fat deposits such as bone marrow fat and intermuscular fat should be minimal because the lumbar vertebrae and pelvic bones are within the abdominal and pelvic wall while visceral fat is located within the abdominal cavity. Also, vertebrae and pelvic bones are cancellous bones containing red bone marrow that have higher CT values than fat.

Despite the success, our algorithm has limitations: First, there is a flaw associated with the bone window thresholding, which was used to separate fat from nonfat compositions. With this method, fat was readily separated from gas, empty space and bone. However, it was observed that fat and soft tissues were not completely separable in some cases due to their closeness in density and partial volume effect. Body fat has a pixel density value of −190∼−30 HU[221], compared to the values of 20∼50 HU for soft tissues[222]. Second, 2D CT slices were used to measure the volume of 3D adipose tissue. Its irregular shape and attachment to the constantly moving gastrointestinal tract makes hard to accurately measure.

BMI is the most widely used measure of body adiposity in clinical practice. Here, the associations of BMI with abdominal fat volumes directly measured by CT were tested with the test cohort. The present result shows that BMI is only moderately associated with abdominal total and subcutaneous fat and has no association with visceral fat. In studies of larger cohorts, BMI has also been found to be more correlated with subcutaneous fat over visceral fat[215,223,224]. These results suggest that BMI is not a reliable marker of abdominal fatness and is a poor parameter of visceral fat.

In summary, we have demonstrated the accuracy of our deep learning based algorithm for quantifying abdominal fat on CT scans. The algorithm has greatly sped up the process of measuring abdominal fat volume. We have shown the possibility of using a relatively small dataset to effectively train a neural network to segment different abdominal

fat deposits. This has important clinical implications as machine learning can be readily applied to other measurable traits using medical imaging. Despite the success, the algorithm has limitations: it performs poorly when an individual has low subcutaneous fat that fails to form a continuous layer and when a body part is interfered with by another body part or foreign object.

7.5 Methods

CT Images

Abdominal CT scans of 110 patients (half men, half women), aged 60 ± 16 years (range: 19-93 years), were retrieved through the Picture Archiving and Communication System (PACS) at the University of Virginia. The scan parameters varied among patients with a tube current of 2-395 mA, slice thicknesses of 2.82 ± 1.75 mm (range: 1.25-5 mm), and a tube voltage of 120 kV. 81% of the CT scans were enhanced with use of iodine-based contrast agents. Average BMI was 28.1 ± 0.8 kg/m^2 (range: 17.2-52.9) for men and 28.8 ± 1.1 (range: 17.2-52.9) kg/m^2 for women. This large variation in tube current, slice thickness, and BMI was purposefully chosen to ensure the algorithm could properly segment any novel CT. All procedures were conducted in compliance with the Health Insurance Portability and Accountability Act (HIPAA) and were included within an IRB approved retrospective study protocol. CT images were de-identified before use to protect patient identity.

Manual segmentation and quantification

The areas corresponding to the subcutaneous fat and the abdominal cavity on each of the training CT images were manually segmented by five coauthors on this article using

ITK-SNAP. CT images were adjusted through windowing at PACS to a greyscale at which fat was visually distinguishable from nonfat components (bone, air, background, soft and watery tissue) and then saved as TIFF (Tagged Image File Format) images for fat quantification. A bone window was found to work better than other windows in distinguishing fat from non-fat tissues on TIFF images.

For quantification of fat, we developed an ImageJ-based strategy using thresholding around a static intensity window corresponding to fat on images adjusted to the bone window. A flowchart explaining the steps needed to quantify total fat can be seen in **Figure 7.5**. Briefly, images were adjusted to binary through thresholding with fat being black and all other tissues white. Black color areas were then calculated using the "Analyze Particles" function of ImageJ. This function also allows all stacked images from a subject to be analyzed for fat areas at once. The resulting fat area for a slice was multiplied by slice thickness to give the fat volume of that slice. The summation of the fat volumes measured from all consecutive slices gave the total fat volume of a subject.

For segmentation of subcutaneous fat (SAT), we manually outlined the area between the skin and the abdominal muscles. Thresholding the image and quantifying the fat area within this selection represents the SAT area. Visceral fat (VAT) is the fat within the abdominal cavity. Due to its irregular shape and extensive distribution in the abdominal cavity, VAT was hard to manually segment. Therefore, the area corresponding to the abdominal cavity was outlined and the VAT volume was calculated though the quantification of fat within this selection.

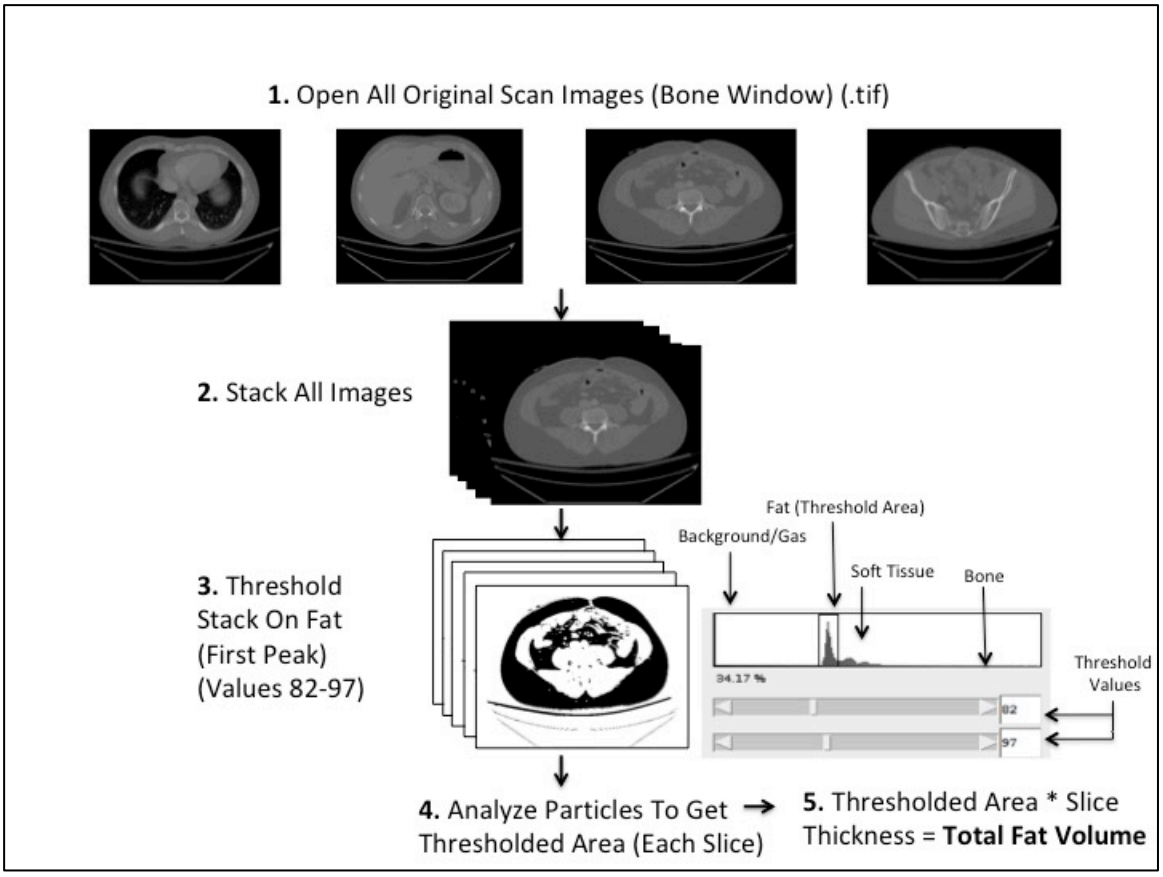

Figure 7.5 *Flowchart illustrating the steps involved in quantification of total fat using CT images. CT images were saved in the preset bone window as .tif files and opened in Fiji. All images are 437 x 437 mm² in size (Pixel width/height = 0.8535, voxel depth = 2.0). All images from a full scan were stacked and thresholded on the first major peak to create a binary image highlighting only fat (thresholding values = 82-97). The "Analyze Particles" function was used to measure the black area corresponding to fat. This area was multiplied by the slice thickness to get the total fat volume for an individual slice in the stack.*

Automatic Measurement

The creation of an automated method for measurement of abdominal fat volume consists of multiple steps, including training data preparation, template-based data augmentation, and fat quantification. We employed a strategy similar to the one previously used for segmenting and quantifying abdominal fat of mice on MR images [214]. The complete flowchart of the process is shown in **Figure 7.6**.

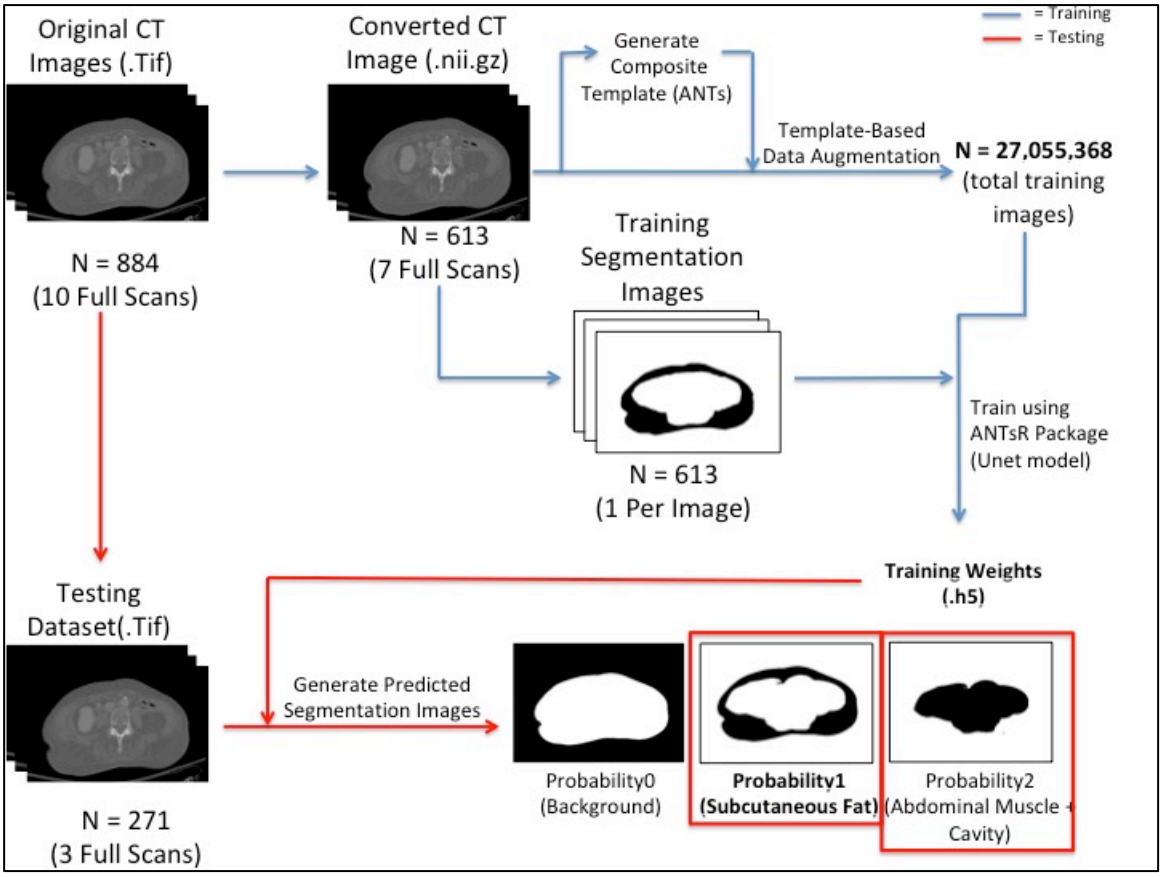

Figure 7.6 *Flowchart illustrating the steps involved in the training of a U-net model for segmentation of SAT in human CT images. CT images were saved in the preset bone window as .tif files and the training set was converted to NifTI format (.nii.gz) for segmentation. A combination of the original .tif images and manually generated*

segmentation images made using ITK-SNAP were augmented using a template-based strategy to create a significantly larger dataset for training. Training was performed using the ANTsR and ANTsRNet R packages. These training weights were then applied to a testing set of novel .tif images to generate predicted segmentation images. Segmentation images used for SAT and VAT quantification are highlighted in red.

