8-2011

Gene By Bmi Interactions Influencing C-Reactive Protein Levels In European-Americans

Sarah Tudor

GENE BY BMI INTERACTIONS INFLUENCING C-REACTIVE PROTEIN LEVELS IN EUROPEAN-AMERICANS

by

Sarah Elizabeth Tudor

APPROVED:

_____
Supervisory Professor
Eric Boerwinkle, PhD

_____
Christie Ballantyne, MD

_____
Maja Barbalic, PhD

_____
Michael Hallman, PhD

_____
Paul Scheet, PhD

APPROVED:

_____
Dean, The University of Texas
Graduate School of Biomedical Sciences at Houston

GENE BY BMI INTERACTIONS INFLUENCING C-REACTIVE PROTEIN LEVELS IN EUROPEAN-AMERICANS

A

THESIS

Presented to the Faculty of
The University of Texas
Health Science Center at Houston
and
The University of Texas
M. D. Anderson Cancer Center
Graduate School of Biomedical Sciences
in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

By

Sarah Elizabeth Tudor, B.S.
Houston, TX

August, 2011

ACKNOWLEDGEMENTS

First and foremost I would like to thank Dr. Eric Boerwinkle and Dr. Maja Barbalic for giving me the opportunity to continue my education in such a supportive environment. Without all of their help and their willingness to take me on as a student for the last ten months I would not be where I am today. I would also like to thank the rest of my committee, Dr. Christie Ballantyne, Dr. Michael Hallman and Dr. Paul Scheet for their input and guidance with this project. Additionally, I would like to thank my fellow students, Jorge Del Auguila, Heather Highland, Jacy Crosby, and Bing Yu for statistical and analytical support throughout this learning process. To the past and present directors (Dr. Cote and Dr. Gambello) and students of the GSBS Human and Molecular Genetics Program, thank you for allowing me to participate in program activities and lectures for my two years here. And thank you especially to Scott, Abby, Danica, Melodie, Mary Anne, and Jennifer. I could not have made it through this process without your unconditional love and support; I will never forget all you've done for me.

GENE BY BMI INTERACTIONS INFLUENCING C-REACTIVE PROTEIN LEVELS IN EUROPEAN-AMERICANS

Publication No._____

Sarah Elizabeth Tudor, B.S.

Supervisory Professor: Eric Boerwinkle, PhD

C-Reactive Protein (CRP) is a biomarker indicating tissue damage, inflammation, and infection. High-sensitivity CRP (hsCRP) is an emerging biomarker often used to estimate an individual's risk for future coronary heart disease (CHD). hsCRP levels falling below 1.00 mg/l indicate a low risk for developing CHD, levels ranging between 1.00 mg/l and 3.00 mg/l indicate an elevated risk, and levels exceeding 3.00 mg/l indicate high risk. Multiple Genome-Wide Association Studies (GWAS) have identified a number of genetic polymorphisms which influence CRP levels. SNPs implicated in such studies have been found in or near genes of interest including: CRP, APOE, APOC, IL-6, HNF1A, LEPR, and GCKR. A strong positive correlation has also been found to exist between CRP levels and BMI, a known risk factor for CHD and a state of chronic inflammation. We conducted a series of analyses designed to identify loci which interact with BMI to influence CRP levels in a subsample of European-Americans in the ARIC cohort. In a stratified GWA analysis, 15 genetic regions were identified as having significantly (p-value < $2.00*10^{-3}$) distinct effects on hsCRP levels between the two obesity strata: lean (18.50 kg/m^2 < BMI < 24.99 kg/m^2) and obese (BMI ≥ 30.00 kg/m^2). A GWA analysis performed on all individuals combined (i.e. not *a priori* stratified for obesity status) with the inclusion of an additional parameter for BMI by gene interaction, identified 11 regions which interact with BMI to influence hsCRP levels. Two regions containing the genes GJA5 and GJA8 (on

chromosome 1) and FBXO11 (on chromosome 2) were identified in both methods of analysis suggesting that these genes possibly interact with BMI to influence hsCRP levels. We speculate that atrial fibrillation (AF), age-related cataracts and the TGF-β pathway may be the biological processes influenced by the interaction of GJA5, GJA8 and FBXO11, respectively, with BMI to cause changes in hsCRP levels. Future studies should focus on the influence of gene x bmi interaction on AF, age-related cataracts and TGF-β.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABREVIATIONS

ADIPOQ	adiponectin
AF	atrial fibrillation
ARIC	Atherosclerosis Risk in Communities Study
APOC1/E	apolipoprotein C1/E
BBS9	Bardet-Biedl syndrome 9
BDP1	transcription factor TFIIIB component B homolog
BMI	body mass index
CDH9	gene cadherin 9 type 2
C/EBP(β/δ)	cytidine-cytidine-adenosine-adenosine-thymidine-enhancer-binding-proteins
CHD	coronary heart disease
CRP	c-reactive protein
CSMD1	CUB and sushi domain-containing protein 1
ELAM	endothelial leukocyte adhesion molecule-1
EPHA8	ephrin type-A receptor 8
FAM75A6	family with sequence similarity 75, member A6

FBXO11	F-box protein 11
FOXN2	forkheadboxN2
FLT1	vascular endothelial growth factor receptor 1
GJA5/8	gap junction protein α5/8
GWAS	genome wide association study
HNF1A	hepatocyte nuclear factor 1 homeobox A
hsCRP	high-sensitivity c-reactive protein
IBD	identity by decent
IBS	identity by state
IFT172	intraflagellar transport 172 homolog
IL-1β/6	interleukin-1 β/6
ITGB1	integrin, β1
kb	kilobase
KLK5	Kallikrein-5
KCNAB1	voltage-gated potassium channel subunit beta-1 gene
LD	linkage disequilibrium
LEPR	leptin receptor

LINGO2	leucine rich repeat and Ig domain containing 2
MANEA	alpha-endomannosidase
MI	myocardial infarction
NRP1	neuropilin 1
NRXN3	neurexin 3
NTM	neurotrimin
OM	otitis media
OSTM1	osteopetrosis associated transmembrane protein 1
PCDH7	protocadherin 8
PDZRN4	PDZ domain containing ring finger 4
pSmad2	phospho Smad2
PTGR1	prostaglandin reductase 1
PVRL2	poliovirus receptor-related 2
Rel	proto-oncogene c-Rel
RFX6	regulator factor X 6
SLCO1B1	solute carrier organic anion transporter family member 1B1
SNP	single nucleotide polymorphism

STAT3	signal transducer and activator of transcription 3
ST3GAL4	CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase
TGF-β1/2	transforming growth factor- β 1/2
TNF-α	tumor necrosis factor alpha
TOMM40	translocase of outer mitochondrial membrane 40 homolog
VCAM-1	vascular cell adhesion molecule-1
VEGF	vascular endothelial growth factor receptor
ZNF	zinc finger nuclease

CHAPTER 1: BACKGROUND AND LITERATURE REVIEW

1.1 C-Reactive Protein

C-Reactive Protein Definition

C-Reactive Protein (CRP), which is named for its ability to "precipitate the somatic C-polysaccharide of *Streptococcus pneumonia"* (1), is an acute-phase protein which is a sensitive systemic marker indicating tissue damage, inflammation, infection and malignant neoplasia. It has also been found to be a highly stable analyte and "immunoassays for it are robust, well-standardized, reproducible, and readily available" (2). Because of these characteristics, CRP has been the most effective laboratory measurement in organic disease screening, monitoring inflammation response, detecting infection for individuals who are immunocompromised and in identifying chronic inflammation. (1)

High-sensitivity CRP (hsCRP) is a nephelometric measurement with sensitivity down to 0.04 mg/l (3), and is the measurement most often used to estimate an individual's risk for future coronary heart disease. The average hsCRP levels of middle-aged Americans fall between 1.00 mg/l and 2.00 mg/l. A single hsCRP measurement above 10.00 mg/l is probably an indication of acute infection. A repeated measurement (2-3 weeks after the initial high measurement) of above 10.00 mg/l is evidence of more serious, underlying inflammation (4). Level of hsCRP increase with age and obesity, have no diurnal or seasonal variation, and are affected by liver failure (2). Interestingly, most individuals will hold a consistent level of plasma hsCRP characteristic of their age and health. Two measurements taken years apart have a self correlation coefficient of 0.5 (2).

CRP Levels as a Response to Inflammation

Inflammation increases CRP synthesis in hepatocytes. CRP synthesis is controlled by cytokines, primarily interleukin 6 (IL-6), originating at the site of inflammation. Values of CRP can rise from less than .05 mg/l to over 5.00 mg/l within six hours. After 48 hours, CRP levels may peak at up to 500.00 mg/l (a 10,000 fold increase) (2). After the inflammation ceases, CRP levels drop quickly. Because the rate of degradation of CRP is constant under all conditions (its half-life is 19 hours), its rate of synthesis determines its circulating level and is a reflection of the stimulus intensity (5).

CRP Production

Following inflammation plasma CRP is synthesized in hepatocytes by the CRP gene (located on the short arm of chromosome 1). This synthesis is transcriptionally regulated by IL-6 and may be increased by the cytokine, interleukin-1β (IL-1β). Both IL-6 and IL-1β are produced at the site of inflammation and activate transcription factors in the cytidine-cytidine-adenosine-adenosine-thymidine-enhancer-binding-proteins (C/EBP) family, C/EBPβ and C/EBPδ. In the proximal promoter region of CRP, C/EBP binding sites interact with binding sites for signal transducer and activator of transcription 3 (STAT3) and Rel (proto-oncogene c-Rel) transcription factors. This interaction stabilizes binding of C/EBPβ and C/EBPδ to the CRP promoter, maximizing CRP production (6).

Tumor necrosis factor α (TNF-α) (7) and transforming growth factor- β (TGF-β) (8) may also help regulate CRP production. CRP production in other cells, such as neurons, monocytes, and lymphocytes has also been reported. However, the mechanisms of

production are unknown, and it is thought that production in these cells does not affect plasma CRP levels significantly (6).

1.2 Coronary Heart Disease

Coronary heart disease (CHD) is the leading cause of death for both men and women in the United States and is caused by atherosclerosis (9). Atherosclerosis is characterized by the accumulation of lipids and other substances (such as leukocytes and smooth muscle cells) on the arterial walls (9, 10). This accumulation, called a plaque or fiberatheroma, consists of a lipid core and fibrous cap (9, 10, 11, 12, 13). Leukocytes cause thinning of the plaques' fibrous cap by inhibiting collagen synthesis. The plaque may weaken and eventually rupture, causing clotting of arterial blood and interrupting blood flow to the heart, causing acute CHD or myocardial infarction (MI) (10, 14).

Coronary Heart Disease and Inflammation

Inflammation accompanies atherosclerosis. Leukocytes, mediators of inflammation mediators, accumulate in early atherosclerotic lesions when the normal endothelial monolayer of the arterial wall becomes inflamed. The cell surfaces express selective adhesion molecules, such as vascular cell adhesion molecule-1 (VCAM-1) (10) and endothelial leukocyte adhesion molecule-1 (ELAM) (14), which subsequently bind to monocytes and T lymphocytes. Leukocytes migrate into the intima in a process called

diapedesis (14). They also propagate an inflammatory response, including secretion of cytokines (IL-6, TNF-α, and IL-1β) (14), which are believed to influence CRP production(10). Because of this, CRP is used as a biomarker for the risk of CHD. Some investigators (14, 15) believe that CRP plays a causal role in atherosclerosis and is not simply a passive biomarker.

Coronary Heart Disease and CRP

In addition to atherosclerosis, CHD, and MI, elevated CRP levels have been associated with type 2 diabetes, stroke, and mortality (16, 17, 18, 19, 20, 21). According to American Heart Association (22), hsCRP levels below 1.00 mg/l indicate a low risk for CHD, levels between 1.00 mg/l and 3.00 mg/l indicate an elevated risk, and levels exceeding 3.00 mg/l indicate high risk. In one meta-analysis, it was shown that individuals with a baseline CRP level in the upper tertile have a relative risk of 2.0 for CHD, compared to individuals in the lowest tertile(20). Both men and women with CRP levels above 10.00 mg/l are 2 to 3 times more likely to experience MI, peripheral artery disease, and ischemic stroke (20, 21, 23, 24. Individuals with elevated CRP levels and unstable angina or MI are more likely to experience cardiac complications (25, 26). It has also been suggested that the susceptibility to coronary events of obese individuals may be attributable to the association of high plasma CRP levels with increased Body Mass Index (BMI) (20).