Training data preparation: 613 CT images for the entire abdominal region of seven individuals were selected for training. Of the seven subjects, two females and three males had a normal BMI and one female and one male were obese. Original CT images were adjusted at the PACS to bone windows and saved as ".tif" files. These images were converted to the Nifti (.nii.gz) format using the ANTs toolkit (https://github.com/ANTsX/ANTs). Each converted image was segmented into two contoured areas, one for SAT and one for abdominal muscle plus its encircled abdominal cavity, using the open source segmentation tool ITK-SNAP and saved as a separate segmentation image.

Template-Based Data Augmentation and Training: The need for large training data sets is a major limitation associated with development of deep learning algorithms[202]. To achieve a training data set size that is sufficient for properly segmenting total and subcutaneous fat, we employed a template-based data augmentation strategy that we previously used for segmenting abdominal fat of mice on MR images[214]. With this strategy, we are able to augment the original 613 CT images to 27,055,368 images. These images are generated by allowing transformations to/from each individual image to the template space as the template is created. This permits a propagation of the training data to the space of each individual image. In the simplest case, the training data is used to construct the template and then each individual training data is propagated to the space of every other

individual training data. In this way, a training data set of size N can be expanded to a data set of size N^2. Using this strategy, we were able to augment our original 613 images to 613^2, or 375,769 level one images. The second level of augmented data created two horizontal flip images per level one image as well as 36 rotation images per flip image. In this way, the 375,769 images were further augmented to the final 27,055,368 level two images (375,769 * 2 horizontal flip images per image * 36 rotation images per flip image).

Training was performed using a U-net-based model and with the ANTsRNet and Keras packages for R using a Tensorflow backend, as was done previously[214].

Validation dataset: The accuracy of the deep learning-based algorithm in segmenting subcutaneous fat was validated with sequential abdominal CT images from three separate subjects and randomly selected CT images for the upper, central and lower abdominal regions of 100 subjects. Manual measurement results were used as the ground truth for comparisons with the automated measurement results. CT images were prepared as described above and input into the trained U-net. Novel segmentation images generated from the algorithm were evaluated for accuracy in quantification of SAT and VAT using a macro developed for the Fiji package for Image J[197]. The steps for quantifying subcutaneous and visceral fat using the macro are depicted in **Figure 7.7**.

Briefly, all original images from a subject were opened and stacked (Stack 1), all segmentation images corresponding to the SAT area (Probability1.nii.gz) were opened and stacked (Stack 2), and all segmentation images corresponding to the abdominal cavity (Probability2.nii.gz) were opened and stacked (Stack 3). For SAT area: Stack 2 was made binary, a selection around the segmentation area was generated using "Create Selection", this selection is placed onto the original image using "Restore Selection", the image is thresholded on static values corresponding to the intensity of fat in bone-windowed images (82-97), and "Analyze Particles" is used to calculate the fat area in the selection. This

process is repeated for all corresponding images from Stack 2 and Stack 1 to generate SAT area calculations for every image in the stack. For VAT area: An identical process is performed with Stack 3 and Stack 1 to generate VAT area calculations for every image. SAT and VAT volumes are calculated as the SAT or VAT area * slice thickness.

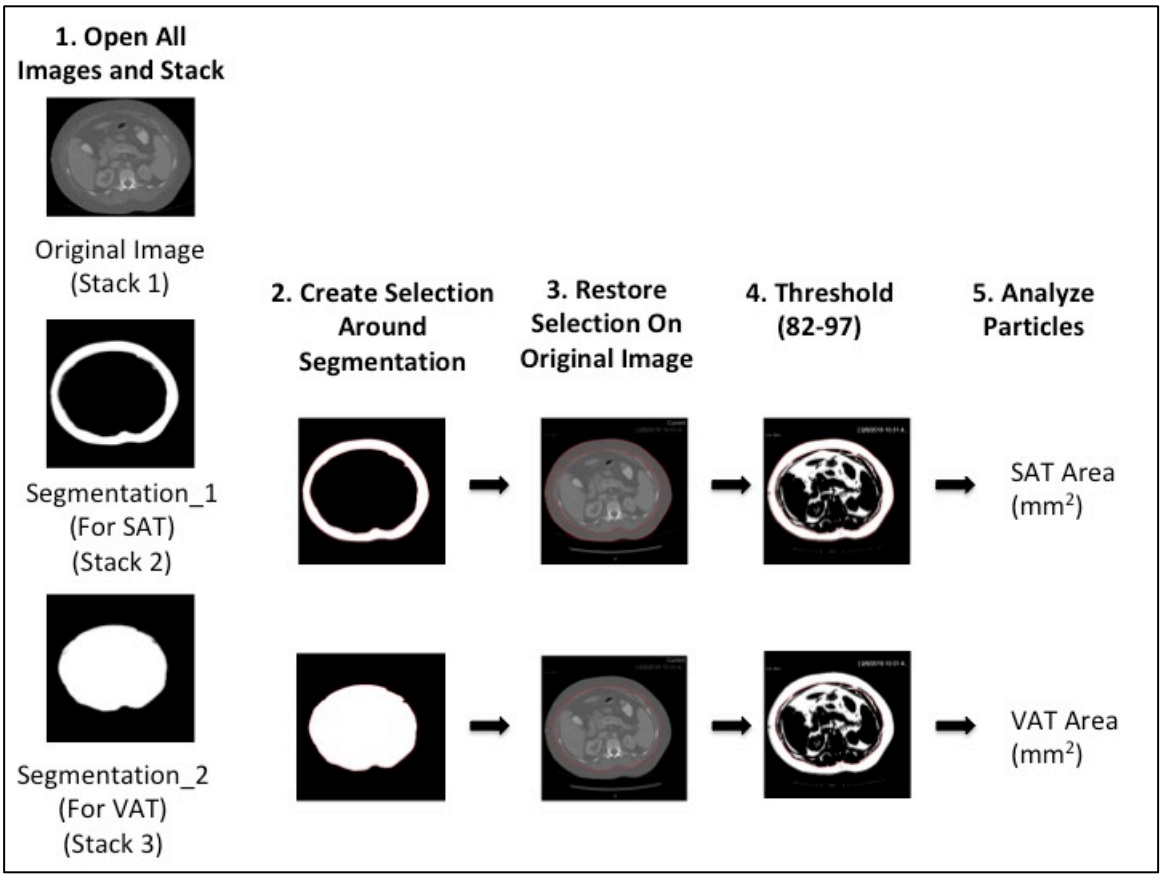

Figure 7.7 *Flowchart describing SAT and VAT area using a macro developed for ImageJ. All original images from a subject were opened and stacked (Stack 1), all segmentation images corresponding to the SAT area (Probability1.nii.gz) were opened and stacked (Stack 2), and all segmentation images corresponding to the abdominal*

*cavity (Probability2.nii.gz) were opened and stacked (Stack 3). For SAT area: Stack 2 was made binary, a selection around the segmentation area was generated using "Create Selection", this selection is placed onto the original image using "Restore Selection", the image is thresholded on static values corresponding to the intensity of fat in bone-windowed images (82-97), and "Analyze Particles" is used to calculate the fat area in the selection. This process is repeated for all corresponding images from Stack 2 and Stack 1 to generate SAT area calculations for every image in the stack. For VAT area: An identical process is performed with Stack 3 and Stack 1 to generate VAT area calculations for every image. SAT and VAT volumes are calculated as the SAT or VAT area * slice thickness.*

Statistical Analysis

Comparisons were made between the automated and manual methods in quantification of visceral and subcutaneous fat volumes. The Dice metric was used to determine the similarity between a manually generated segmentation image and an automatically generated one. If two segmentations completely overlap, the Dice score is 1; it is 0 if there is no overlap. The residual was determined from the difference between manually measured and automatically measured fat volume for each slice. In addition, Pearson's correlation analysis was done to determine correlations between the manually and automatically generated fat volumes, as reported[214].

Chapter 8

Closing Remarks and Future Directions

8.1 Summary

A great deal of work has been done to develop a thorough understanding of the general processes involved in atherosclerosis, implicating pathways in blood vessel morphogenesis, lipid metabolism, nitric oxide signaling and inflammation, as well as many others[30,39,147,225–231]. However, despite knowing a large number of the players involved in plaque formation, our current preventative therapeutic strategies remain largely unsuccessful. Therefore, more work is required to help develop novel therapeutics.

The most effective way of improving our understanding of cardiometabolic disorders is through discerning the genetic architecture underlying the genetic associations observed in mouse linkage analyses and GWAS. However at present, this remains a difficult task. This is particularly true in mice where the QTLs have large confidence intervals spanning tens of megabases. More recent advances in technology has allowed for high-throughput dense genotyping and gene expression in mice. Because of this, our ability to prioritize candidate genes for aortic lesion QTLs has greatly improved. Despite the quantity of QTLs that have been generated, a large amount of work remains uncovering the genes causing the observed linkage. Therefore, a large amount of the heritable component of atherosclerosis in mice remains undiscovered. In these studies, we hypothesized that through generating novel linkage data using rarely used inbred strains, we would be able to discover novel insights into the heritable components of atherosclerosis and its associated cardiometabolic syndromes. We accomplished this through QTL analysis of aortic lesion size, carotic lesion size, hyperlipidemia, and type 2 diabetes. Furthermore, we hypothesized that through modern strategies utilizing the large amounts of publicly available data, we could uncover novel genes influencing atherosclerosis in a fashion hitherto unutilized. We accomplished this through a combination of haplotype analysis and a novel strategy employing gene expression, aortic lesion size correlation, and eQTL data to discover *Mep1α.*

Finally, we hypothesized that by developing novel, more accurate ways to quantify abdominal fat deposits, we could create a platform in which to improve our understanding of obesity, the cardiometabolic disease with the highest co-occurrence with atherosclerosis. We accomplished this by utilizing a combination of machine learning-based segmentation and medical imaging in the form of MRIs and CTs to quantify SAT and VAT fat volumes in mice and humans.

8.2 An F2 Intercross Using A Rare Susceptible Strain Discovered Many Novel Insights For Dyslipidemia, T2D, Carotid Atherosclerosis, and Aortic Atherosclerosis

In a F2 intercross between BALB/cJ and SM/J mice, QTL analysis for plasma lipids and glucose revealed one significant QTL on chromosome 9 for fasting glucose, named *Bglu17*, that coincides with a significant QTL for HDL (high-density lipoprotein) and a suggestive QTL for non-HDL cholesterol levels. Haplotype analysis revealed that "lipid genes" Sik3, Apoa1, and Apoc3 were probable candidates for *Bglu17*. Our findings strongly suggest that one or more of these “lipid genes” might be the causal gene(s) of *Bglu17*, contributing to variation in fasting glucose levels. Although it is unknown how they affect glucose homeostasis, one probable effect path is through the influence on plasma lipid levels, which then predispose variation in glucose-related traits. The current observation of the significant correlations of fasting glucose levels with HDL, non-HDL cholesterol, and triglyceride levels in this cross supports this speculation. Plasma lipid levels, especially non-HDL cholesterol, of the F2 mice were significantly elevated on the Western diet, so were the fasting glucose levels. When fed the Western diet, *Apoe*$^{-/-}$ mice display a rapid rise in non-HDL cholesterol levels, often reaching their peak within a couple of weeks (unpublished data), whereas their blood glucose levels rise more slowly and gradually within 12 weeks[133,134]. This difference in onset suggests a causal role for plasma lipids in the rise of

blood glucose in the *Apoe*$^{-/-}$ mouse model. From these findings we have found that the colocalization of QTLs for both phenotypes and the sharing of potential candidate genes demonstrate genetic connections between dyslipidemia and type 2 diabetes, and have furthered our understanding of how genetics loci can influence multiple cardiometabolic disorders. With this study, we have confirmed our hypothesis that though using a lesser-used inbred strain, we were able to uncover novel insight into the genetics of dyslipidemia and type 2 diabetes.

With the BALB/cJ x SM/J F2 intercross, we performed linkage analysis for carotid atherosclerosis, which is a major cause of stroke. QTL analysis revealed eight loci for carotid lesion sizes. *Cath1* on chromosome 12, *Cath2* on chromosome 5, *Cath3* on chromosome 13, and *Cath4* on chromosome 6 are four significant QTL for carotid atherosclerosis thus far mapped in two *Apoe*$^{-/-}$ mouse intercrosses[122,128]. Performing combined cross-linkage analysis using data from this cross and two previous F2 crosses derived from BALB, C57BL/6J and C3H/HeJ strains, we identified five significant and nine suggestive QTLs. Of them *Cath1*, the QTL on chromosome 12, was the most prominent. Of these QTLs, four novel QTLs, two significant QTLs named *Cath7* on chromosome 5 and *Cath8* on chromosome 9 and two suggestive QTLs, *Cath5* and *Cath6* on chromosomes 15 and 18 respectively, were mapped. Nearly all of these QTLs were mapped in one or more individual crosses, but the combined cross analysis had an increased power of detecting shared QTLs by two or more crosses. Indeed, all five significant QTLs had a higher LOD score than that achieved in an individual cross. The significant correlations of carotid lesion sizes with HDL and non- HDL cholesterol levels suggest that some loci exert effects on carotid atherosclerosis partially through action on lipoproteins. Bioinformatics-based analysis prioritized nine probable candidate genes for *Cath1*. From these findings we demonstrated the polygenic control of carotid atherosclerosis in mice and show that with the integration of multiple intercrosses

sharing the same QTLs, we can enhance our ability to discover loci that are influencing a phenotype. Through this study we further confirm our first hypothesis by showing that with data generated through a rarely used inbred strain, we were able to identify hitherto undiscovered genetic architecture for carotid atherosclerosis.

We also performed linkage analysis in the BALB/cJ x SM/J F2 intercross for aortic lesion size. We discovered that a significant locus, named *Ath49,* was mapped to the H2 complex [mouse major histocompatibility complex (MHC)] on chromosome 17. As the MHC region harbors genes that play an important role in both innate and adaptive immunity, these findings highlight the significance of inflammation in atherogenesis and its potential for developing anti-atherosclerotic therapy. We developed a novel bioinformatics-based strategy using publicly available data from the mouse HDMP project for developing candidate genes at a mouse locus with overlapping aortic lesion size QTLs. Through this strategy we provide a methodology for prioritizing candidate genes at any mouse aortic QTL. With the integration of this method with mouse-human comparative genomics and haplotype analysis, two classical methods of candidate gene selection, we identified a single likely candidate gene for *Ath49*, *Mep1α*. With these findings we have identified the MHC as a major genetic determinant of atherosclerosis, highlighting the importance of inflammation in atherogenesis, and show that using a combination of gene expression, comparative genomics, and haplotype analysis, we can uncover candidate genes at any loci. Through this study we further confirm our hypothesis and show that data generated using a rare inbred strain was integral in discovering novel genetic architecture underlying aortic atherosclerosis.

An observation of particular interest is the difference in QTLs mapped for carotid and aortic lesion size we discovered in these studies. This phenomenon has been a recurring observation in our labs' other linkage studies for carotid and aortic lesion size and

leads to the hypothesis that there is unique genetic architecture affecting plaque formation depending on the location of the atherosclerotic lesion. Cellular components of atherosclerotic lesions in the carotid differ significantly from those seen in the aorta, further supporting this hypothesis[232]. Moreover lesion formation is significantly less common in the mouse carotid, even when fed Western diet, as seen from the F2 population's lesion size distributions in Chapters 3 and 4. This phenomenon has remained consistent across the other F2 intercrosses our lab has performed and suggests differing mechanisms required for lesion formation depending on location. Through this study we have provided additional support for these hypotheses and helped discover differences in the genetic architecture that could be unique to carotid and aortic atherosclerosis.

8.3 Modern Bioinformatics Strategies Uncovered a Novel Gene Influencing Atherosclerosis

Our most exciting findings came from investigating *Mep1α,* the candidate gene for *Ath49*. Double knockout mice for *Mep1α* and *Apoe* discplayed a markedly different phenotype in regards to atherosclerosis. Double nockout mice had decreased aortic lesion size, alterations in plaque stability markers, and significant shifts in the plaque's cellular composition. Through our findings, we have discovered a novel gene negatively influencing atherosclerosis. With this study we have the confirmation of our second hypothesis, as through the integration of publicly available bioinformatics data and linkage data from a lesser used inbred strain, we were able to successfully uncover a novel gene influencing atherosclerosis.

8.4 Novel Methods For Fat Quantification Provides a Platform For Improving Our Understanding of Obesity

There is a difference in body weight between C57BL/6 and chromosome 9 congenic mice, created by inserting the C3H/HeJ genome into a B6 background at 15.6–115.6 Mb. To test whether this difference is attributable to body fat, we developed a machine learning-based method of segmenting and quantifying abdominal fat into subcutaneous (SAT) and visceral (VAT) fat volumes using MRI. After proving that the method works, we successfully used it to compare the SAT and VAT volume and distribution between the C57BL/6 and chromosome 9 congenic mice. We showed that chromosome 9 congenic mice have a significant decrease in total SAT and VAT and that this volume difference occurs primarily around the abdomen. With these findings, we have shown that the region on chromosome 9 linked to aortic lesion size, plasma lipid and glucose levels, and body weight is definitively affecting fat volume. Moreover, we have created a novel strategy for quantifying abdominal fat in mice and present a platform for accurately and non-invasively studying obesity in mouse models. Through this study we have confirmed our third hypothesis by showing that we can more accurately discern the difference in fat composition between two groups of mice, and that genetics plays a role in this fat distribution.

CT is a more widely used imaging technique compared to MRI, so when quantifying abdominal SAT and VAT in humans we adapted this machine learning-based strategy we have developed in mice for use with CTs. After proving that this method works we then used it to show that BMI, the current most commonly used metric for diagnosis of obesity in humans, is correlated with total fat and SAT, but not VAT. With these findings, we provide a novel tool for studying obesity in humans by providing a feasible way to quantify a more accurate and potentially more informative phenotype that until now has been unquantifiable.

8.5 Future Directions

With the findings from these collected works, we have opened a large number of doors moving forward. Through the BALB/cJ x SM/J F2 intercross we have uncovered both novel and recurring QTLs for plasma lipids, glucose, and carotid atherosclerosis, however the genes underlying the linkage of these QTLs remain undiscovered. *Cath1* for carotid atherosclerosis is of particular interest, as it was replicated in every F2 intercross' linkage analysis our lab has performed and is the most prominent QTL of the combined cross analysis. We have found an effective strategy for prioritizing candidate genes at overlapping loci and have successfully employed this strategy for aortic QTLs. Moving forward, this strategy has the potential to work for plasma lipids, glucose, and carotid atherosclerosis as well. We have performed one of the three candidate gene selection methods for these QTLs through the use of haplotype analysis for these traits. But candidate genes from mouse-human comparative genomics based on overlap with GWAS data for type 2 diabetes, lipid levels, stroke, or BMI have not been generated. Moreover, candidate genes prioritized based off of gene expression data, phenotype correlation, and eQTLs remain to be created. At the current moment, no publicly available data has been generated on a large cohort of mice for lipid levels, glucose, or carotid atherosclerosis in the same was as it has been for aortic atherosclerosis. Thus the same strategies we used for gene expression-based candidate genes of aortic QTLs cannot be repeated for other cardiometabolic phenotypes at present. If this data became available, we would have the ability to uncover genes that haplotype analysis and mouse-human comparative genomics cannot.

The additional candidate genes discovered by using mouse-human comparative genomics has worked with success in mouse HDL studies, and is a viable method of candidate gene selection[70]. We have generated a comprehensive list of aortic QTLs and looked at each of the syntenic regions in humans for suggestive and significant GWAS associations. However, this has not been done for glucose or carotid QTLs and their

respective GWAS loci. In order to generate candidate genes in this manner, this process can be repeated for plasma glucose and carotid atherosclerosis.

Our gene that we investigated for its impact on atherosclerosis, *Mep1α*, has a much vaster effect that previously anticipated. Most often, a single gene can have significant impact, yet this impact is small[233]. We have observed an uncommon phenotype in which the impact is both significant and large. The primary influence on atherosclerosis appears to be through widespread changes in plaque composition, however the specific pathways this gene is integral for and the cell types contributing to this phenotype remain unclear. Sudies on *Mep1α's* predicted degradome show many potential substrates, so discerning which of these downstream proteins and which cell types are affecting plaque formation will be difficult[184]. Because of this, an unbiased comparison of the proteome of the aortic lesions of *Mep1α*$^{-/-}$.*Apoe*$^{-/-}$ and B6.*Apoe*$^{-/-}$ mice might help highlight which proteins are not being degraded and can aid in uncovering the mechanisms underlying the change we observe. Additionally, conditional knockouts of *Mep1α* in endothelial cells, macrophages, and smooth muscle cells could help eleucidate which cell types are the pirmary contributors to the phenotype we are observing. Much remains on the further charactersiation of this gene's impact on atherosclerosis, but this gene provides a promising novel target for therapeutics.