1.3 BMI and Obesity

Definition of Obesity

Obesity is defined by the World Health Organization (27)as having a BMI of over 30.00 $kg/m^{2;}$ a person with a BMI between 25 and 30 kg/m^2 is considered overweight. A normal weight, or lean, individual will have a BMI which falls between 18.50 and 24.99 kg/m^2. In 2007 and 2008, the overall age-adjusted prevalence of obesity in the United States was 33.8% (32.3% in men, 35.5% in women) (28). The overall prevalence of those either overweight or obese was 68.0%, 72.3% in men and 64.1% in women (28). In the past 30 years, the global prevalence of obesity has tripled (27, 29, 30). Around the world, one billion adults and 155 million children are overweight, and 300 million adults and 40 million children are obese (27, 31). The causes of obesity are both environmental and genetic in origin. Lack of physical activity and consumption of energy-dense foods with high saturated fat content are two of the major environmental factors influencing obesity (27). Smoking, alcohol consumption, and depression have also been suggested as causal factors (32, 33)

Obesity and CRP

A strong positive correlation exists between CRP levels and BMI (and other measures of obesity, including waist-to-hip ratio (20) and those used to estimate visceral adiposity (26)). This correlation could not be explained by other diseases or conditions known to increase CRP levels (34). When a person is obese, they are considered to be in a state of low-grade, chronic inflammation, marked by high CRP levels (7, 35). In obese individuals, elevated

CRP is caused by adipokines, chemokines, and cytokines secreted by adipocytes (36). Elevated cytokine levels, including IL-6 and TNF-α levels, have been observed in both obese adults and obese children (7), and it is estimated that 25% of systemic IL-6 is released into the circulation by adipose tissue (34). This means that in addition to being a state of inflammation, obesity may also be also a promoter of inflammation and is a potential mediator in the inflammatory process of coronary heart disease.

1.4 Genetics of CRP

Heritability

CRP heritability (the percentage of variation in CRP levels due to genetic variation) is approximately 0.3 – 0.4 in multiple populations, including Caucasian Americans, Caucasian Europeans, Japanese Americans, and Native Americans (37). A weakness of most heritability studies is that they typically assume pure additivity among genes and among genes and the environment. This thesis will investigate whether interactions between genes and obesity influence CRP levels.

Heritability estimates indicate that genes affect a trait, but they do not identify individual genes. Genome-Wide Association Studies (GWAS) have been used to identify single nucleotide polymorphisms (SNPs) associated with CRP levels and their corresponding genes. Table 1 is a list of SNPs associated with CRP in published genome-wide association studies (38).

Initial Population	Replication Population	Region	Reported Gene(s)	SNP – risk allele	P-Value	Reference
66,185 Individuals of European ancestry	16,540 Individuals of European ancestry	19q13.32	*APOC1*	rs4420638-A	9×10^{-139}	39
		1q23.2	*CRP*	rs2794520-C	2×10^{-186}	
		12q24.31	*HNF1A*	rs1183910-G	2×10^{-124}	
		1p31.3	*LEPR*	rs4420065-C	4×10^{-62}	
		1q21.3	*IL6R*	rs4129267-C	2×10^{-48}	
		2p23.3	*GCKR*	rs1260326-T	5×10^{-40}	
		2q13	*IL1F10*	rs6734238-G	2×10^{-17}	
		1q44	*NLRP3*	rs12239046-C	1×10^{-15}	
		8p23.1	*PPP1R3B*	rs9987289-A	3×10^{-13}	
		16q12.1	*SALL1*	rs10521222-C	9×10^{-13}	
		12q23.2	*ASCL1*	rs10745954-A	2×10^{-11}	
		1p34.3	*PABPC4*	rs12037222-A	6×10^{-11}	
		20q13.12	*HNF4A*	rs1800961-C	2×10^{-9}	
		15q22.2	*RORA*	rs340029-T	4×10^{-9}	
		7q11.23	*BCL7B*	rs13233571-C	4×10^{-9}	
		5q31.1	*IRF1*	rs4705952-G	1×10^{-8}	
		18p11.21	*PTPN2*	rs2847281-A	2×10^{-8}	
		6q22.1	*GPRC6A*	rs6901250-A	5×10^{-8}	
		21q22.2	*PSMG1*	rs2836878-G	2×10^{-7}	
		14q24.2	*RGS6*	rs4903031-G	5×10^{-6}	
10,112 Japanese Individuals	2,742 Japanese Individuals	1q23.2	CRP	rs3093059-G	4×10^{-21}	40
		7p15.3	IL6	rs2097677-A	4×10^{-11}	
		12q24.31	HNF1A	rs7310409-G	3×10^{-8}	
		19q13.32	APOE-CI-CII-cluster	rs4420638-A	3×10^{-7}	
17,967 European and Indian Asian men and women	17,967 European and Indian Asian men and women	1q23.2	*CRP*	rs7553007-A	8×10^{-44}	41
		12q24.31	*HNF1A*	rs1183910-T	1×10^{-30}	
		19q13.32	*APOE, APOC1, APOCII*	rs4420638-G	5×10^{-27}	
		1q21.3	*IL6R*	rs4537545-T	2×10^{-14}	
		1p31.3	*LEPR*	rs6700896-T	3×10^{-14}	
909 individuals	5,106 individuals	12q24.31	*HNF1A*	rs1169310-A	2×10^{-8}	42
		19q13.32	*APOE*	rs2075650-?	1×10^{-7}	
		1q23.2	*CRP*	rs11265260-?	7×10^{-6}	
6,345 women	NR	1q23.2	*CRP*	rs3091244-?	6×10^{-28}	43
		1p31.3	*LEPR*	rs1892534-A	7×10^{-21}	
		19q13.32	*APOE*	rs769449-?	9×10^{-21}	
		12q24.31	*HNF1A*	rs7310409-A	7×10^{-17}	
		2p23.3	*GCKR*	rs780094-A	7×10^{-15}	
		12q23.2	*Unknown*	rs10778213-G	1×10^{-10}	
		1q21.3	*IL6R*	rs8192284-?	2×10^{-8}	

Table 1: Genes associated with CRP in Published Genome-Wide Association Studies

CRP Gene

The most frequently studied gene in relation to CRP levels is the CRP gene itself, on chromosome 1q21-23. Variants in this gene have been found in multiple GWA and candidate gene studies to be associated with CRP levels (39, 40, 41, 42, 43, 36, 35, 44, 45). One GWA study found that 3.4% of residual variation in CRP levels could be attributed to the CRP gene (43). In another GWAS, *rs7553007* in the *CRP* gene was the single nucleotide polymorphism (SNP) found to be most associated with CRP (each minor allele lowered CRP levels by 21%) (41). Bochud et al. also reported CRP to be influenced by *rs7553007* which tags multiple SNPs in the CRP gene. In women only, this SNP was also positively associated with BMI, waist circumference, and fat mass (35).

IL-6 Gene

The IL6 gene on chromosome 7p15 has been studied in relation to CRP levels. A SNP in the IL6 promoter region, rs2097677, was significantly associated with CRP levels in a Japanese population (40). Five GWA studies have found variants in IL-6 to be associated with CRP levels (39, 40, 41, 42, 43). One study (43) found the contribution of IL-6 to variation in CRP levels to be 0.6%. A G-C transversion at -174, 5' to IL6, has been inversely associated with CRP levels (44).

*HNF1*A *Gene*

Hepatocyte nuclear factor 1 homeobox A (HNF1A), located on chromosome 12q24.31, was reported in the five GWA studies to be significantly associated with CRP levels. In addition to IL-6, nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), is involved in transcriptional control of CRP synthesis. NF-κB binding sites interact with HNF-1α binding sites in the CRP gene promoter to help regulate CRP synthesis. One GWAS (42) found five SNPs in HNF1A to be associated with CRP levels including two nonsynonymous coding SNPs, *rs1169288* and *rs2464196.* The strongest significantly associated SNP, however, was *rs1139310* in the HNF1A 3' untranslated region. Addition of each copy of the minor allele of this SNP was associated with "0.13 mg/l lower mean log(CRP) levels" (42). A common haplotype (frequency = 30%) consisting of the minor alleles of *rs1169288, rs1169286, rs2464196, and rs1169310* was also found to be associated with lower CRP levels. Ridker et al. found 8 SNPS in HNF1A to be significantly associated with CRP levels, and that 1.1% of residual variation in CRP levels could be attributed to these polymorphisms.

APOE and LEPR Genes

Genes that have been previously found to be associated with obesity and atherosclerosis is of particular interest in CRP GWA studies because of CRP's known association with BMI, waist circumference and adiposity. Such polymorphisms are contained in the apolipoprotein E gene (APOE) (Lange) and the leptin receptor gene (LEPR) (35). APOE, in addition to being involved in obesity, also plays a role in coronary heart

disease development. Its roles include "macrophage cholesterol efflux, platelet aggregation, and allele-specific antioxidant and immune activities" (46, 47, 48). APOE was significantly associated with CRP levels in 4 GWA studies (40, 41, 42, 43). Ridker and colleagues and Reiner and colleagues. both found two variants at 19q13.32 near the APOE gene to be associated with CRP (42, 43); 1.5% of variation in CRP levels was attributable to APOE variants (43). Ridker et al. also noted that "two highly significant SNPs near the APOE locus are in a block that includes APOC1 [apolipoprotein C1 gene] and TOMM40 [Translocase of outer mitochondrial membrane 40 homolog] loci" (43). Three GWA studies found LEPR to be associated with CRP levels (35, 43, 49, 39, 41). Ridker et al. estimated that 1.6% of variation in CRP levels was due to nine variants at the LEPR locus (1q31.3) (43). CRP expression is induced by leptin receptor activation in vascular endothelial cells (50). These data indicate that leptin may be a mediating factor in the relationship between CRP and BMI/adiposity. One GWAS found a significant interaction ($p < 2.9x10^{-6}$) between the LEPR gene and BMI that was associated with CRP levels.

GCKR

Two GWA studies (39, 43) found CPR levels to be associated with variants in the glucokinase (hexokinase 4) regulator (GCKR) at chromosome 2p23.2, which is expressed in hepatocytes. Ridker et al. estimated that 1.1% of variation in CRP levels is due to the SNPs in the GCKR gene. The GCKR gene codes for a regulatory protein which inhibits glucokinase in liver cells, possibly resulting in limited accumulation of glucose as glycogen in hepatocytes (43).

CHAPTER 2: RESEARCH DESIGN AND METHODS

2.1 Significance of Research

Obesity phenotypes (such as BMI) may interact with genes to increase a person's risk for many life-threatening diseases. Such diseases include atherosclerosis, CHD, type 2 diabetes, hypertension, and stroke (27). Obesity is also known to be associated with high plasma CRP levels (which may be an additional risk factor for cardiovascular disease). Defining the interaction between obesity and genes that affect CRP levels will help in better characterizing risk factors for the aforementioned diseases.

2.2 Hypothesis and Specific Aims

The main hypothesis of this thesis is that BMI interacts with genetic variation to influence CRP levels. This hypothesis will be tested by performing genome-wide association analyses on data from the Atherosclerosis Risk in Communities (ARIC) study.

Aim 1: Investigate whether body mass index interacts with genes in a GWA study to influence C-reactive protein levels in European-Americans.

Exploratory Aim 2: Compare and contrast two different statistical methods of analysis for assessing gene x environment interaction used to investigate the previous aim.

2.3 Genome Wide Association Studies

Background

A genome wide association study (GWAS) is a hypothesis-free analysis which takes between 300K and 1,000K SNPs into account. Many factors including the Human Genome Project (2003), the international HapMap project, improved genotyping technologies, and fast statistical genetic analysis technologies have all contributed significantly to the GWAS movement in recent years.