The strategies we have developed for quantifying SAT and VAT in mice and humans provide an entirely new avenue for studying obesity. Before, the ability to utilize subcutaneous and visceral fat volume, the most accurate phenotype for obesity, was unfeasible due to the time it took to manually segment and quantify. We have created a way to expedite the process to a manageable timeframe, and with this give future studies the possibility to study how genes or pathways affect subcutaneous or visceral fat volume. The applications of this strategy are vast, as any study that has previously used less accurate measurements can supplant them with SAT and VAT volume. Moreover in humans, this

quantification method has the potential to be a novel predictor of metabolic disorders, more accurate than the currently used BMI.

In conclusion, these studies have opened many avenues for the further investigation of the genetic components of atherosclerosis and its related cardiometabolic disorders. We have significantly contributed to this ongoing investigation by providing multiple novel QTLs for plasma lipids, glucose, carotid lesion size, and aortic lesion size. We have developed strategies for accomplishing the future task of discovering the heritable components of atherosclerosis, dyslipidemia, type 2 diabetes, and obesity including candidate gene generation for linkage QTLs and quantification of SAT and VAT volume. Moreover, we have significantly contributed to the understanding of aortic atherosclerosis by implementing our candidate gene generation strategy for an aortic QTL locus on mouse chromosome 17 to uncover a novel and important gene for atherosclerotic plaque development. Our work presented here lays the foundation for future studies to continue discovering the genetic architecture of cardiometabolic disorders.

Appendix A

Supplemental Data

All supplemental data are available at:

https://github.com/atg3qz/Dissertation-Supplementary-Data

Supplemental File 1.1 UCSC Genome Browser track for visualizing the locations of published aortic QTLs.

Supplemental File 2.1 Raw data used for plasma lipid and glucose QTL and correlation analysis in a BALB x SM female F2 intercross.

Supplemental File 3.1 Raw data used for carotid lesion size QTL analysis using data from a BALB x SM female F2 intercross and from combined cross analysis.

Supplemental File 4.1 Raw data used for aortic lesion size QTL analysis in a BALB x SM female F2 intercross.

Supplemental File 5.1 Raw data used for investigation of Mep*1α*'s role in atherosclerosis

Supplemental File 6.1 Raw data used for validation of the machine learning-based model for fat quantification in mouse MRI and for investigation of the difference in fat content between B6 and chromosome 9 congenic mice.

Supplemental File 7.1 Raw data used for validation of the machine learning-based model for fat quantification in human CTs and analysis of the relationship between SAT or VAT and BMI.

References

1. Heart Disease and Stroke Statistics—2017 Update. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5408160/. Accessed September 25, 2019.

2. Heron M. Deaths: Leading Causes for 2016. :77.

3. Ajani UA, Ford ES, Mokdad AH. Prevalence of high C-reactive protein in persons with serum lipid concentrations within recommended values. *Clin Chem*. 2004;50(9):1618-1622. doi:10.1373/clinchem.2004.036004

4. Brons C, Grunnet LG. Mechanisms in Endocrinology: Skeletal muscle lipotoxicity in insulin resistance and type 2 diabetes: A causal mechanism or an innocent bystander? *Eur J Endocrinol*. 2017;176:67-78.

5. St-Pierre J, Lemieux I, Vohl MC, et al. Contribution of abdominal obesity and hypertriglyceridemia to impaired fasting glucose and coronary artery disease. *Am J Cardiol*. 2002;90:15-18.

6. Chan JM, Rimm EB, Colditz GA, Stampfer MJ, Willett WC. Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men. *Diabetes Care*. 1994;17:961-969.

7. Seabolt LA, Welch EB, Silver HJ. Imaging methods for analyzing body composition in human obesity and cardiometabolic disease. *Ann N Acad Sci*. 2015;1353:41-59.

8. Positano V, Gastaldelli A, Sironi AM, Santarelli MF, Lombardi M, Landini L. An accurate and robust method for unsupervised assessment of abdominal fat by MRI. *J Magn Reson Imaging*. 2004;20:684-689.

9. Demerath EW, Ritter KJ, Couch WA, et al. Validity of a new automated software program for visceral adipose tissue estimation. *Int J Obes*. 2007;31:285-291.

10. Kullberg J, Angelhed JE, Lonn L, et al. Whole-body T1 mapping improves the definition of adipose tissue: Consequences for automated image analysis. *J Magn Reson Imaging*. 2006;24:394-401.

11. Kn BP, Gopalan V, Lee SS, Velan SS. Quantification of abdominal fat depots in rats and mice during obesity and weight loss interventions. *PLoS One*. 2014;9:e108979.

12. Burke GL, Savage PJ, Sprafka JM, et al. Relation of risk factor levels in young adulthood to parental history of disease. The CARDIA study. *Circulation*. 1991;84(3):1176-1187. doi:10.1161/01.cir.84.3.1176

13. Sarlund H, Pyörälä K, Penttilä I, Laakso M. Early abnormalities in coronary heart disease risk factors in relatives of subjects with non-insulin-dependent diabetes. *Arterioscler Thromb J Vasc Biol*. 1992;12(6):657-663.

14. Schumacher MC, Maxwell TM, Wu LL, Hunt SC, Williams RR, Elbein SC. Dyslipidemias among normoglycemic members of familial NIDDM pedigrees. *Diabetes Care*. 1992;15(10):1285-1289. doi:10.2337/diacare.15.10.1285

15. Williams RR, Hunt SC, Hasstedt SJ, et al. Are there interactions and relations between genetic and environmental factors predisposing to high blood pressure? *Hypertens Dallas Tex 1979*. 1991;18(3 Suppl):I29-37. doi:10.1161/01.hyp.18.3_suppl.i29

16. Samson SL, Garber AJ. Metabolic syndrome. *Endocrinol Metab Clin North Am*. 2014;43(1):1-23. doi:10.1016/j.ecl.2013.09.009

17. Fihn SD, Gardin JM, Abrams J, et al. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease: A report of the American College of Cardiology Foundation/American Heart Association task force on practice guidelines, and the American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. *Circulation*. 2012;126(25):e354-471. doi:10.1161/CIR.0b013e318277d6a0

18. Rhee E-J, Kim HC, Kim JH, et al. 2018 Guidelines for the management of dyslipidemia. *Korean J Intern Med*. 2019;34(4):723-771. doi:10.3904/kjim.2019.188

19. Kinoshita M, Yokote K, Arai H, et al. Japan Atherosclerosis Society (JAS) Guidelines for Prevention of Atherosclerotic Cardiovascular Diseases 2017. *J Atheroscler Thromb*. 2018;25(9):846-984. doi:10.5551/jat.GL2017

20. Rothwell PM, Cook NR, Gaziano JM, et al. Effects of aspirin on risks of vascular events and cancer according to bodyweight and dose: Analysis of individual patient data from randomised trials. *Lancet Lond Engl*. 2018;392(10145):387-399. doi:10.1016/S0140-6736(18)31133-4

21. Gulizia MM, Colivicchi F, Abrignani MG, et al. Consensus Document ANMCO/ANCE/ARCA/GICR-IACPR/GISE/SICOA: Long-term Antiplatelet Therapy in Patients with Coronary Artery Disease. *Eur Heart J Suppl J Eur Soc Cardiol*. 2018;20(Suppl F):F1-F74. doi:10.1093/eurheartj/suy019

22. Foley DP, Bonnier H, Jackson G, et al. Prevention of restenosis after coronary balloon angioplasty: Rationale and design of the Fluvastatin Angioplasty Restenosis (FLARE) Trial. The FLARE Study Group. *Am J Cardiol*. 1994;73(14):50D-61D. doi:10.1016/0002-9149(94)90633-5

23. Wu C, Camacho FT, Wechsler AS, et al. Risk score for predicting long-term mortality after coronary artery bypass graft surgery. *Circulation*. 2012;125(20):2423-2430. doi:10.1161/CIRCULATIONAHA.111.055939

24. Parikh S, Cohen JR. Perioperative stroke after general surgical procedures. *N Y State J Med*. 1993;93(3):162-165.

25. Lusis AJ. Atherosclerosis. *Nature*. 2000;407(6801):233-241. doi:10.1038/35025203

26. Libby P. Vascular biology of atherosclerosis: Overview and state of the art. *Am J Cardiol*. 2003;91(3A):3A-6A. doi:10.1016/s0002-9149(02)03143-0

27. Libby P, Ridker PM, Hansson GK. Progress and challenges in translating the biology of atherosclerosis. *Nature*. 2011;473(7347):317-325.

28. Lee D-Y, Chiu J-J. Atherosclerosis and flow: Roles of epigenetic modulation in vascular endothelium. *J Biomed Sci*. 2019;26(1):56. doi:10.1186/s12929-019-0551-8

29. Wang L, Luo J-Y, Li B, et al. Integrin-YAP/TAZ-JNK cascade mediates atheroprotective effect of unidirectional shear flow. *Nature*. 2016;540(7634):579-582. doi:10.1038/nature20602

30. Chistiakov DA, Melnichenko AA, Myasoedova VA, Grechko AV, Orekhov AN. Mechanisms of foam cell formation in atherosclerosis. *J Mol Med Berl Ger*. 2017;95(11):1153-1165. doi:10.1007/s00109-017-1575-8

31. Galkina E, Ley K. Vascular adhesion molecules in atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2007;27(11):2292-2301. doi:10.1161/ATVBAHA.107.149179

32. Khan BV, Parthasarathy SS, Alexander RW, Medford RM. Modified low density lipoprotein and its constituents augment cytokine-activated vascular cell adhesion molecule-1 gene expression in human vascular endothelial cells. *J Clin Invest*. 1995;95(3):1262-1270. doi:10.1172/JCI117776

33. Weber C, Schober A, Zernecke A. Chemokines: Key regulators of mononuclear cell recruitment in atherosclerotic vascular disease. *Arterioscler Thromb Vasc Biol*. 2004;24(11):1997-2008. doi:10.1161/01.ATV.0000142812.03840.6f

34. Bobryshev YV. Monocyte recruitment and foam cell formation in atherosclerosis. *Micron Oxf Engl 1993*. 2006;37(3):208-222. doi:10.1016/j.micron.2005.10.007

35. Quehenberger O. Thematic review series: The immune system and atherogenesis. Molecular mechanisms regulating monocyte recruitment in atherosclerosis. *J Lipid Res*. 2005;46(8):1582-1590. doi:10.1194/jlr.R500008-JLR200

36. Bouhlel MA, Derudas B, Rigamonti E, et al. PPARgamma activation primes human monocytes into alternative M2 macrophages with anti-inflammatory properties. *Cell Metab*. 2007;6(2):137-143. doi:10.1016/j.cmet.2007.06.010

37. Gisterå A, Hansson GK. The immunology of atherosclerosis. *Nat Rev Nephrol*. 2017;13(6):368-380. doi:10.1038/nrneph.2017.51

38. Björkbacka H, Kunjathoor VV, Moore KJ, et al. Reduced atherosclerosis in MyD88-null mice links elevated serum cholesterol levels to activation of innate immunity signaling pathways. *Nat Med*. 2004;10(4):416-421. doi:10.1038/nm1008

39. Kavurma MM, Rayner KJ, Karunakaran D. The walking dead: Macrophage inflammation and death in atherosclerosis. *Curr Opin Lipidol*. 2017;28(2):91-98. doi:10.1097/MOL.0000000000000394