In ARIC

Genotyping was performed in the ARIC study via the Affymetrix Genome-Wide Human SNP Array 6.0 genotyping chip and quality control was performed (crucial because of the huge amount of data in the study). Of those European Americans genotyped, 9,345 individuals remained after initial quality control exclusions. Subjects were excluded if they did not allow use of their DNA. Individuals with genotype success rates < 95% and those with a greater percentage of heterozygous genotypes than anticipated were also removed from the sample. Relatedness was assessed using estimated identity by decent (IBD), and one suspected first-degree relative of an included individual was removed from the sample. (The subject with less missing data was retained). Other individuals were excluded because of suspected mixed/contaminated samples, problems with one DNA plate, mismatches between sex estimated from genotype data and phenotypic sex, mismatches with 39 previously genotyped SNPs, and identification as a genetic outlier based on average identity by state (IBS) statistics and principal components analysis. SNPs were excluded based on

no chromosome location, being monomorphic, being out of Hardy-Weinburg Equilibrium (HWE) (p-value < 10^{-5}), having a minor allele frequency (MAF) < 1%, and having a call rate < 95%. Out of the original 841,820 called SNPs, 669,450 made it past quality control and were used for imputation. Autosomal SNPs were imputed from HapMap Phase II CEU samples (consisting of Utah residents with ancestry from northern and western Europe), using Mach v1.0.16 software. A total of 2,543,887 SNPs will be used in this analysis.

2.4 Data Analysis Methods

Subjects

The GWA analyses for this project will be done in European Americans from the Atherosclerosis Risk in Communities (ARIC) study. The study, sponsored by the National Heart, Lung and Blood Institute (NHLBI), seeks to investigate atherosclerotic diseases and cardiovascular risk factors by race, gender, location and date. The initial cohort comprised 15,792 participants between the ages of 45 and 64 who were ascertained in 1987 from four different communities in the US. Roughly 4,000 individuals were recruited from each community. These communities included Forsyth County, NC; the city of Jackson, MS (where only African American residents were sampled); the northwestern suburbs of Minneapolis, MN; and Washington County, MD. Individuals in the cohort were examined every three years from 1987 until 1998 and continue to have yearly telephone follow-ups. Information from the initial exams included medical, social and demographic data. This thesis will use information from the 4th examination year (1996-1998), in which hsCRP, BMI, age, sex and genotype data were collected for 7,675 European American individuals.

Summary statistics will be calculated including mean, standard deviation, and median. Simple comparison statistics will be done using a student's t-test.

Determining Significance Levels

Due to the large number of tests required to run a GWAS, by chance alone, as many as 50,000 SNPs out of 1,000,000 will be significant at the 0.05 level. For this reason, a method such as Bonferroni's correction or permutation must be used to determine a significant p-value. It is typical to correct for 1,000,000 SNPs allowing for a nominal significance value of 0.05/2,500,000 or 5×10^{-8} (51).

Aim 1: Method 1

The first method used to approach aim 1 will be a stratified analysis. Individuals will be separated into two categories based on their BMI: obese and lean. Regardless of gender, an individual with a BMI ≥ 30.00 kg/m^2 will be placed in the obese group where as an individual with a BMI ≥ 18.50 kg/m^2 and less than 25.00 kg/m^2 will be placed in the lean group. The raw measurement of hsCRP will not be directly used, but will instead be natural log transformed to normalize the data. Genome wide association analyses will be performed in each group to identify genetic variations associated with ln(hsCRP). Each analysis will be done adjusting for sex and age. An additive genetic model for the alleles at each locus will be assumed.

Because ln(hsCRP) is a quantitative trait, a linear regression model for this analysis will be used.

$$\ln(\text{hsCRP}) = \mu + \beta_1 \text{age} + \beta_2 \text{sex} + \beta_3 \text{SNP} + \varepsilon$$

The output of the analysis will supply the minor allele, the minor allele frequency, the number of individuals used in the analysis, the β value for each SNP, and the standard error for each β value. Each β value is tested against the null hypothesis that $\beta = 0$ (or that hsCRP levels do not change with genotype) using a chi squared test. A corresponding p-value will be calculated for each test.

Once the analysis is complete in both categories of obesity, genetic variants with a significant association with ln(hsCRP) (p-value less than $5.00*10^{-8}$) will be examined. To ensure that SNPs with possible moderate effects on hsCRP are not disregarded, those with p-values less than 10^{-5} will also be inspected. The significant SNPs in each category will be separated into regions of 800 kilobases (kb) and the most significant SNP in each region will be chosen as the index SNP.

The difference between the index SNPs significant in one or both categories will be tested using the T-statistic:

$$(b_o - b_l) / \sqrt{(se_o^2 + se_l^2 - 2*\text{corr}(b_o,b_l)*se_o*se_l)}$$

where $b_{o/l}$ is the beta estimate for each SNP in the obese and lean groups, $se_{o/l}$ is the standard error of the β estimates for each SNP in the obese and lean groups and $\text{corr}(b_o,b_l)$ is the correlation between the lean and obese β estimates for all SNPs in the GWA analysis. Because of the large sample size, the t-statistic will be treated as a z-score. The p-value of

significance will be determined by dividing 0.05 by the number of unique regions being examined. This method will help identify those regions which have diverse effects on CRP levels for individuals in different weight classes.

Aim 1: Method 2

In the second method used to approach aim 1, the genome wide association analysis will be performed on all individuals combined (i.e. not a priori stratified for obesity status) with the inclusion of an additional parameter for BMI by gene interaction. The linear regression model for this analysis is:

$$\ln(\text{hsCRP}) = \mu + \beta_1 \text{age} + \beta_2 \text{sex} + \beta_3 \text{BMI} + \beta_4 \text{SNP} + \beta_4 \text{SNPxBMI} + \varepsilon$$

The output of the analysis for each SNP will provide a β-value of interaction and a corresponding p-value. A p-value of less than 10^{-5} will be considered to be statistically significant for this interaction analysis. The top SNPs will be separated into 800 kb regions with the most significant SNP in each region being chosen as the index SNP. Regions for which this interaction is significant can be considered to have an interactive effect with BMI on CRP levels.

Aim 2

Exploratory Aim 2 will be a comparison of the previously described methods. This comparison will identify which genetic regions are most likely to interact with BMI to influence CRP levels in European-Americans. In general, an indication of this will be

consistent effects across the two methods. Top regions that are significantly different between the two classes of obesity from the first method and those which show significant interaction with BMI from the second method will be identified. If these regions interact with BMI/obesity in the same direction (indicated by β values) then they can be considered to have an interactive effect on BMI. In each of the regions identified by both methods, genes near the index SNP or containing the index SNP will be considered as having possible interactive effects with BMI on hsCRP levels in European-Americans. The investigation of gene x BMI interactions using these methods are initial studies and results will need to be replicated in subsequent genome wide association studies.

CHAPTER 3: RESULTS

3.1 Summary Statistics

Of the 15,792 participants examined at baseline in the ARIC cohort, 11,440 are self-reported European-Americans. Of these individuals, 9,345 were genotyped and provided informed consent to use their genetic information. At the fourth examination, hsCRP levels were measured in 7,675 of the 9,345 genotyped individuals. One or more of BMI, age and sex were not reported in 12 of the 7,675 individuals with hsCRP levels. Table 2 contains summary statistics from the fourth examination for the 7,663 individuals with information including genotype, BMI, age, sex, and hsCRP levels.

The sample included 46.59% (n = 3,570) males and 53.41% (n = 4,093) females. The participants' ages ranged between 53 and 75 and the average age was 63.11 (SD = 5.63). Underweight (BMI < 18.50 kg/m^2) individuals made up 0.74% (n = 57) of the sample; 27.18% (n = 2,083) of the individuals were lean (18.50 $kg/m^2 \leq$ BMI < 25.00 kg/m^2); 40.61% (n = 3,112) of the individuals were overweight (25.00 $kg/m^2 \leq$ BMI < 30.00 kg/m^2); and 31.46% (n = 2,411) of the individuals were obese (BMI $\geq$ 30.00 kg/m^2). Mean BMI was 28.31 kg/m^2 (n = 7,663) with the average for men being 28.45 kg/m^2 and the average for women being 28.18 kg/m^2.

Levels of hsCRP ranged between 0.15 mg/l and 142.74 mg/l and the average level was 4.13 (SD = 6.24) mg/l. Individuals whose hsCRP levels fell below 1.00 mg/l made up 23.72%. The average BMI of these individuals (n = 1818) was 25.78 kg/m^2 (SD = 3.90). Individuals who had hsCRP levels between 1.00 mg/l and 3.00 mg/l made up 35.50% of the sample (average BMI 28.00 kg/m^2, SD = 4.55). Most individuals (40.98%), however, had hsCRP levels above 3.00 mg/l. This group's average BMI was 30.03 kg/m^2 (SD = 5.88). A

total of 479 individuals had hsCRP levels above 10.00 mg/l. These individuals had an average BMI of 30.90 kg/m^2 (SD = 5.68).

On average, men had a significantly lower hsCRP level (3.33 mg/l, SD = 5.94) than women (4.83 mg/l, SD = 6.43) ($p < 0.0001$). Mean hsCRP levels increased with BMI (Table 2). In all categories of BMI, mean hsCRP levels were above the average for American middle-aged adults (between 1.00 mg/l and 2.00 mg/l) (4). This was to be expected due to the age range of the cohort and the fact that CRP increases with age (2). Only individuals in the underweight category had a mean hsCRP level (2.11 mg/l, SD = 4.45) between 1.00 mg/l and 3.00 mg/l. The means in the lean (3.08 mg/l, SD = 5.70), overweight (3.85 mg/l, SD = 6.53) and obese (5.45 mg/l, SD = 6.15) categories were all above 3.00 mg/l.

Variable	N	Mean	Median	Std. Dev.	Minimum	Maximum
Age (years)	7,663	63.11		5.63	53.00	75.00
BMI (kg/m^2)	7,663	28.31		5.27	13.52	55.62
hsCRP (mg/l)	7,663	4.13	2.23	6.25	0.15	142.75

Sex	Frequency	Percent	BMI (kg/m^2)	Age (years)	hsCRP (mg/l)	Median hsCRP (mg/l)
Male	3,570	46.59	28.45(SD=4.42)	63.50(SD=5.64)	3.33(SD=5.94)	1.72
Female	4,093	53.41	28.18(SD=5.12)	62.77(SD=5.59)	4.83(SD=6.43)	2.99
Difference between sexes:			p = 0.0245	p < 0.0001	p < 0.0001	

Obesity Status	Frequency	Percent	BMI (kg/m^2)	Age (years)	hsCRP (mg/l)	
Underweight	57	0.74	17.41(SD=1.01)	64.39(SD=5.61)	2.11(SD=4.45)	0.69
Lean	2,083	27.18	22.87(SD=1.55)	63.15(SD=5.73)	3.08(SD=5.70)	1.44
Overweight	3,112	40.61	27.42(SD=1.40)	63.36(SD=5.66)	3.85(SD=6.53)	2.00
Obese	2,411	31.46	34.41(SD=4.16)	62.73(SD=5.47)	5.45(SD=6.15)	3.68

hsCRP Level (mg/l)	CHD Risk	Freq.	Percent	hsCRP (mg/l)	BMI (kg/m^2)	
≤ 1.00	Low	1818	23.72	0.61(SD=0.24)	25.78(SD=3.90)	0.63
1.00 – 3.00	Elevated	2705	35.3	1.83(SD=0.57)	28.00(SD=4.55)	1.72
≥ 3.00	High	3140	40.98	8.15(SD=8.20)	30.03(SD=5.88)	5.80
≥ 10.00	----------	479	6.25	21.84(SD=14.06)	30.90(SD=5.68)	18.02

Table 2: Exam Year 4 Summary Statistics

Frequencies and means of select exam year 4 variables in European-American participants in the ARIC cohort who had been previously genotyped. Basic mean comparison statistics are also provided.

3.2 AIM 1: Method 1

Summary Statistics

Since the first series of analyses uses information from only lean and obese individuals, summary statistics are provided for each group separately. There were 2,083 lean individuals with BMI values between 18.50 kg/m^2 and 24.99 kg/m^2. There were 2,411 obese individuals (BMI $\geq$ 30.00 kg/m^2). Table 3 contains summary statistics for lean individuals; Table 4 contains statistics for obese individuals.

The mean age in the lean sample was 63.15 years (SD = 5.72) while the mean age in the obese sample was 62.73 years (SD = 5.47) ($p = 0.0114$). The lean individuals in the sample had an average BMI of 22.87 (SD = 1.55) kg/m^2 and the average BMI for obese individuals was 34.41 (SD = 4.16) kg/m^2 ($p < 0.0001$). Lean individuals had an average hsCRP level of 3.06 (SD = 5.66) mg/l and obese individuals had an average level of 5.45 (SD = 6.15) mg/l ($p < 0.0001$).