40. Allahverdian S, Chehroudi AC, McManus BM, Abraham T, Francis GA. Contribution of intimal smooth muscle cells to cholesterol accumulation and macrophage-like cells in human atherosclerosis. *Circulation*. 2014;129(15):1551-1559. doi:10.1161/CIRCULATIONAHA.113.005015

41. Andreeva ER, Pugach IM, Orekhov AN. Subendothelial smooth muscle cells of human aorta express macrophage antigen in situ and in vitro. *Atherosclerosis*. 1997;135(1):19-27. doi:10.1016/s0021-9150(97)00136-6

42. Rong JX, Shapiro M, Trogan E, Fisher EA. Transdifferentiation of mouse aortic smooth muscle cells to a macrophage-like state after cholesterol loading. *Proc Natl Acad Sci U S A*. 2003;100(23):13531-13536. doi:10.1073/pnas.1735526100

43. Frontini MJ, O'Neil C, Sawyez C, Chan BMC, Huff MW, Pickering JG. Lipid incorporation inhibits Src-dependent assembly of fibronectin and type I collagen by vascular smooth muscle cells. *Circ Res*. 2009;104(7):832-841. doi:10.1161/CIRCRESAHA.108.187302

44. Shah PK. Biomarkers of plaque instability. *Curr Cardiol Rep*. 2014;16(12):547. doi:10.1007/s11886-014-0547-7

45. Libby P. Molecular and cellular mechanisms of the thrombotic complications of atherosclerosis. *J Lipid Res*. 2009;50 Suppl:S352-357. doi:10.1194/jlr.R800099-JLR200

46. Marenberg ME, Risch N, Berkman LF, Floderus B, de Faire U. Genetic susceptibility to death from coronary heart disease in a study of twins. *N Engl J Med*. 1994;330(15):1041-1046. doi:10.1056/NEJM199404143301503

47. Lloyd-Jones DM, Nam BH, D'Agostino S, Levy D, Murabito JM, Wang TJ. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: A prospective study of parents and offspring. *JAMA*. 2004;291:2204-2211.

48. Bachmann JM, Willis BL, Ayers CR, Khera A, Berry JD. Association between family history and coronary heart disease death across long-term follow-up in men: The Cooper Center Longitudinal Study. *Circulation*. 2012;125(25):3092-3098. doi:10.1161/CIRCULATIONAHA.111.065490

49. Holmes MV, Harrison S, Talmud PJ, Hingorani AD, Humphries SE. Utility of genetic determinants of lipids and cardiovascular events in assessing risk. *Nat Rev Cardiol*. 2011;8(4):207-221. doi:10.1038/nrcardio.2011.6

50. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science*. 2008;322(5903):881-888. doi:10.1126/science.1156409

51. Bronte-Stewart B, Botha MC, Krut LH. ABO blood groups in relation to ischaemic heart disease. *Br Med J*. 1962;1(5293):1646-1650. doi:10.1136/bmj.1.5293.1646

52. Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet*. 2014;46(11):1173-1186. doi:10.1038/ng.3097

53. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491(7422):119-124.

54. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421-427. doi:10.1038/nature13595

55. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. *PLoS Comput Biol*. 2012;8(12):e1002822. doi:10.1371/journal.pcbi.1002822

56. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76-82. doi:10.1016/j.ajhg.2010.11.011

57. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*. 2010;42(7):565-569. doi:10.1038/ng.608

58. Manolio TA. Bringing genome-wide association findings into clinical use. *Nat Rev Genet*. 2013;14(8):549-558. doi:10.1038/nrg3523

59. CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45(1):25-33. doi:10.1038/ng.2480

60. Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet*. 2011;43(4):339-344. doi:10.1038/ng.782

61. Nikpay M, Goel A, Won HH, al et. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary arterydisease. *Nat Genet*. 2015;47(10):1121-1130.

62. Nelson CP, Goel A, Butterworth AS, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet*. 2017;49(9):1385-1391. doi:10.1038/ng.3913

63. Liu X, Li YI, Pritchard JK. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell*. 2019;177(4):1022-1034.e6. doi:10.1016/j.cell.2019.04.014

64. Broman KW. Review of statistical methods for QTL mapping in experimental crosses. *Lab Anim*. 2001;30(7):44-52.

65. Doerge RW. Mapping and analysis of quantitative trait loci in experimental populations. *Nat Rev Genet*. 2002;3(1):43-52. doi:10.1038/nrg703

66. Abiola O, Angel JM, Avner P, et al. The nature and identification of quantitative trait loci: A community's view. *Nat Rev Genet*. 2003;4(11):911-916. doi:10.1038/nrg1206

67. Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. *Bioinforma Oxf Engl*. 2003;19(7):889-890. doi:10.1093/bioinformatics/btg112

68. Smith R, Sheppard K, DiPetrillo K, Churchill G. Quantitative trait locus analysis using J/qtl. *Methods Mol Biol Clifton NJ*. 2009;573:175-188. doi:10.1007/978-1-60761-247-6_10

69. Meir KS, Leitersdorf E. Atherosclerosis in the apolipoprotein-E-deficient mouse: A decade of progress. *Arter Thromb Vasc Biol*. 2004;24(6):1006-1014.

70. Rollins J, Chen Y, Paigen B, Wang X. In Search of New Targets for Plasma High-Density Lipoprotein Cholesterol Levels: Promise of Human–Mouse Comparative Genomics. *Trends Cardiovasc Med*. 2006;16(7):220-234.

71. Nakashima Y, Plump AS, Raines EW, Breslow JL, Ross R. ApoE-deficient mice develop lesions of all phases of atherosclerosis throughout the arterial tree. *Arter Thromb*. 1994;14(1):133-140.

72. Ishibashi S, Goldstein JL, Brown MS, Herz J, Burns DK. Massive xanthomatosis and atherosclerosis in cholesterol-fed low density lipoprotein receptor-negative mice. *J Clin Invest*. 1994;93(5):1885-1893. doi:10.1172/JCI117179

73. Dansky HM, Shu P, Donavan M, et al. A phenotype-sensitizing Apoe-deficient genetic background reveals novel atherosclerosis predisposition loci in the mouse. *Genetics*. 2002;160(4):1599-1608.

74. Kayashima Y, Tomita H, Zhilicheva S, et al. Quantitative trait loci affecting atherosclerosis at the aortic root identified in an intercross between DBA2J and 129S6 apolipoprotein E-null mice. *PLoS One*. 2014;9(2):e88274.

75. Wang SS, Shi W, Wang X, et al. Mapping, genetic isolation, and characterization of genetic loci that determine resistance to atherosclerosis in C3H mice. *Arter Thromb Vasc Biol*. 2007;27(12):27671-2676.

76. Zhang Z, Rowlan JS, Wang Q, Shi W. 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):190-201. doi:10.1161/CIRCGENETICS.111.961649

77. Burkhardt R, Sündermann S, Ludwig D, et al. Cosegregation of aortic root atherosclerosis and intermediate lipid phenotypes on chromosomes 2 and 8 in an intercross of C57BL/6 and BALBc/ByJ low-density lipoprotein receptor-/- mice. *Arter Thromb Vasc Biol*. 2011;31(4):775-784.

78. Seidelmann SB, Luca CD, Leibel RL, Breslow JL, Tall AR, Welch CL. Quantitative trait locus mapping of genetic modifiers of metabolic syndrome and atherosclerosis in low-density lipoprotein receptor-deficient mice: Identification of a locus for metabolic syndrome and increased atherosclerosis on chromosome 4. *Arter Thromb Vasc Biol*. 2005;25(1):204-210.

79. Rowlan JS, Li Q, Manichaikul A, Wang Q, Matsumoto AH, Shi W. Atherosclerosis susceptibility Loci identified in an extremely atherosclerosis-resistant mouse strain. *J Am Heart Assoc*. 2013;2(4):e000260.

80. Smith JD, Bhasin JM, Baglione J, Settle M, Xu Y, Barnard J. Atherosclerosis susceptibility loci identified from a strain intercross of apolipoprotein E-deficient mice via a high-density genome scan. *Arter Thromb Vasc Biol*. 2006;26(3):597-603.

81. Teupser D, Tan M, Persky AD, Breslow JL. Atherosclerosis quantitative trait loci are sex- and lineage-dependent in an intercross of C57BL/6 and FVB/N low-density lipoprotein receptor-/- mice. *Proc Natl Acad Sci U A*. 2006;103(1):123-128.

82. Korstanje R, Li R, Howard T, et al. Influence of sex and diet on quantitative trait loci for HDL cholesterol levels in an SM/J by NZB/BlNJ intercross population. *J Lipid Res*. 2004;45(5):881-888.

83. Wang SS, Schadt EE, Wang H, et al. Identification of pathways for atherosclerosis in mice: Integration of quantitative trait locus analysis andglobal gene expression data. *Circ Res*. 2007;101(3):11-30.

84. Welch CL, Bretschger S, Latib N, et al. Localization of atherosclerosis susceptibility loci to chromosomes 4 and 6 using the Ldlr knockout mousemodel. *Proc Natl Acad Sci U A*. 2001;98(14):7946-7951.

85. Ghazalpour A, Wang X, Lusis AJ, Mehrabian M. Complex inheritance of the 5-lipoxygenase locus influencing atherosclerosis in mice. *Genetics*. 2006;173(2):943-951.

86. Mehrabian M, Wong J, Wang X, et al. Genetic locus in mice that blocks development of atherosclerosis despite extreme hyperlipidemia. *Circ Res*. 2001;89:125-130.

87. Colinayo VV, Qiao JH, Wang X, et al. Genetic loci for diet-induced atherosclerotic lesions and plasma lipids in mice. *Mamm Genome*. 2003;14(7):464-471.

88. Su Z, Li Y, James JC, et al. Quantitative trait locus analysis of atherosclerosis in an intercross between C57BL/6 and C3H mice carrying the mutant apolipoprotein E gene. *Genetics*. 2006;172(3):1799-1807.

89. Ishimori N, Li R, Kelmenson PM, et al. Quantitative trait loci analysis for plasma HDL-cholesterol concentrations and atherosclerosis susceptibility between inbred mouse strains C57BL/6J and 129S1/SvImJ. *Arter Thromb Vasc Biol*. 2004;24(1):161-166.

90. Mu JL, Naggert JK, Svenson KL, et al. Quantitative trait loci analysis for the differences in susceptibility to atherosclerosis and diabetes between inbred mouse strains C57BL/6J and C57BLKS/J. *J Lipid Res*. 1999;40(7):1328-1335.

91. Bennett BJ, Davis RC, Civelek M, et al. Genetic Architecture of Atherosclerosis in Mice: A Systems Genetics Analysis of Common Inbred Strains. *PLoS Genet*. 2015;11(12).

92. Wang X, Paigen B. Genetics of variation in HDL cholesterol in humans and mice. *Circ Res*. 2005;96(1):27-42.

93. Smith JD, James D, Dansky HM, Wittkowski KM, Moore KJ, Breslow JL. In silico quantitative trait locus map for atherosclerosis susceptibility in apolipoprotein E-deficient mice. *Arter Thromb Vasc Biol*. 2003;23(1):117-122.

94. Burgess-Herbert SL, Cox A, Tsaih SW, Paigen B. Practical applications of the bioinformatics toolbox for narrowing quantitative trait loci. *Genetics*. 2008;180(4):2227-2235.