In the lean group, the men (making up 35.67% of the sample) had a mean BMI (23.31 kg/m^2, SD = 1.37) that was significantly greater than that of the women in the group (22.63 kg/m^2, SD = 1.60) ($p < 0.0001$). The men, however, had a significantly lower mean hsCRP level (2.41 mg/l, SD = 3.87) than the women (3.45 mg/l, SD = 6.47) ($p < 0.0001$).

The obese women (making up 46.12% of the sample) had a significantly greater average BMI (35.14 kg/m^2, SD = 4.55) than the obese men (33.54 kg/m^2, SD = 3.46) ($p < 0.0001$). The women also had a significantly higher average hsCRP level (6.80 mg/l, SD = 6.80 mg/l) as compared to the level in men (3.88 mg/l, SD = 4.83) ($p < 0.0001$). Individuals

with hsCRP levels falling below 1.00 mg/l made up 37.25% (n = 776) of the lean subsample and only 9.87% (n = 238) of the obese sample. The percentage of individuals with hsCRP levels between 1.00 mg/l and 3.00 mg/l (elevated CHD risk) was similar between the lean group (33.75%, n = 703) and obese group (32.73%, n = 789). There was a very large difference, however, in the percentages of lean (29.00%, n = 604) and obese (57.40%, n = 1384) individuals with hsCRP levels above 3.00 mg/l, indicating high CHD risk. In the lean group, 82 individuals (3.94%) had hsCRP levels above 10.00 mg/l. A greater percentage (9.50%, n = 22.9) of obese individuals were also in this category.

Variable	Category	N	Mean	Median	Std. Dev.	Minimum	Maximum
Age (years)	Lean	2,083	63.15		5.72	53.00	74.00
	Obese	2,411	62.73		5.47	53.00	75.00
Difference between obesity strata : p = 0.0114							
BMI (kg/m^2)	Lean	2,083	22.87		1.55	18.53	24,99
	Obese	2,411	34.41		4.16	30	55.62
Difference between obesity strata : p < 0.0001							
hsCRP (mg/l)	Lean	2,083	3.08	1.44	5.7	0.15	111.44
	Obese	2,411	5.45	3.68	6.15	0.17	91.09
Difference between obesity strata : p < 0.0001							
ln(hsCRP) (ln(mg/l))	Lean	2,083	0.45		1.11	-1.87	4.71
	Obese	2,411	1.26		0.96	-1.78	4.51
Difference between obesity strata : p < 0.0001							

Table 3: Summary Statistics for Lean and Obese Individuals

Frequencies and means of select exam year 4 variables in lean and obese European-American participants in the ARIC cohort who had been previously genotyped. Basic mean comparison statistics are also provided.

Sex	Category	Frequency	Percent	BMI (kg/m^2)	Age (years)	hsCRP (mg/l)	Median hsCRP (mg/l)
Male	Lean	743	35.67	23.31(SD=1.37)	64.20(SD=5.73)	2.41(SD=3.87)	1.19
Female	Lean	1340	64.33	22.63(SD=1.60)	62.57(SD=5.65)	3.45(SD=6.46)	1.68
Difference between sexes in lean subsample:				p < 0.0001	p < 0.0001	p < 0.0001	

Sex	Category	Frequency	Percent	BMI(kg/m^2)	Age (years)	hsCRP(mg/l)	
Male	Obese	1112	46.12	33.54(3.46)	62.96(5.50)	3.88(4.83)	2.50
Female	Obese	1299	53.88	35.14(4.55)	62.52(5.44)	6.80(6.80)	4.98
Difference between sexes in obese subsample:				p < 0.0001	p = 0.0490	p < 0.0001	
Difference between obesity strata by sex:			**Male**	p < 0.0001	p < 0.0001	p < 0.0001	
			Female	p < 0.0001	p = 0.8370	p < 0.0001	

hsCRP Level (mg/l)	Category	CHD Risk	Freq./Perdent	BMI (kg/m^2)	Age (years)	hsCRP (mg/l)	
≤ 1.00	Lean	Low	776/37.25	22.65(SD=5.81)	62.90(SD=5.81)	0.57(SD=0.24)	0.57
1.00–3.00	Lean	Elevated	703/33.75	23.03(SD=5.65)	63.28(SD=5.65)	1.77(SD=0.58)	1.64
≥ 3.00	Lean	High	604/29.00	22.97(SD=1.52)	63.32(SD=5.71)	7.82(SD=8.89)	5.47
≥ 10.00	Lean	----------	82/3.94	22.88(SD=5.63)	63.17(SD=5.63)	23.15(SD=17.05)	17.60

hsCRP Level (mg/l)	Category	CHD Risk	Freq./Percent	BMI(kg/m^2)	Age (years)	hsCRP (mg/l)	
≤ 1.00	Obese	Low	238.9.87	32.67(SD=2.54)	62.79(SD=5.55)	0.68(SD=0.22)	0.70
1.00–3.00	Obese	Elevated	789/32.73	33.42(SD=3.42)	62.61(SD=5.44)	1.93(SD=0.57)	1.89
≥ 3.00	Obese	High	1384/57.40	35.27(SD=4.54)	62.78(SD=5.48)	8.28(SD=6.83)	6.26
≥ 10.00	Obese	----------	229/9.50	36.40(SD=5.48)	62.48(SD=5.81)	20.13(SD=9.65)	17.63
Difference between obesity strata by hsCRP level:			**≤ 1.00**	p < 0.0001	p = 0.7710	p < 0.0001	
			1.00–3.00	p < 0.0001	p = 0.0213	p < 0.0001	
			≥ 3.00	p < 0.0001	p = 0.0520	p = 0.2550	
			≥ 10.00	p < 0.0001	p = 0.3491	p < 0.0001	

Table 4: Summary Statistics for Lean and Obese Individuals by Sex and hsCRP Level

Frequencies and means of select exam year 4 variables in lean and obese European-American participants in the ARIC cohort who had been previously genotyped. These statistics are stratified for each obesity category by sex and hsCRP level (separately). Basic mean comparison statistics are also provided.

Natural Log Transformation

Histograms were plotted for hsCRP levels in both the lean and obese subsamples to investigate their distribution. Neither distribution was normal so the hsCRP level for each individual was natural log transformed for use in the linear regression analyses. Figures 1 and 2 show the distributions before and after transformation for the lean and obese subsamples respectively.

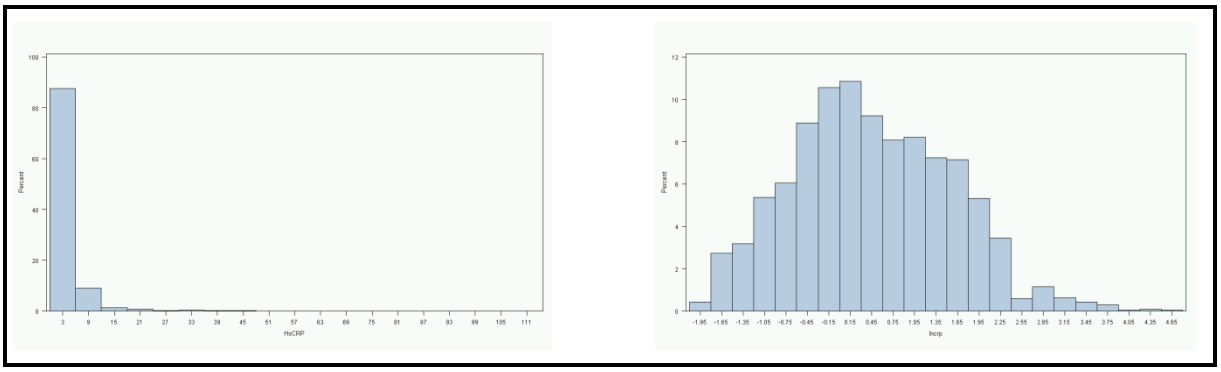


Figure 1: Lean Subsample Distribution of hsCRP Levels Before and After Natural Log Transformation

Percent of individuals is on the y-axis and hsCRP or ln(hsCRP) levels are on the x-axis.

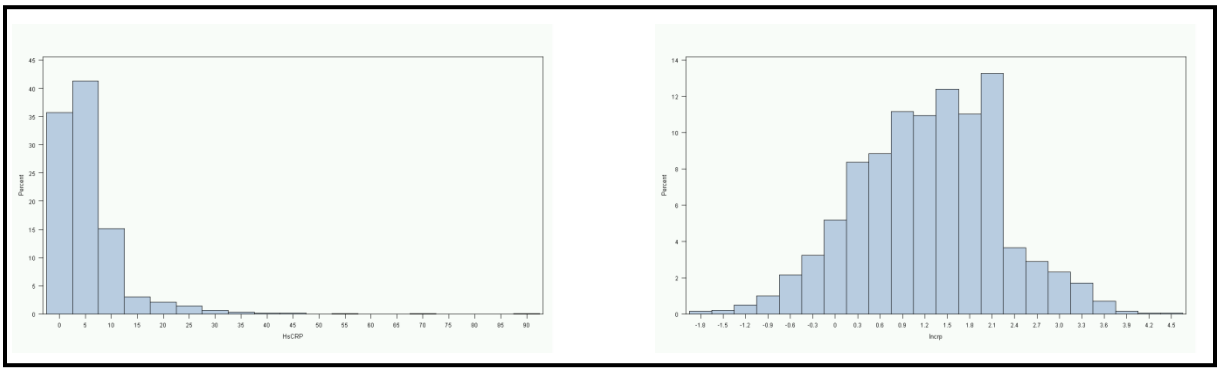


Figure 2: Obese Subsample Distribution of hsCRP Levels Before and After Natural Log Transformation

Percent of individuals is on the y-axis and hsCRP or ln(hsCRP) levels are on the x-axis.

GWA Results in Lean Individuals

This thesis first examined the association of 2,543,887 genotyped and imputed SNPs with natural log transformed hsCRP levels using linear regression in 2,083 lean European-Americans (18.50 $kg/m^2 \leq BMI < 25.00\ kg/m^2$). An additive genetic model was assumed, adjusted for age and sex. No SNP met the pre-determined genome-wide significant p-value of $2.00*10^{-8}$ and only one (rs2272417, $p = 2.67*10^{-8}$) had a p-value less than 10^{-7}. However, 145 SNPs reached moderate levels of significance (defined as $p < 10^{-5}$) (Supplementary Table 1). These 145 SNPs fell into 16 separate 800 kb regions across the genome. The SNPs with the most significant p-values of association in each region were selected as index SNPs. Table 5 contains information about these index SNPs.

name (rs)	chr	position	alleles	function	nearby genes	MAF	n	β	p-value
2272417	2	27560144	C/T	intron	IFT172(i), GCKR	0.4134	2083	-0.2026	2.670E-08
11265260**	1	157966663	A/G	unknown	CRP	0.0601	2083	-0.3639	1.450E-07
12463210	19	35788102	A/G	unknown	ZNF536	0.0075	2083	-0.9934	4.280E-07
4448101	6	96095326	C/T	unknown	MANEA	0.3587	2083	-0.1764	6.900E-07
157582	19	50088059	C/T	intron	TOMM40(i), APOE/C	0.2291	2083	0.3098	7.590E-07
663002	1	22798200	A/C	intron	EPHA8(i)	0.2174	2083	-0.2035	1.850E-06
12260720	10	33415966	C/T	unknown	ITGB1, NRP1	0.2135	2083	-0.2013	4.210E-06
7662351	4	30831931	C/T	unknown	PCDH7	0.0502	2083	0.3464	4.290E-06
12713007	2	48337971	C/T	unknown	FBXO11, FOXN2	0.4905	2083	0.1589	4.870E-06
10841753	12	21212637	C/T	intron	SLCO1B1(i)	0.1880	2083	0.2000	5.020E-06
4335431	1	216624232	C/T	intron	TGFB2(i)	0.0842	2083	-0.2775	5.870E-06
17052680	6	117407601	A/C	unknown	RFX6	0.2335	2083	-0.1814	6.870E-06
12568235	1	145785795	C/T	unknown	GJA5, GJA8	0.0092	2083	2.2839	6.910E-06
7026971	9	113371319	A/T	intron	PTGR1(i), ZNF483(i)	0.0427	2083	-0.3672	7.130E-06
13093806	3	157680330	A/C	intron	KCNAB1(i)	0.3751	2083	0.1672	7.860E-06
13041068	20	3109994	A/T	unknown	DDRGK1	0.0140	2083	-1.2894	8.940E-06

Table 5: Top Regions in Lean Individuals (Method 1)

Index SNPs from the top 16 genomic regions associated with hsCRP levels in European-American lean individuals. Only the most significant SNP, rs2272417, met the pre-determined genome wide significance value of $5*10^{-8}$. SNPs with function listed as unknown have nearby gene(s) reported. Those with function listed as intron have the gene they are in indicated with an (i) and any nearby gene reported (if relevant). ** = found to be previously associated with CRP levels in a GWAS (31)

The most significant SNP found to be associated with hsCRP levels in 2,083 lean individuals was rs2272417. Its significant negative β value indicates that with addition of each minor allele (C) hsCRP levels are lowered by 0.2026. This SNP is in the intraflagellar transport 172 homolog (IFT172) gene and is also in a region (see Figure 1) containing the GCKR gene which was reported in two previous GWAS to be associated with CRP levels (31, 56). Out of the 145 SNPs with a p-value of less than 10^{-5}, 51 are in the 800 kb region flanking IFT172 and GCKR (on chromosome 2). Along with rs2272417, 8 other SNPs are contained in IFT172. Eight SNPs in the region are found in the GCKR gene and 3 of these have p-values of less than $3*10^{-7}$ (rs11681351, rs4425043, and rs2293571).