95. Manichaikul A, Wang Q, Shi YL, Zhang Z, Leitinger N, Shi W. Characterization of Ath29, a major mouse atherosclerosis susceptibility locus, and identification of Rcn2as a novel regulator of cytokine expression. *Am J Physiol Heart Circ Physiol*. 2011;301(3):1056-1061.

96. Leduc MS, Hageman RS, Verdugo RA, et al. Integration of QTL and bioinformatic tools to identify candidate genes for triglycerides in mice. *J Lipid Res*. 2011;52(9):1672-1682.

97. Farber CR. Systems Genetics: A Novel Approach to Dissect the Genetic Basis of Osteoporosis. *Curr Osteoporos Rep*. 2012;10(3):228-235. doi:10.1007/s11914-012-0112-5

98. Farber CR, van Nas A, Ghazalpour A, et al. An Integrative Genetics Approach to Identify Candidate Genes Regulating BMD: Combining Linkage, Gene Expression, and Association. *J Bone Miner Res*. 2009;24(1):105-116. doi:10.1359/JBMR.080908

99. Calabrese GM, Mesner LD, Stains JP, et al. Integrating GWAS and Co-expression Network Data Identifies Causal Bone Mineral Density Genes SPTBN1 and MARK3 and an Osteoblast Functional Module. *Cell Syst*. 2017;4(1):46-59.e4. doi:10.1016/j.cels.2016.10.014

100. Li N, Fu J, Koonen DP, Kuivenhoven JA, Snieder H, Hofker MH. Are hypertriglyceridemia and low HDL causal factors in the development of insulin resistance? *Atherosclerosis*. 2014;233(1):130-138.

101. Saleheen D, Nazir A, Khanum S, Haider SR, Frossard PM. R1615P: A novel mutation in ABCA1 associated with low levels of HDL and type II diabetes mellitus. *Int J Cardiol*. 2006;110(2):259-260. doi:10.1016/j.ijcard.2005.06.059

102. Albert JS, Yerges-Armstrong LM, Horenstein RB, et al. Null mutation in hormone-sensitive lipase gene and risk of type 2 diabetes. *N Engl J Med*. 2014;370(24):2307-2315. doi:10.1056/NEJMoa1315496

103. Hu Y, Ren Y, Luo RZ, et al. Novel mutations of the lipoprotein lipase gene associated with hypertriglyceridemia in members of type 2 diabetic pedigrees. *J Lipid Res*. 2007;48(8):1681-1688. doi:10.1194/jlr.M600382-JLR200

104. Mani A, Radhakrishnan J, Wang H, et al. LRP6 mutation in a family with early coronary disease and metabolic risk factors. *Science*. 2007;315(5816):1278-1282. doi:10.1126/science.1136370

105. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, al et. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466(7307):707-713.

106. Consortium GLG. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45(11):1274-1283.

107. Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet*. 2010;42(2):105-116. doi:10.1038/ng.520

108. Soranzo N, Sanna S, Wheeler E, et al. Common variants at 10 genomic loci influence hemoglobin A_1(C) levels via glycemic and nonglycemic pathways. *Diabetes*. 2010;59(12):3229-3239. doi:10.2337/db10-0502

109. Manning AK, Hivert M-F, Scott RA, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet*. 2012;44(6):659-669. doi:10.1038/ng.2274

110. Li N, van der Sijde MR, LifeLines Cohort Study Group, et al. Pleiotropic effects of lipid genes on plasma glucose, HbA1c, and HOMA-IR levels. *Diabetes*. 2014;63(9):3149-3158. doi:10.2337/db13-1800

111. Shi W, Wang NJ, Shih DM, Sun VZ, Wang X, Lusis AJ. Determinants of atherosclerosis susceptibility in the C3H and C57BL/6 mouse model: Evidence for involvement of endothelial cells but not blood cells or cholesterol metabolism. *Circ Res*. 2000;86:1078-1084.

112. Tian J, Pei H, James JC, et al. Circulating adhesion molecules in apoE-deficient mouse strains with different atherosclerosis susceptibility. *Biochem Biophys Res Commun*. 2005;329(3):1102-1107.

113. Shi W, Zhang Z, Chen MH, Angle JF, Matsumoto AH. Genes within the MHC region have a dramatic influence on radiation-enhanced atherosclerosis in mice. *Circ Cardiovasc Genet*. 2010;3(5):409-413.

114. Liu S, Li J, Chen MH, Liu Z, Shi W. Variation in Type 2 Diabetes-Related Phenotypes among Apolipoprotein E-Deficient Mouse Strains. *PLoS One*. 2015;10(5).

115. Zhang Y, Kundu B, Zhong M, et al. PET imaging detection of macrophages with a formyl peptide receptor antagonist. *Nucl Med Biol*. 2015;42(4):381-386. doi:10.1016/j.nucmedbio.2014.12.001

116. Breslow JL. Genetic differences in endothelial cells may determine atherosclerosis susceptibility. *Circulation*. 2000;102(1):5-6. doi:10.1161/01.cir.102.1.5

117. Su Z, Li Y, James JC, et al. Genetic linkage of hyperglycemia, body weight and serum amyloid-P in an intercross between C57BL/6 and C3H apolipoprotein E-deficient mice. *Hum Mol Genet*. 2006;15(10):1650-1658.

118. Li J, Wang Q, Chai W, Chen MH, Liu Z, Shi W. Hyperglycemia in apolipoprotein E-deficient mouse strains with different atherosclerosis susceptibility. *Cardiovasc Diabetol*. 2011;10(117).

119. Lyons MA, Wittenburg H, Li R, et al. Quantitative trait loci that determine lipoprotein cholesterol levels in an intercross of 129S1/SvImJ and CAST/Ei inbred mice. *Physiol Genomics*. 2004;17(1):60-68. doi:10.1152/physiolgenomics.00142.2003

120. Wittenburg H, Lyons MA, Li R, et al. QTL mapping for genetic determinants of lipoprotein cholesterol levels in combined crosses of inbred mouse strains. *J Lipid Res*. 2006;47(8):1780-1790. doi:10.1194/jlr.M500544-JLR200

121. Sehayek E, Duncan EM, Yu HJ, Petukhova L, Breslow JL. Loci controlling plasma non-HDL and HDL cholesterol levels in a C57BL /6J x CASA /Rk intercross. *J Lipid Res*. 2003;44(9):1744-1750. doi:10.1194/jlr.M300139-JLR200

122. Li Q, Li Y, Zhang Z, et al. Quantitative trait locus analysis of carotid atherosclerosis in an intercross between C57BL/6 and C3H apolipoprotein E-deficient mice. *Stroke*. 2008;39(1):166-173.

123. Ishimori N, Li R, Kelmenson PM, et al. Quantitative trait loci that determine plasma lipids and obesity in C57BL/6J and 129S1/SvImJ inbred mice. *J Lipid Res*. 2004;45(9):1624-1632. doi:10.1194/jlr.M400098-JLR200

124. Stylianou IM, Langley SR, Walsh K, Chen Y, Revenu C, Paigen B. Differences in DBA/1J and DBA/2J reveal lipid QTL genes. *J Lipid Res*. 2008;49(11):2402-2413. doi:10.1194/jlr.M800244-JLR200

125. Yaguchi H, Togawa K, Moritani M, Itakura M. Identification of candidate genes in the type 2 diabetes modifier locus using expression QTL. *Genomics*. 2005;85(5):591-599. doi:10.1016/j.ygeno.2005.01.006

126. Lyons MA, Korstanje R, Li R, et al. Genetic contributors to lipoprotein cholesterol levels in an intercross of 129S1/SvImJ and RIIIS/J inbred mice. *Physiol Genomics*. 2004;17(2):114-121. doi:10.1152/physiolgenomics.00168.2003

127. Su Z, Wang X, Tsaih S-W, et al. Genetic basis of HDL variation in 129/SvImJ and C57BL/6J mice: Importance of testing candidate genes in targeted mutant mice. *J Lipid Res*. 2009;50(1):116-125. doi:10.1194/jlr.M800411-JLR200

128. Rowlan JS, Zhang Z, Wang Q, Fang Y, Shi W. New quantitative trait loci for carotid atherosclerosis identified in an intercross derived from apolipoprotein E-deficient mouse strains. *Physiol Genomics*. 2013;(45):332.

129. Ko A, Cantor RM, Weissglas-Volkov D, et al. Amerindian-specific regions under positive selection harbour new lipid variants in Latinos. *Nat Commun*. 2014;5:3983. doi:10.1038/ncomms4983

130. Comuzzie AG, Cole SA, Laston SL, et al. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLoS One*. 2012;7(12).

131. Purcell-Huynh DA, Weinreb A, Castellani LW, Mehrabian M, Doolittle MH, Lusis AJ. Genetic factors in lipoprotein metabolism. Analysis of a genetic cross between inbred mouse strains NZB/BINJ and SM/J using a complete linkage map approach. *J Clin Invest*. 1995;96(4):1845-1858. doi:10.1172/JCI118230

132. Willer CJ, Mohlke KL. Finding genes and variants for lipid levels after genome-wide association analysis. *Curr Opin Lipidol*. 2012;23(2):98-103. doi:10.1097/MOL.0b013e328350fad2

133. Li J, Lu Z, Wang Q, Su Z, Bao Y, Shi W. Characterization of Bglu3, a mouse fasting glucose locus, and identification of Apcs as an underlying candidate gene. *Physiol Genomics*. 2012;44(6):345-351. doi:10.1152/physiolgenomics.00087.2011

134. Zhou W, Chen M-H, Shi W. Influence of phthalates on glucose homeostasis and atherosclerosis in hyperlipidemic mice. *BMC Endocr Disord*. 2015;15:13. doi:10.1186/s12902-015-0015-4

135. Wilson PW, Kannel WB, Anderson KM. Lipids, glucose intolerance and vascular disease: The Framingham Study. *Monogr Atheroscler*. 1985;13:1-11.

136. Wilson PWF, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB. Prediction of incident diabetes mellitus in middle-aged adults: The Framingham Offspring Study. *Arch Intern Med*. 2007;167(10):1068-1074. doi:10.1001/archinte.167.10.1068

137. Drew BG, Rye K-A, Duffy SJ, Barter P, Kingwell BA. The emerging role of HDL in glucose metabolism. *Nat Rev Endocrinol*. 2012;8(4):237-245. doi:10.1038/nrendo.2011.235

138. Hwang Y-C, Ahn H-Y, Yu S-H, Park S-W, Park C-Y. Atherogenic dyslipidaemic profiles associated with the development of Type 2 diabetes: A 3.1-year longitudinal study. *Diabet Med J Br Diabet Assoc*. 2014;31(1):24-30. doi:10.1111/dme.12278

139. Hwang Y-C, Ahn H-Y, Park S-W, Park C-Y. Apolipoprotein B and non-HDL cholesterol are more powerful predictors for incident type 2 diabetes than fasting glucose or glycated hemoglobin in subjects with normal glucose tolerance: A 3.3-year retrospective longitudinal study. *Acta Diabetol*. 2014;51(6):941-946. doi:10.1007/s00592-014-0587-x

140. Ley SH, Harris SB, Connelly PW, et al. Association of apolipoprotein B with incident type 2 diabetes in an aboriginal Canadian population. *Clin Chem*. 2010;56(4):666-670. doi:10.1373/clinchem.2009.136994

141. Rütti S, Ehses JA, Sibler RA, et al. Low- and high-density lipoproteins modulate function, apoptosis, and proliferation of primary human and murine pancreatic beta-cells. *Endocrinology*. 2009;150(10):4521-4530. doi:10.1210/en.2009-0252

142. Visscher PM, Thompson R, Haley CS. Confidence intervals in QTL mapping by bootstrapping. *Genetics*. 1996;143(2):1013-1020.

143. Shepherd M, Ellis I, Ahmad AM, et al. Predictive genetic testing in maturity-onset diabetes of the young (MODY). *Diabet Med J Br Diabet Assoc*. 2001;18(5):417-421.

144. Wang X, Korstanje R, Higgins D, Paigen B. Haplotype analysis in multiple crosses to identify a QTL gene. *Genome Res*. 2004;14(9):1767-1772. doi:10.1101/gr.2668204

145. Yuan Z, Pei H, Roberts DJ, et al. Quantitative trait locus analysis of neointimal formation in an intercross between C57BL/6 and C3H/HeJ apolipoprotein E-deficient mice. *Circ Cardiovasc Genet*. 2009;2(3):220-228.

146. Mozaffarian D, Benjamin EJ, Go AS, et al. Heart Disease and Stroke Statistics—2015 Update: A Report From the American Heart Association. *Circulation*. 2015;131(4):29-322.

147. Donnan GA, Fisher M, Macleod M, Davis SM. Stroke. *Lancet*. 2008;371(9624):1612-1623.

148. Markus H, Cullinane M. Severely impaired cerebrovascular reactivity predicts stroke and TIA risk in patients with carotid artery stenosis and occlusion. *Brain*. 2001;124(Pt 3):457-467.

149. Matteis M, Vernieri F, Caltagirone C, Troisi E, Rossini PM, Silvestrini M. Patterns of cerebrovascular reactivity in patients with carotid artery occlusion and severe contralateralstenosis. *J Neurol Sci*. 1999;168(1):47-51.

150. Sacco RL, Blanton SH, Slifer S, et al. Heritability and linkage analysis for carotid intima-media thickness: The family study of stroke risk andcarotid atherosclerosis. *Stroke*. 2009;40(7):2307-2312.

151. Swan L, Birnie DH, Inglis G, Connell JM, Hillis WS. The determination of carotid intima medial thickness in adults–a population-based twin study. *Atherosclerosis*. 2003;166(1):137-141.

152. Zhao J, Cheema FA, Bremner JD, et al. Heritability of carotid intima-media thickness: A twin study. *Atherosclerosis*. 2008;197(2):814-820.

153. Bis JC, Kavousi M, Franceschini N, et al. Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. *Nat Genet*. 2011;43(10):940-947.

154. Du R, Zhou J, Lorenzano S, et al. Integrative Mouse and Human Studies Implicate ANGPT1 and ZBTB7C as Susceptibility Genes to Ischemic Injury. *Stroke*. 2015;46(12):3514-3522.

155. Xie G, Myint PK, Voora D, et al. Genome-wide association study on progression of carotid artery intima media thickness over 10 years in a Chinese cohort. *Atherosclerosis*. 2015;243(1):30-37. doi:10.1016/j.atherosclerosis.2015.08.034

156. Li R, Lyons MA, Wittenburg H, Paigen B, Churchill GA. Combining data from multiple inbred line crosses improves the power and resolution of quantitative trait loci mapping. *Genetics*. 2005;169(3):1699-1709.

157. Wang X, Ria M, Kelmenson PM, et al. Positional identification of TNFSF4, encoding OX40 ligand, as a gene that influences atherosclerosis susceptibility. *Nat Genet*. 2005;37(4):365-372. doi:10.1038/ng1524

158. Grainger AT, Jones MB, Li J, Chen M-H, Manichaikul A, Shi W. Genetic analysis of atherosclerosis identifies a major susceptibility locus in the major histocompatibility complex of mice. *Atherosclerosis*. 2016;254:124–132. doi:10.1016/j.atherosclerosis.2016.10.011

159. Wang Q, Grainger AT, Manichaikul A, Farber E, Onengut-Gumuscu S, Shi W. Genetic linkage of hyperglycemia and dyslipidemia in an intercross between BALB/cJ and SM/J Apoe-deficient mouse strains. *BMC Genet*. 2015;16:133. doi:10.1186/s12863-015-0292-y

160. Tomita H, Zhilicheva S, Kim S, Maeda N. Aortic arch curvature and atherosclerosis have overlapping quantitative trait loci in a cross between 129S6/SvEvTac and C57BL/6J apolipoprotein E-null mice. *Circ Res*. 2010;106(6):1052-1060.

161. Bennett BJ, Wang SS, Wang X, Wu X, Lusis AJ. Genetic regulation of atherosclerotic plaque size and morphology in the innominate artery of hyperlipidemic mice. *Arterioscler Thromb Vasc Biol*. 2009;29(3):348-355. doi:10.1161/ATVBAHA.108.176685

162. Meschia JF, Worrall BB, Rich SS. Genetic susceptibility to ischemic stroke. *Nat Rev Neurol*. 2011;7(7):369-378. doi:10.1038/nrneurol.2011.80

163. Frazier-Wood AC, Manichaikul A, Aslibekyan S, et al. Genetic variants associated with VLDL, LDL and HDL particle size differ with race/ethnicity. *Hum Genet*. 2013;132(4):405-413. doi:10.1007/s00439-012-1256-1

164. Mora S, Szklo M, Otvos JD, et al. LDL particle subclasses, LDL particle size, and carotid atherosclerosis in the Multi-Ethnic Study of Atherosclerosis (MESA). *Atherosclerosis*. 2007;192:211-217.

165. Sanossian N, Saver JL, Navab M, Ovbiagele B. High-density lipoprotein cholesterol: An emerging target for stroke treatment. *Stroke*. 2007;38(1104):1109.

166. De Michele M, Panico S, Celentano E, et al. Association of impaired glucose homeostasis with preclinical carotid atherosclerosis in women: Impact of the new American Diabetes Association criteria. *Metabolism*. 2002;51(1):52-56. doi:10.1053/meta.2002.29025

167. Wergedal JE, Ackert-Bicknell CL, Beamer WG, Mohan S, Baylink DJ, Srivastava AK. Mapping genetic loci that regulate lipid levels in a NZB/B1NJxRF/J intercross and a combined intercross involving NZB/B1NJ, RF/J, MRL/MpJ, and SJL/J mouse strains. *J Lipid Res*. 2007;48(8):1724-1734.

168. Wade CM, Kulbokas CJ, Kirby AW, et al. The mosaic structure of variation in the laboratory mouse genome. *Nature*. 2002:574-578.

169. Peters LL, Robledo RF, Bult CJ, Churchill GA, Paigen BJ, Svenson KL. The mouse as a model for human biology: A resource guide for complex trait analysis. *Nat Rev Genet*. 2007;8:58-69.

170. Banerjee SS, Thirunavukkarasu M, Rishi MT, Sanchez J, Maulik N, Maulik G. HIF-prolyl hydroxylases and cardiovascular diseases. *Toxicol Mech Methods*. 2012;22(5):347-358.

171. Holmes MV, Asselbergs FW, Palmer TM, et al. Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J*. 2015;36(9):539-550. doi:10.1093/eurheartj/eht571

172. White J, Sofat R, Hemani G, et al. Plasma urate concentration and risk of coronary heart disease: A Mendelian randomisation analysis. *Lancet Diabetes Endocrinol*. 2016;4(4):327-336. doi:10.1016/S2213-8587(15)00386-1

173. Shi W, Wang Q, Choi W, Li J. Mapping and Congenic Dissection of Genetic Loci Contributing to Hyperglycemia and Dyslipidemia in Mice. *PloS One*. 2016;11(2):e0148462. doi:10.1371/journal.pone.0148462

174. Grainger AT, Jones MB, Li J, Chen M-H, Manichaikul A, Shi W. Genetic analysis of atherosclerosis identifies a major susceptibility locus in the major histocompatibility complex of mice. *Atherosclerosis*. 2016;254:124–132. doi:10.1016/j.atherosclerosis.2016.10.011

175. Arnold P, Otte A, Becker-Pauly C. Meprin metalloproteases: Molecular regulation and function in inflammation and fibrosis. *Biochim Biophys Acta Mol Cell Res*. 2017;1864(11 Pt B):2096-2104. doi:10.1016/j.bbamcr.2017.05.011

176. Herzog C, Haun RS, Kaushal GP. Role of meprin metalloproteinases in cytokine processing and inflammation. *Cytokine*. 2019;114:18-25. doi:10.1016/j.cyto.2018.11.032

177. Broder C, Becker-Pauly C. The metalloproteases meprin α and meprin β: Unique enzymes in inflammation, neurodegeneration, cancer and fibrosis. *Biochem J*. 2013;450(2):253-264. doi:10.1042/BJ20121751

178. Broder C, Arnold P, Vadon-Le Goff S, et al. Metalloproteases meprin α and meprin β are C- and N-procollagen proteinases important for collagen assembly and tensile strength. *Proc Natl Acad Sci U S A*. 2013;110(35):14219-14224. doi:10.1073/pnas.1305464110

179. Herzog C, Haun RS, Kaushal V, Mayeux PR, Shah SV, Kaushal GP. Meprin A and meprin alpha generate biologically functional IL-1beta from pro-IL-1beta. *Biochem Biophys Res Commun*. 2009;379(4):904-908. doi:10.1016/j.bbrc.2008.12.161

180. Herzog C, Haun RS, Shah SV, Kaushal GP. Proteolytic processing and inactivation of CCL2/MCP-1 by meprins. *Biochem Biophys Rep*. 2016;8:146-150. doi:10.1016/j.bbrep.2016.08.019

181. Arnold P, Boll I, Rothaug M, et al. Meprin Metalloproteases Generate Biologically Active Soluble Interleukin-6 Receptor to Induce Trans-Signaling. *Sci Rep*. 2017;7:44053. doi:10.1038/srep44053

182. Biasin V, Wygrecka M, Bärnthaler T, et al. Docking of Meprin α to Heparan Sulphate Protects the Endothelium from Inflammatory Cell Extravasation. *Thromb Haemost*. 2018;118(10):1790-1802. doi:10.1055/s-0038-1670657

183. Gao P, Guo R, Chen J, et al. A meprin inhibitor suppresses atherosclerotic plaque formation in ApoE-/- mice. *Atherosclerosis*. 2009;207(1):84-92. doi:10.1016/j.atherosclerosis.2009.04.036

184. Jefferson T, Auf dem Keller U, Bellac C, et al. The substrate degradome of meprin metalloproteases reveals an unexpected proteolytic link between meprin β and ADAM10. *Cell Mol Life Sci CMLS*. 2013;70(2):309-333. doi:10.1007/s00018-012-1106-2

185. Jian D, Wang W, Zhou X, et al. Interferon-induced protein 35 inhibits endothelial cell proliferation, migration and re-endothelialization of injured arteries by inhibiting the nuclear factor-kappa B pathway. *Acta Physiol Oxf Engl*. 2018;223(3):e13037. doi:10.1111/apha.13037

186. Derda AA, Woo CC, Wongsurawat T, et al. Gene expression profile analysis of aortic vascular smooth muscle cells reveals upregulation of cadherin genes in myocardial infarction patients. *Physiol Genomics*. 2018;50(8):648-657. doi:10.1152/physiolgenomics.00042.2017

187. Imanaka-Yoshida K. Tenascin-C in cardiovascular tissue remodeling: From development to inflammation and repair. *Circ J Off J Jpn Circ Soc*. 2012;76(11):2513-2520.