The top SNP in the second most hsCRP-associated region in lean individuals is rs11265260 (p-value = $1.54*10^{-7}$). Figure 1 shows the region of association (on chromosome 1), containing the CRP gene, with rs11265260 as the index SNP. The SNP rs11265260 specifically has been previously reported and replicated in one previous CRP GWAS (59. Reiner). With addition of each minor allele (A) at the rs11265260 locus, hsCRP levels decrease, on average, by 0.36.

rs12463210 (p-value = $4.28*10^{-7}$) on chromosome 19 is in third most significant region associated with hsCRP levels in lean individuals. This SNP and 11 others with p-values < 10^{-6} are in the region (Figure 1) containing zinc finger nuclease 536 (ZNF536). Each additional A allele causes hsCRP levels to decrease, on average, by 0.99.

The SNP rs4448101 on chromosome 6 is also associated with hsCRP levels in lean individuals with a significance of $6.90*10^{-7}$. Each C allele leads to an average change in hsCRP of -0.18. Two other SNPs, rs2613562 (p-value = $3.39*10^{-6}$) and rs2612565 (p-value

= $3.50*10^{-6}$), met moderate levels of significance in the same region. The gene alpha-endomannosidase (MANEA) is contained in this region (Figure 1).

The fifth most significant associated region in lean individuals contains the SNP rs157582 (p-value = $7.59*10^{-7}$). Its significant positive β value indicates that each C allele increases hsCRP levels by 0.31. Genes in the 800 kb region (Figure 1) surrounding the SNP include TOMM40, APOE and APOC1. Two other SNPs in the region reached the moderate significance level of 10^{-6} (rs4420628 and rs6857).

Figure 1 also shows plots of 11 regions containing moderately significant top SNPs (p-values of 10^{-5} to 10^{-7}). Three of these regions were found in chromosome 1 and one was found in each of chromosomes 2, 3, 4, 6, 9, 10, 12, and 20. Gene found in these regions include: ephrin type-A receptor 8 (EPHA8), integrin β1 (ITGB1), neuropilin 1 (NRP1), protocadherin 7 (PCDH7), F-box protein 11 (FBXO11), forkhead box N2 (FOXN2), gap junction protein α5/8 (GJA5 and GJA8), voltage-gated potassium channel subunit beta-1 (KCNAB1), DDRGK domain containing 1 (DDRGK1), solute carrier organic anion transporter family member 1B1 (SLCO1B1), TGF-B2, regulator factor X 6 (RFX6), prostaglandin reductase 1 (PTGR1), and zinc finger nuclease 483 (ZNF483).

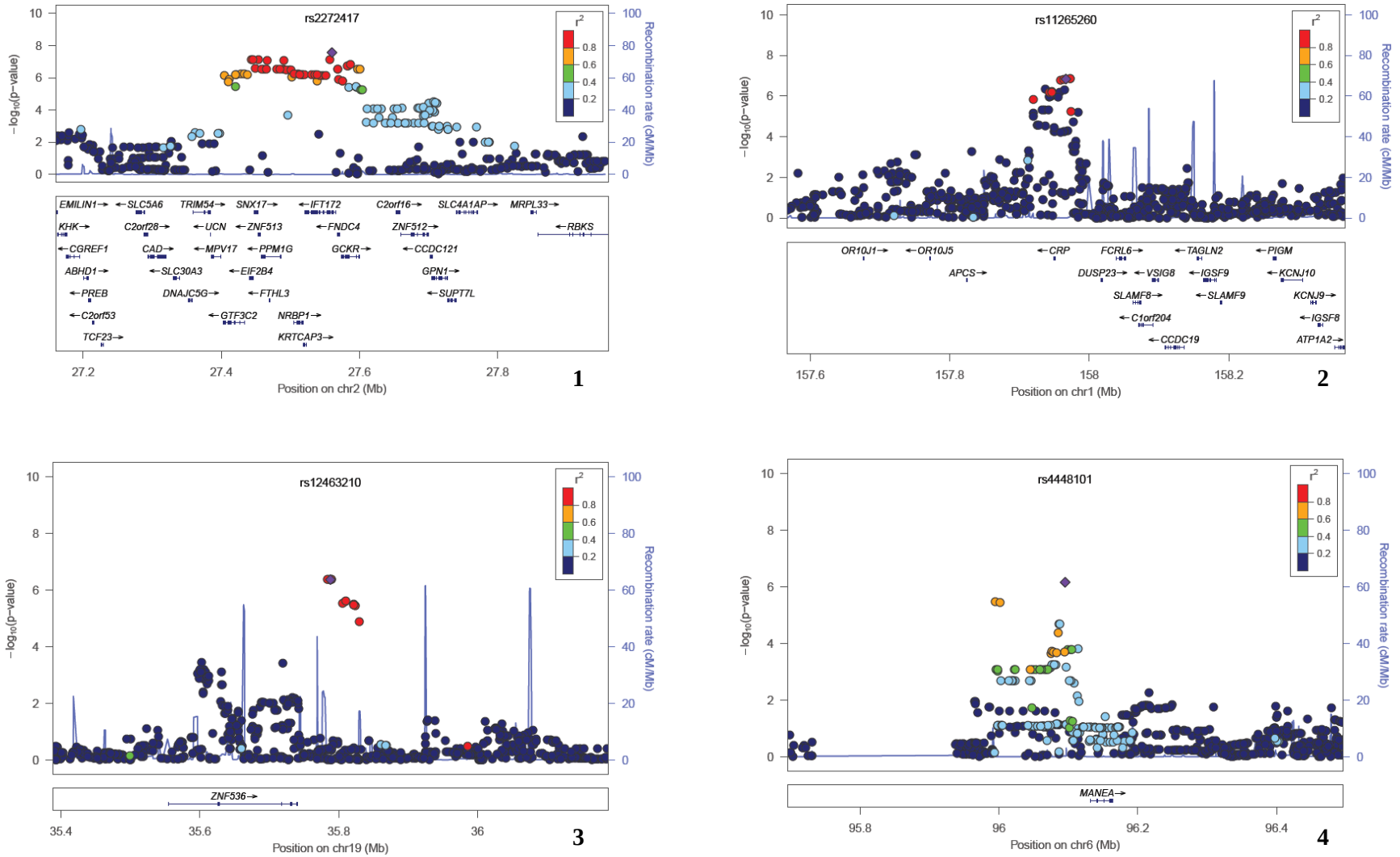


Figure 3: Plots of Top Regions in Lean Individuals

Regional plots of 16 loci showing moderate genome-wide significant association with hsCRP levels in European-American lean individuals. This plot shows the top 4 regions (refer to table 5). Each region's rank is at the bottom right corner of each plot. In each plot, the top SNP is designated by a purple diamond. $-\log_{10}$ p-values are on the y-axis and SNP genomic positions are on the x-axis. Genes which flank the 800 kb regions are reported under the x-axis. The estimated rates of recombination are marked in blue. Each SNP is colored to show linkage disequilibrium (LD) with the top SNP.

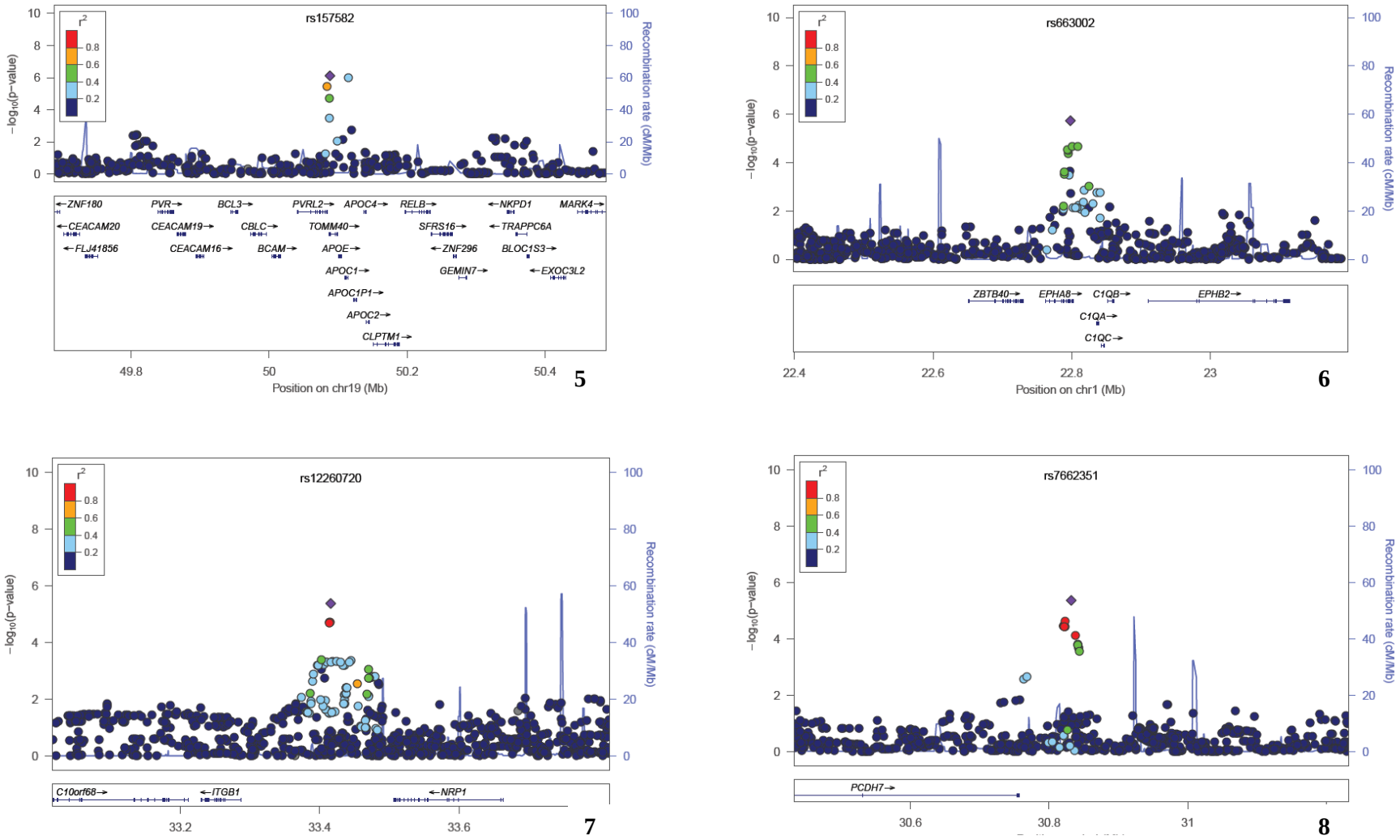


Figure 1 continued: Regional plots of 16 loci showing moderate genome-wide significant association with hsCRP levels in European-American lean individuals. This plot shows the 5th through 8th most significant regions (refer to table 5). Each region's rank is at the bottom right corner of each plot. In each plot, the top SNP is designated by a purple diamond. $-\log_{10}$ p-values are on the y-axis and SNP genomic positions are on the x-axis. Genes which flank the 800 kb regions are reported under the x-axis. The estimated rates of recombination are marked in blue. Each SNP is colored to show linkage disequilibrium (LD) with the top SNP.