188. Angsana J, Chen J, Smith S, et al. Syndecan-1 modulates the motility and resolution responses of macrophages. *Arterioscler Thromb Vasc Biol*. 2015;35(2):332-340. doi:10.1161/ATVBAHA.114.304720

189. Voyvodic PL, Min D, Liu R, et al. Loss of syndecan-1 induces a pro-inflammatory phenotype in endothelial cells with a dysregulated response to atheroprotective flow. *J Biol Chem*. 2014;289(14):9547-9559. doi:10.1074/jbc.M113.541573

190. Baeyens N, Mulligan-Kehoe MJ, Corti F, et al. Syndecan 4 is required for endothelial alignment in flow and atheroprotective signaling. *Proc Natl Acad Sci U S A*. 2014;111(48):17308-17313. doi:10.1073/pnas.1413725111

191. Borland SJ, Morris TG, Borland SC, et al. Regulation of vascular smooth muscle cell calcification by syndecan-4/FGF-2/PKCα signalling and cross-talk with TGFβ. *Cardiovasc Res*. 2017;113(13):1639-1652. doi:10.1093/cvr/cvx178

192. Shami A, Gonçalves I, ardh-Nilsson AH. Collagen and related extracellular matrix proteins in atherosclerotic plaque development. *Curr Opin Lipidol*. 2014;25(5):394-399.

193. Nadkarni SK, Bouma BE, de Boer J, Tearney GJ. Evaluation of collagen in atherosclerotic plaques: The use of two coherent laser-based imaging methods. *Lasers Med Sci*. 2009;24(3):439-445. doi:10.1007/s10103-007-0535-x

194. Ungvari Z, Valcarcel-Ares MN, Tarantini S, et al. Connective tissue growth factor (CTGF) in age-related vascular pathologies. *GeroScience*. 2017;39(5-6):491-498. doi:10.1007/s11357-017-9995-5

195. Lin J, Kakkar V, Lu X. Impact of matrix metalloproteinases on atherosclerosis. *Curr Drug Targets*. 2014;15(4):442-453.

196. Oksala N, Levula M, Airla N, et al. ADAM-9, ADAM-15, and ADAM-17 are upregulated in macrophages in advanced human atherosclerotic plaques in aorta and carotid and femoral arteriesâ€"Tampere vascular study. *Ann Med*. 2009;41(4):279-290. doi:10.1080/07853890802649738

197. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: An open-source platform for biological-image analysis. *Nat Methods*. 2012;9(7):676-682.

198. Tremmel M, Gerdtham UG, Nilsson PM, Saha S. Economic Burden of Obesity: A Systematic Literature Review. *Int J Env Res Public Health*. 2017;14:10.3390/ijerph14040435.

199. Ogden C, Carroll M. Prevalence of Overweight, Obesity, and Extreme Obesity Among Adults: United States, Trends 1960–1962 Through 2007–2008. *NCHS Data Brief*. 201:1-6.

200. Baum T, Cordes C, Dieckmeyer M, et al. MR-based assessment of body fat distribution and characteristics. *Eur J Radiol*. 2016;85(8):1512-1518. doi:10.1016/j.ejrad.2016.02.013

201. Schär M, Eggers H, Zwart NR, Chang Y, Bakhru A, Pipe JG. Dixon water-fat separation in PROPELLER MRI acquired with two interleaved echoes. *Magn Reson Med*. 2016;75(2):718-728. doi:10.1002/mrm.25656

202. McBee MP, Awan OA, Colucci AT, et al. Deep Learning in Radiology. *Acad Radiol*. 2018.

203. Avants BB, Tustison NJ, Wu J, Cook PA, Gee JC. An open source multivariate framework for n-tissue segmentation with evaluation on public data. *Neuroinformatics*. 2011;9:381-400.

204. Manjón JV, Coupé P, Martí-Bonmatí L, Collins DL, Robles M. Adaptive non-local means denoising of MR images with spatially varying noise levels. *J Magn Reson Imaging JMRI*. 2010;31(1):192-203. doi:10.1002/jmri.22003

205. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: Improved N3 bias correction. *IEEE Trans Med Imaging*. 2010;29(6):1310-1320. doi:10.1109/TMI.2010.2046908

206. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*. 2006;31:1116-1128.

207. Avants BB, Yushkevich P, Pluta J, et al. The optimal template effect in hippocampus studies of diseased populations. *NeuroImage*. 2010;49(3):2457-2466. doi:10.1016/j.neuroimage.2009.09.062

208. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60-88.

209. Norman B, Pedoia V, Majumdar S. Use of 2D U-Net Convolutional Neural Networks for Automated Cartilage and Meniscus Segmentation of Knee MR Imaging Data to Determine Relaxometry and Morphometry. *Radiology*. 2018;288(1):177-185. doi:10.1148/radiol.2018172322

210. Blanc-Durand P, Van Der Gucht A, Schaefer N, Itti E, Prior JO. Automatic lesion detection and segmentation of 18F-FET PET in gliomas: A full 3D U-Net convolutional neural network study. *PloS One*. 2018;13(4):e0195798. doi:10.1371/journal.pone.0195798

211. Li J, Sarma KV, Chung Ho K, Gertych A, Knudsen BS, Arnold CW. A Multi-scale U-Net for Semantic Segmentation of Histological Images from Radical Prostatectomies. *AMIA Annu Symp Proc AMIA Symp*. 2017;2017:1140-1148.

212. Hales C, Carroll M, Fryar C, Ogden C. Prevalence of Obesity Among Adults and Youth: United States, 2015–2016. *NCHS Data Brief*. 2017;288.

213. Tustison NJ, Avants BB, Lin Z, et al. Convolutional Neural Networks with Template-Based Data Augmentation for Functional Lung Image Quantification. *Acad Radiol*. 2019;26(3):412-423. doi:10.1016/j.acra.2018.08.003

214. Grainger AT, Tustison NJ, Qing K, Roy R, Berr SS, Shi W. Deep learning-based quantification of abdominal fat on magnetic resonance images. *PLoS One*. 2018;13(9):e0204071.

215. Snell-Bergeon JK, Hokanson JE, Kinney GL, et al. Measurement of abdominal fat by CT compared to waist circumference and BMI in explaining the presence of coronary calcium. *Int J Obes Relat Metab Disord J Int Assoc Study Obes*. 2004;28(12):1594-1599. doi:10.1038/sj.ijo.0802796

216. Ryo M, Kishida K, Nakamura T, Yoshizumi T, Funahashi T, Shimomura I. Clinical significance of visceral adiposity assessed by computed tomography: A Japanese perspective. *World J Radiol*. 2014;6(7):409-416. doi:10.4329/wjr.v6.i7.409

217. Weston AD, Korfiatis P, Kline TL, et al. Automated Abdominal Segmentation of CT Scans for Body Composition Analysis Using Deep Learning. *Radiology*. 2019;290(3):669-679. doi:10.1148/radiol.2018181432

218. Commandeur F, Goeller M, Betancur J, et al. Deep Learning for Quantification of Epicardial and Thoracic Adipose Tissue From Non-Contrast CT. *IEEE Trans Med Imaging*. 2018;37(8):1835-1846. doi:10.1109/TMI.2018.2804799

219. Wang Y, Qiu Y, Thai T, Moore K, Liu H, Zheng B. A two-step convolutional neural network based computer-aided detection scheme for automatically segmenting adipose tissue volume depicting on CT images. *Comput Methods Programs Biomed*. 2017;144:97-104. doi:10.1016/j.cmpb.2017.03.017

220. Potretzke AM, Schmitz KH, Jensen MD. Preventing overestimation of pixels in computed tomography assessment of visceral fat. *Obes Res*. 2004;12(10):1698-1701. doi:10.1038/oby.2004.210

221. Kvist H, Chowdhury B, Grangård U, Tylén U, Sjöström L. Total and visceral adipose-tissue volumes derived from measurements with computed tomography in adult men and women: Predictive equations. *Am J Clin Nutr*. 1988;48(6):1351-1361. doi:10.1093/ajcn/48.6.1351

222. Lamba R, McGahan JP, Corwin MT, et al. CT Hounsfield numbers of soft tissues on unenhanced abdominal CT scans: Variability between two different manufacturers' MDCT scanners. *AJR Am J Roentgenol*. 2014;203(5):1013-1020. doi:10.2214/AJR.12.10037

223. Camhi SM, Bray GA, Bouchard C, et al. The relationship of waist circumference and BMI to visceral, subcutaneous, and total body fat: Sex and race differences. *Obes Silver Spring Md*. 2011;19(2):402-408. doi:10.1038/oby.2010.248

224. Nattenmueller J, Hoegenauer H, Boehm J, et al. CT-based compartmental quantification of adipose tissue versus body metrics in colorectal cancer patients. *Eur Radiol*. 2016;26(11):4131-4140. doi:10.1007/s00330-016-4231-8

225. Cho KW, Morris DL, Lumeng CN. Flow Cytometry Analyses of Adipose Tissue Macrophages. *Methods Enzymol*. 2014;537:297-314. doi:10.1016/B978-0-12-411619-1.00016-1

226. Chinetti-Gbaguidi G, Colin S, Staels B. Macrophage subsets in atherosclerosis. *Nat Rev Cardiol*. 2015;12(1):10-17. doi:10.1038/nrcardio.2014.173

227. Allen CL, Bayraktutan U. Risk factors for ischaemic stroke. *Int J Stroke*. 2008;3(2). http://onlinelibrary.wiley.com/doi/10.1111/j.1747-4949.2008.00187.x/epdf.

228. Leopold JA, Loscalzo J. Oxidative Risk for Atherothrombotic Cardiovascular Disease. *Free Radic Biol Med*. 2009;47(12):1673-1706. doi:10.1016/j.freeradbiomed.2009.09.009

229. Bessueille L, Magne D. Inflammation: A culprit for vascular calcification in atherosclerosis and diabetes. *Cell Mol Life Sci*. 2015;72(13):2475-2489.

230. Frostegård J. Immunity, atherosclerosis and cardiovascular disease. *BMC Med*. 2013;11:117. doi:10.1186/1741-7015-11-117

231. Chistiakov DA, Grechko AV, Myasoedova VA, Melnichenko AA, Orekhov AN. The role of monocytosis and neutrophilia in atherosclerosis. *J Cell Mol Med*. 2018;22(3):1366-1382. doi:10.1111/jcmm.13462

232. Chang Z, Huangfu C, Grainger AT, Zhang J, Guo Q, Shi W. Accelerated atherogenesis in completely ligated common carotid artery of apolipoprotein E-deficient mice. *Oncotarget*. 2017;8(66):110289–110299. doi:10.18632/oncotarget.22685

233. Stylianou IM, Bauer RC, Reilly MP, Rader DJ. Genetic basis of atherosclerosis: Insights from mice and humans. *Circ Res*. 2012;110(2):337-355.