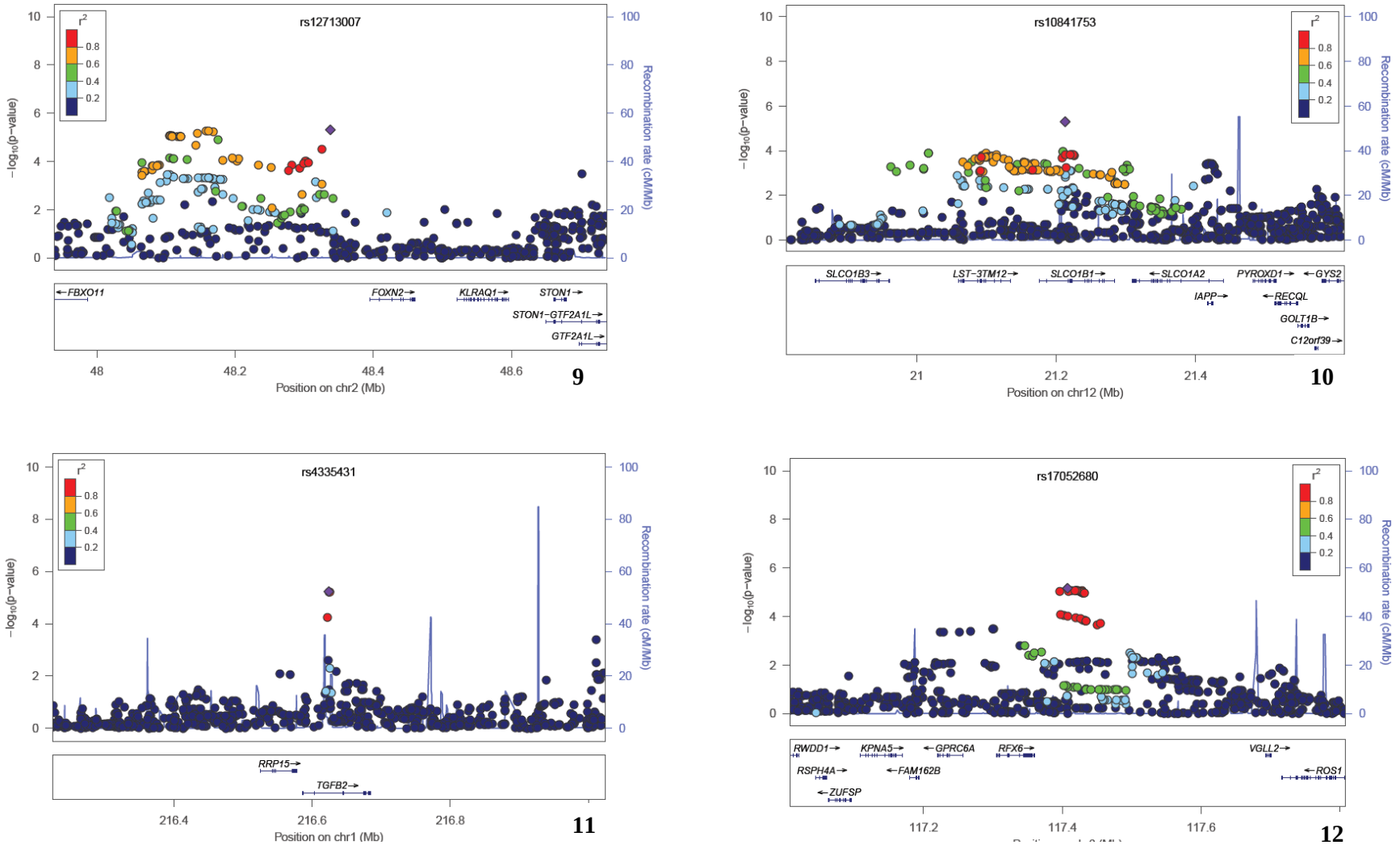


Figure 1 continued: Regional plots of 16 loci showing moderate genome-wide significant association with hsCRP levels in European-American lean individuals. This plot shows the 9th through 12th most significant regions (refer to table 5). Each region's rank is at the bottom right corner of each plot. In each plot, the top SNP is designated by a purple diamond. $-\log_{10}$ p-values are on the y-axis and SNP genomic positions are on the x-axis. Genes which flank the 800 kb regions are reported under the x-axis. The estimated rates of recombination are marked in blue. Each SNP is colored to show linkage disequilibrium (LD) with the top SNP.

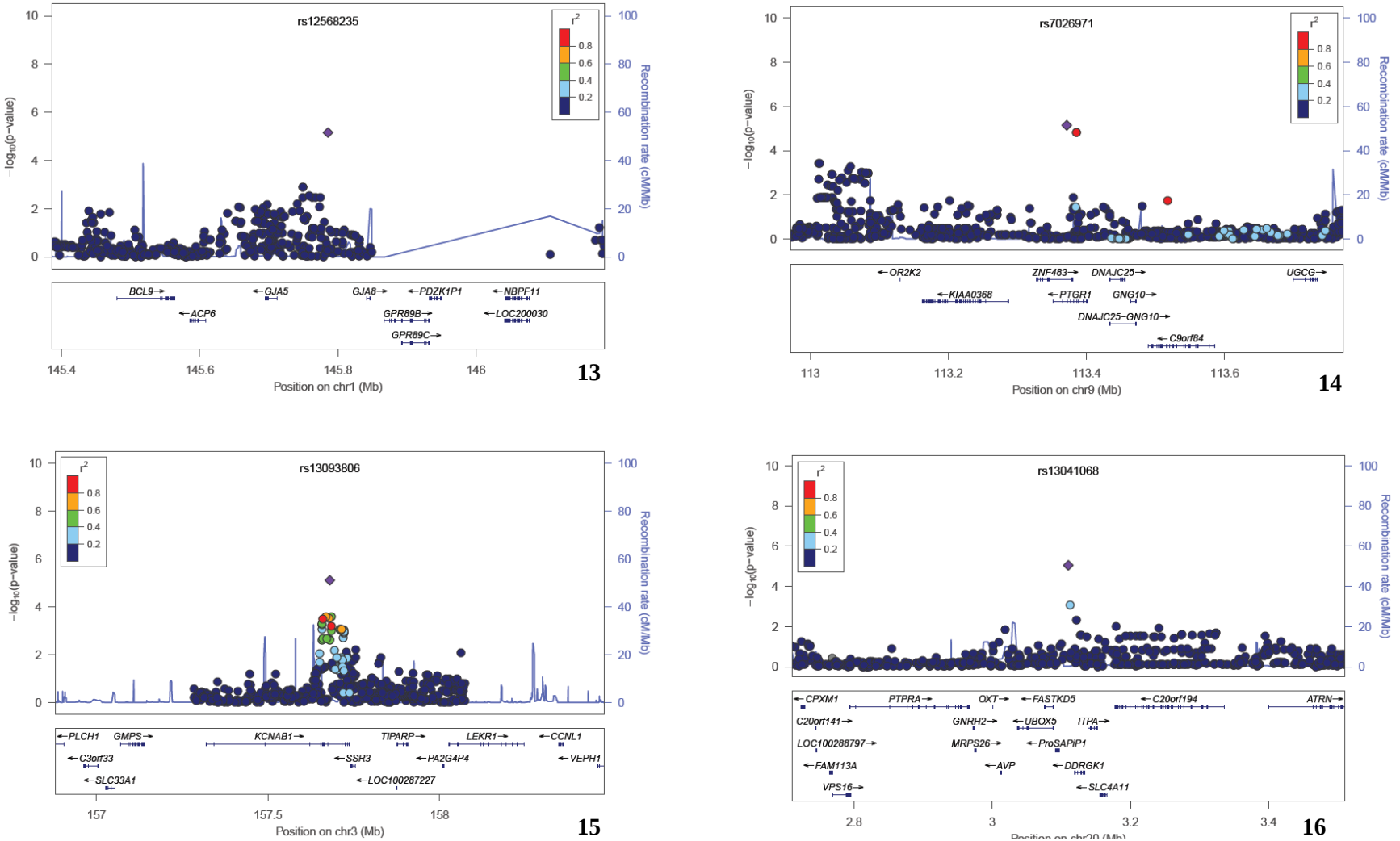


Figure 1 continued: Regional plots of 16 loci showing moderate genome-wide significant association with hsCRP levels in European-American lean individuals. This plot shows the 13th through 16th most significant regions (refer to table 5). Each region's rank is at the bottom right corner of each plot. In each plot, the top SNP is designated by a purple diamond. $-\log_{10}$ p-values are on the y-axis and SNP genomic positions are on the x-axis. Genes which flank the 800 kb regions are reported under the x-axis. The estimated rates of recombination are marked in blue. Each SNP is colored to show linkage disequilibrium (LD) with the top SNP.

GWA Results in Obese Individuals

Genome-wide linear regression was subsequently performed to investigate the association of 2,543,887 genotyped and imputed SNPs with ln(hsCRP) in 2,411 obese (BMI $\geq$ 30.00 kg/m^2) European-Americans. Just as in the lean group, the analysis was adjusted for age and sex and an additive genetic model was assumed. Significant or suggestive association was observed in only 36 SNPs (p-value < 10^{-5}) (see supplementary table 2). The top 4 most significantly associated SNPs all fell into the same region on chromosome 19 flanking the genes TOMM40, APOE, and APOC. Two of these polymorphisms, rs4420638 (p-value = $1.86*10^{-9}$) and rs6857 (p-value = $1.31*10^{-8}$), reached the pre-determined genome-wide significance level of $2.00*10^{-8}$. The 5th most associated SNP (rs12068753, p-value = $8.07*10^{-7}$) was located on chromosome 1 in a region containing the CRP gene. These SNPs and the 31 others reaching suggestive significance levels were grouped into 11 representative 800 kb regions of association across the genome. The SNPs with the most significant p-value in each region were selected as index SNPs (see table 6).

name (rs)	chr	position	alleles	function	nearby genes	MAF	n	β	p-value
4420638	19	50114786	A/G	intron	LOC100129500(i), APOE/C	0.1736	2411	0.2155	1.860E-09
12068753	1	157959161	A/T	unknown	CRP	0.0623	2411	0.2711	8.020E-07
13330968	16	79898765	G/T	unknown	BCMO1, GAN	0.0141	2411	-1.5731	1.490E-06
4278077	7	33356392	C/G	intron	BBS9(i)	0.0158	2411	0.4425	6.190E-06
2808898	9	43432012	C/T	unknown	FAM75A6	0.1265	2411	0.9764	6.700E-06
5008004	12	40283747	C/T	unknown	PDZRN4	0.3251	2411	0.1286	8.490E-06
17086609	13	27827711	A/G	intron	FLT1(i)	0.2992	2411	-0.1530	8.560E-06
629882	11	125747907	C/T	intron	ST3GAL4(i)	0.1278	2411	-0.1736	8.890E-06
4903731	14	77765959	A/G	unknown	NRXN3	0.0743	2411	0.2201	9.160E-06
10408037	19	56133595	A/G	unknown	KLK5	0.0090	2411	-1.0801	9.940E-06
7019284	9	28384358	G/T	intron	LINGO2	0.3167	2411	0.1260	9.950E-06

Table 6: Top Regions in Obese Individuals (Method 1)

Index SNPs from 11 genomic regions associated with hsCRP levels in obese European-Americans. Only the top SNP, rs4420638, reached the set genome wide significance level of $5.00*10^{-8}$. SNPs with function listed as unknown have nearby gene(s) reported. Those with function listed as intron have the gene they are in indicated with an (i) and any nearby gene reported (if relevant).

The most significant region of association is represented by the SNP rs4420638 (p-value $1.86*10^{-9}$) on chromosome 19 (figure 2). This SNP has a β value of 0.22. In the top 36 significantly associated SNPs, only the top four are in this region. rs6857 (p-value = $1.31*10^{-8}$) is located in the poliovirus receptor-related 2 (PVRL2) gene. The third most significant SNP (p-value = $3.84*10^{-7}$) is rs2075650 and it is contained in the TOMM40 gene. The SNP rs439401 (p-value = $4.71*10^{-7}$) in LOC100129500 (also the gene in which rs4420638 is located) is the fourth most significantly hsCRP associated SNP.

23 SNPs located in chromosome 1 in the region containing the CRP gene are moderately associated with hsCRP levels in the 2,411 obese individuals. This region is represented in figure 2 with rs12068753 as the index SNP (p-value = $8.02*10^{-7}$). With addition of each A allele at the rs12068753 locus, hsCRP levels increase by 0.27. rs11265260, the top SNP in this region in lean individuals, is associated with hsCRP levels in obese individuals with a significance level of $7.15*10^{-6}$.

The other 9 SNPs in the top 36 were the only SNPs in their respective regions (see figure 2) found to be associated with hsCRP levels in obese individuals. These SNPs were found in chromosome 16 (rs13330968, p-value = $1.49*10^{-6}$), chromosome 7 (rs4278077, p-value = $6.19*10^{-6}$), chromosome 9 (rs2808988, p-value = $6.70*10^{-7}$ and rs7019284, p-value = $9.95*10^{-6}$), chromosome 12 (rs5008004, p-value = $8.49*10^{-6}$), chromosome 13 (rs17086609, p-value = $8.56*10^{-6}$), chromosome 11 (rs629882, p-value = $8.89*10^{-6}$), chromosome 14 (rs4903731, p-value = $9.16*10^{-6}$), and chromosome 19 (rs10408037, p-value = $9.94*10^{-6}$). Genes in these regions include Bardet-Biedl syndrome 9 (BBS9), leucine rich repeat and Ig domain containing 2 (LINGO2). PDZ domain containing ring finger 4 (PDZRN4), vascular endothelial growth factor receptor 1 (FLT1), CMP-N-

acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase (ST3GAL4), neurexin 3 (NRXN3), Kallikrein-5 (KLK5), and family with sequence similarity 75, member A6 (FAM75A6).

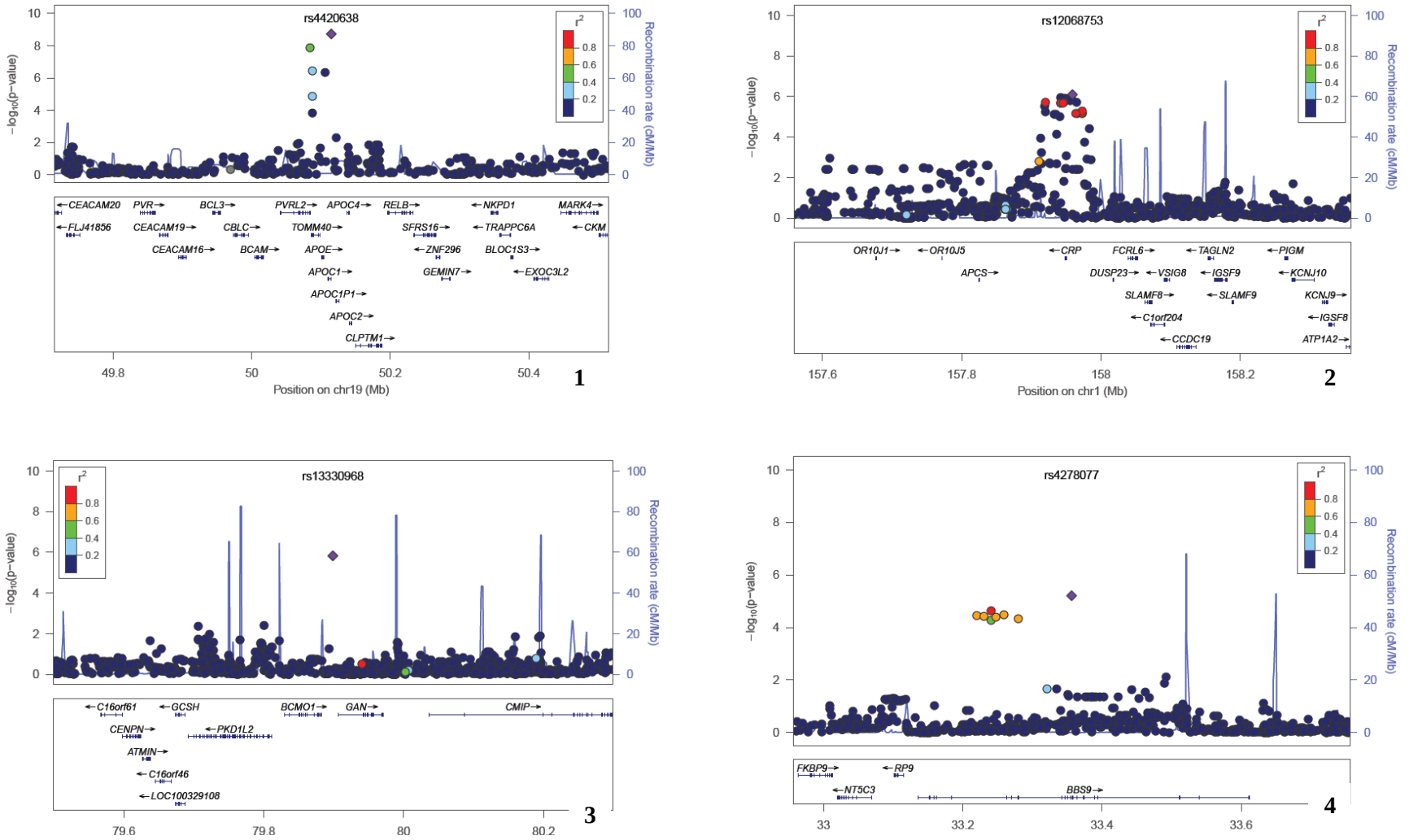


Figure 4: Plots of Top Regions in Lean Individuals

Regional plots of 11 loci showing moderate genome-wide significant association with hsCRP levels in European-American obese individuals. This plot shows the four most significant regions (refer to table 6). Each region's rank is at the bottom right corner of each plot. In each plot, the top SNP is designated by a purple diamond. $-\log_{10}$ p-values are on the y-axis and SNP genomic positions are on the x-axis. Genes which flank the 800 kb regions are reported under the x-axis. The estimated rates of recombination are marked in blue. Each SNP is colored to show linkage disequilibrium (LD) with the top SNP.

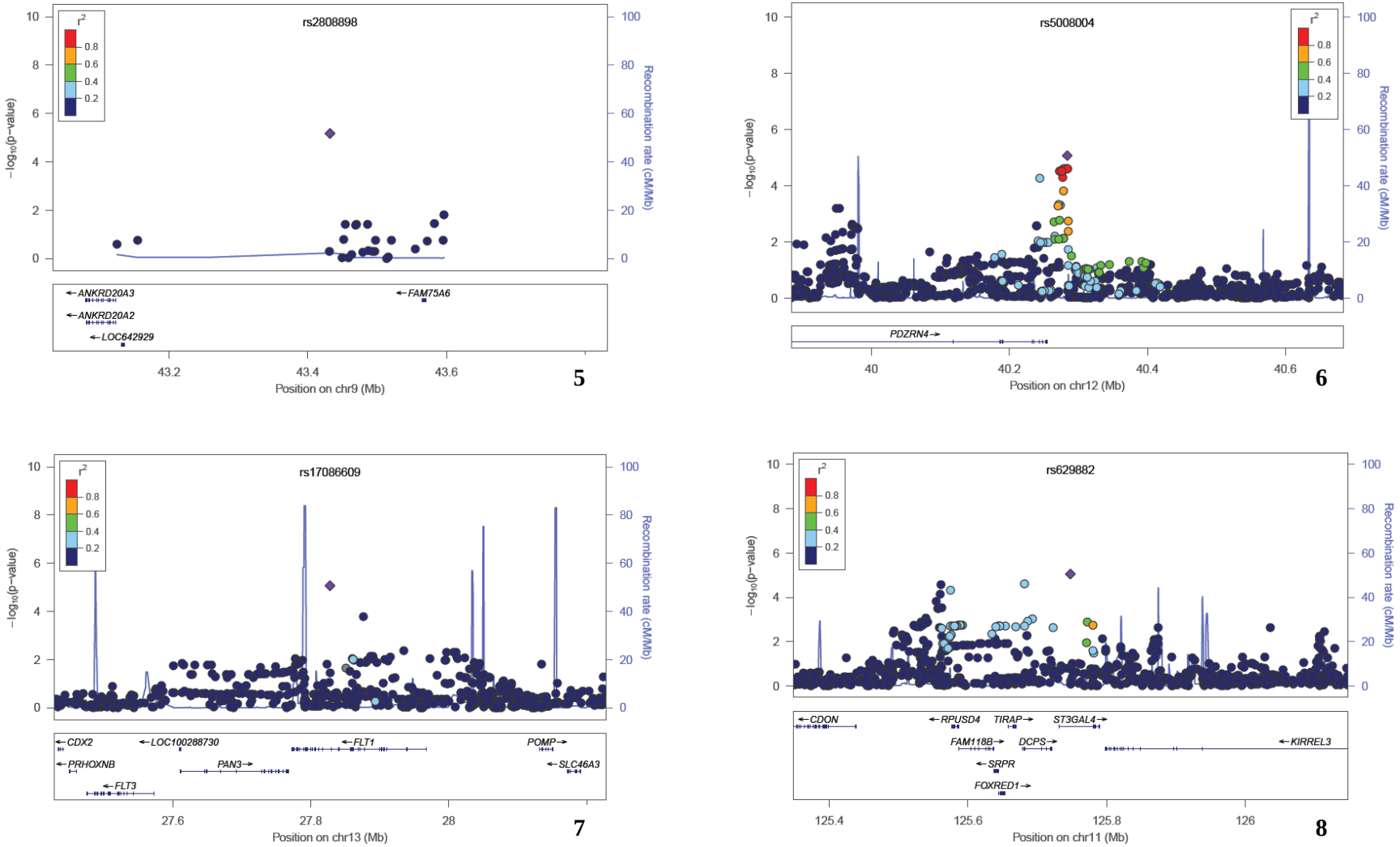
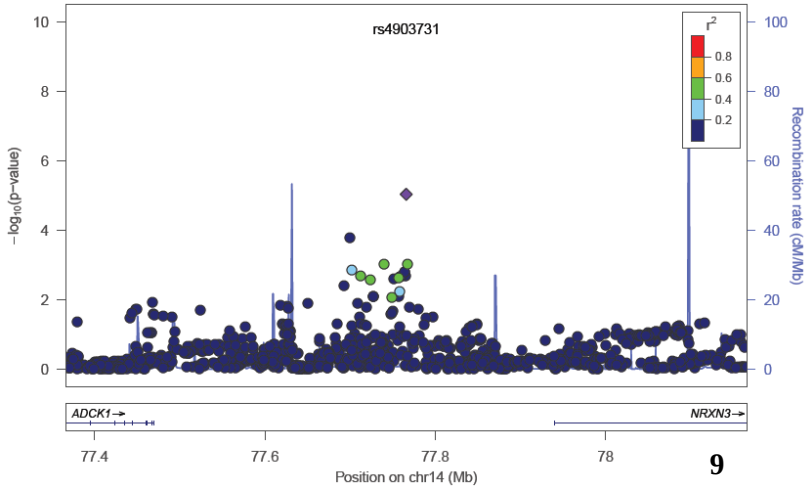
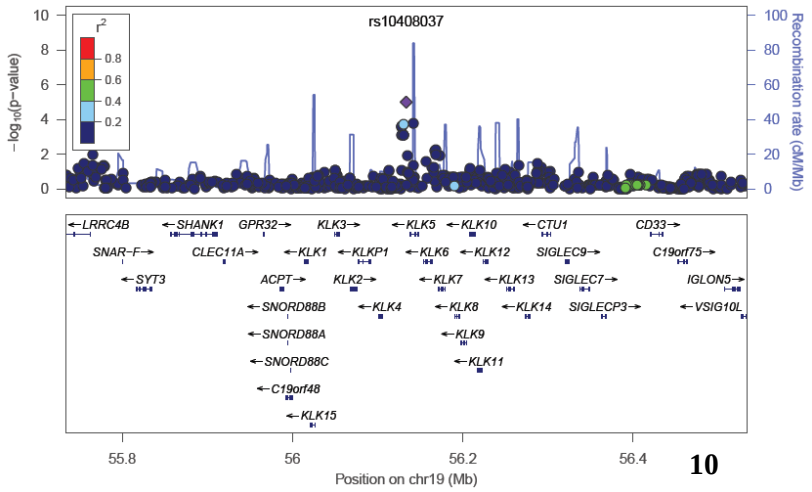


Figure 2 continued: Regional plots of 11 loci showing moderate genome-wide significant association with hsCRP levels in European-American obese individuals. This plot shows the 5th through 8th most significant regions (refer to table 6). Each region's rank is at the bottom right corner of each plot. In each plot, the top SNP is designated by a purple diamond. $-\log_{10}$ p-values are on the y-axis and SNP genomic positions are on the x-axis. Genes which flank the 800 kb regions are reported under the x-axis. The estimated rates of recombination are marked in blue. Each SNP is colored to show linkage disequilibrium (LD) with the top SNP.





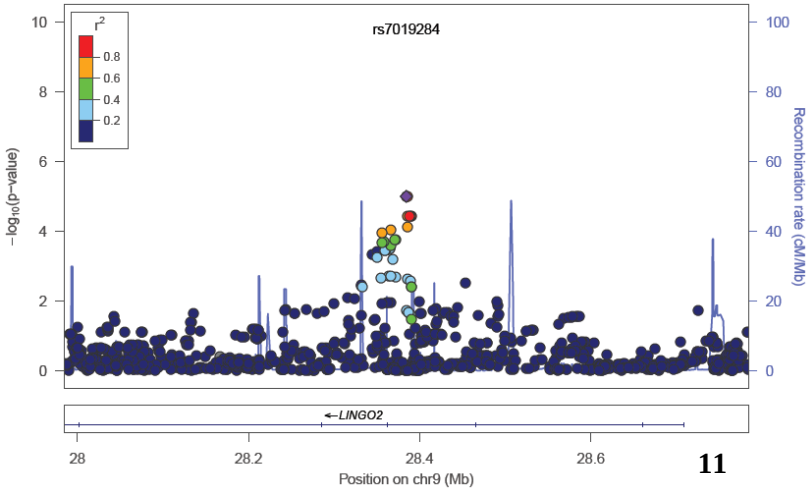


Figure 2 continued: Regional plots of 11 loci showing moderate genome-wide significant association with hsCRP levels in European American obese individuals. This plot shows the 9th through 11th most significant regions (refer to table 6). Each region's rank is at the bottom right corner of each plot. In each plot, the top SNP is designated by a purple diamond. $-\log_{10}$ p-values are on the y-axis and SNP genomic positions are on the x-axis. Genes which flank the 800 kb regions are reported under the x-axis. The estimated rates of recombination are marked in blue. Each SNP is colored to show linkage disequilibrium (LD) with the top SNP.

Comparison of GWA Results between Lean and Obese Individuals

The first aim of this thesis was to investigate which SNPs interact with BMI to influence the outcome of hsCRP levels. The first method used to do this was to compare the SNP association results from the separate GWA analyses done in lean and obese individuals (results above). Index SNPs from the 16 regions of significant association with hsCRP levels in lean individuals and the 11 regions of significant association with hsCRP levels in obese individuals were used in this analysis. Two of those regions were similar in both the lean and obese individuals leaving 25 unique regions.

To compare the effects of each region between BMI strata, the β value for each index SNP in lean individuals was compared to the β value for each index SNP in obese individuals using a t-statistic:

$$(b_o - b_l) / \sqrt{(se_o^2 + se_l^2 - 2*corr(b_o,b_l)*se_o*se_l)}$$

The Pearson correlation between the lean and obese β estimates for all SNPs in the GWAS was calculated to be 0.0305. Because of the large sample size (2,083 lean individuals, 2,411 obese individuals), the t-test is treated as a z-test. The p-value for significance was set at 0.05 divided by the number of independent SNPs being investigated ($0.05/25 = 2.00*10^{-3}$). This 2-sided p-value corresponds to the z-test statistic of ± 3.09. In any SNP comparison with a test statistic greater than 3.09 and less than -3.09 the null hypothesis is rejected and that SNP can be considered to differ significantly between obesity strata in their influence of hsCRP levels. Table 7 provides information on the 16 regions which were found to be significant by this test.

Of the 16 regions found to differ significantly in their effects on lean and obese individuals, the most significant was on chromosome 2. The index SNP was rs12713007 and had a p-value of $2.00*10^{-6}$. The region was identified in the GWA performed in the lean subsample. It contained 18 SNPs which all had moderately significant ($10^{-6} < 10^{-7}$) p-values of association. None of these SNPs were in a known gene, however, all 18 SNPs are contained in the 400 kb region between the genes FBXO11 and FOXN2.

The 15 other regions found to differ significantly between lean and obese individuals had index SNPs with p-values ranging between $1.88*10^{-3}$ and $6.00*10^{-6}$. These include regions located in or near KCNAB1 (rs13093806, p-value = $6.00*10^{-6}$) on chromosome 3, TGFB2 (rs4335431, p-value = $3.00*10^{-5}$) on chromosome 1, ITGB1 and NRP1 (rs12260720, p-value = $3.20*10^{-5}$) on chromosome 10, PTGR1 and ZNF483 (rs7026971, p-value = $5.00*10^{-5}$) on chromosome 9, GJA5 and GJA8 (rs12568235, p-value = $5.80*10^{-5}$) on chromosome 1, NRXN3 (rs4903731, p-value = $9.40*10^{-5}$) on chromosome 14, ZNF536 (rs12463210, p-value = $1.22*10^{-4}$) on chromosome 19, EPHA8 (rs663002, p-value = $2.30*10^{-4}$) on chromosome 1, PCDH7 (rs7662351, p-value = $4.43*10^{-4}$) on chromosome 4, IFT172 and GCKR (rs2272417, p-value = $4.68*10^{-4}$) on chromosome 2, MANEA (rs4448101, p-value = $5.06*10^{-4}$) on chromosome 6, SLCO1B1 (rs10841753, p-value = $6.08*10^{-4}$) on chromosome 12, PDZRN4 (rs5008004, p-value = $7.16*10^{-4}$) on chromosome 12, KLK5 (rs10408037, p-value = $7.76*10^{-4}$) on chromosome 19, and FLT1 (rs17086609, p-value = $1.88*10^{-3}$) on chromosome 13.

name (rs)	chr	position	alleles	function	nearby genes	β_lean	se_β_lean	p-val_lean	β_obese	se_β_obese	p-val_obese	t-stat	2-sided p-val	Regional Plot
12713007	2	48337971	C/T	unknown	FBXO11, FOXN2	0.1589	0.0348	4.87E-06	-0.0418	0.0261	0.1095	-4.6837	2.00E-06	Figure 3 #9
13093806	3	157680330	A/C	intron	KCNAB1(i)	0.1672	0.0374	7.86E-06	-0.0413	0.0280	0.1405	-4.5278	6.00E-06	Figure 3 #15
4335431	1	216624232	C/T	intron	TGFB2(i)	-0.2775	0.0612	5.87E-06	0.0460	0.0492	0.3500	4.1799	3.00E-05	Figure 3 #11
12260720	10	33415966	C/T	unknown	ITGB1, NRP1	-0.2013	0.0437	4.21E-06	0.0253	0.0340	0.4562	4.1518	3.20E-05	Figure 3 #7
7026971	9	113371319	A/T	intron	PTGR1(i),ZNF483(i)	-0.3672	0.0818	7.13E-06	0.0524	0.0659	0.4265	4.0556	5.00E-05	Figure 3 #14
12568235	1	145785795	C/T	unknown	GJA5, GJA8	2.2839	0.5079	6.91E-06	-0.1563	0.3471	0.6526	-4.0240	5.80E-05	Figure 3 #13
4903731	14	77765959	A/G	unknown	NRXN3	-0.0875	0.0627	0.1629	0.2201	0.0496	9.16E-06	3.9053	9.40E-05	Figure 4 #9
12463210	19	35788102	A/G	unknown	ZNF536	-0.9934	0.1965	4.28E-07	-0.0034	0.1730	0.9845	3.8405	1.22E-04	Figure 3 #3
663002	1	22798200	A/C	intron	EPHA8(i)	-0.2035	0.0427	1.85E-06	-0.0100	0.0320	0.7553	3.6832	2.30E-04	Figure 3 #6
7662351	4	30831931	C/T	unknown	PCDH7	0.3464	0.0753	4.29E-06	0.0163	0.0582	0.7800	-3.5184	4.34E-04	Figure 3 #8
2272417	2	27560144	C/T	intron	IFT172(i), GCKR	-0.2026	0.0364	2.67E-08	-0.0461	0.0271	0.0889	3.4979	4.68E-04	Figure 3 #1
4448101	6	96095326	C/T	unknown	MANEA	-0.1764	0.0355	6.90E-07	-0.0231	0.0272	0.3943	3.4770	5.06E-04	Figure 3 #4
10841753	12	21212637	C/T	intron	SLCO1B1(i)	0.2000	0.0438	5.02E-06	0.0165	0.0321	0.6081	-3.4281	6.08E-04	Figure 3 #10
5008004	12	40283747	C/T	unknown	PDZRN4	-0.0301	0.0378	0.4265	0.1286	0.0289	8.49E-06	3.3833	7.16E-04	Figure 4 #6
10408037	19	56133595	A/G	unknown	KLK5	0.3790	0.3663	0.3008	-1.0801	0.2445	9.94E-06	-3.3612	7.76E-04	Figure 4 #10
17086609	13	27827711	A/G	intron	FLT1(i)	0.0219	0.0456	0.6318	-0.1530	0.0344	8.56E-06	-3.1083	1.88E-03	Figure 4 #7

Table 7: Top Regions Identified by Method 1 Comparison

Information on 16 regions which represent the regions found through comparison of obese strata-specific β values to have significantly different effects on hsCRP levels. SNPs with function listed as unknown have nearby gene(s) reported. Those with function listed as intron have the gene they are in indicated with an (i) and any nearby gene reported (if relevant).

3.3 AIM 1: Method 2

The second series of analyses used a SNP by BMI interaction term to investigate the pairwise interaction between the 2,543,887 genetic polymorphisms and BMI in all genotyped individuals (i.e. not a priori stratified for obesity status).. The linear regression was adjusted for age, sex and SNP, and an additive genetic model was assumed. 40 SNPs reached the set significance level of less than 10^{-5} (see table 8). These 40 SNPs were contained in 11 genomic regions.

The index SNP with the most significant interaction term (p-value = $6.710*10^{-7}$) was found on chromosome 5 in the gene encoding transcription factor TFIIIB component B homolog (BDP1). The index SNPs in the remaining 10 regions found to interact significantly with BMI to influence hsCRP levels include: rs4265829 (p-value = $1.980*10^{-6}$) on chromosome 8 near LOC100132891, rs1481852 (p-value = $2.310*10^{-6}$) on chromosome 5 in the gene cadherin 9 type 2 (CDH9), rs7476844 (p-value = $3.32*10^{-6}$) on chromosome 10 in no known gene, rs12568235 (p-value = $3.450*10^{-6}$) near GJA5 and GJA8 on chromosome 1, rs12713007 (p-value = $5.11*10^{-6}$) near FXO11 and FOXN2 on chromosome 2, rs1955377, p-value = $6.200*10^{-6}$ on chromosome 3, rs1793615 (p-value = $8.330*10^{-6}$) near the neurotrimin gene (NTM) on chromosome 11, rs11708430 (p-value = $8.63*10^{-6}$) on chromosome 3 in no known gene, rs11136787 (p-value = $9.190*10^{-6}$) in CUB and sushi domain-containing protein 1 (CSMD1) on chromosome 8, and rs9386702 (p-value = $9.88*10^{-6}$) in osteopetrosis associated transmembrane protein 1 (OSTM1) on chromosome 6.

name (rs)	chr	position	alleles	function	nearby genes	MAF	n	β_interaction	p-value_interaction
4365829	5	70894524	G/T	intron	BDP1(i)	0.0056	7663	0.321544	6.71E-07
1481852	8	72914781	A/G	near-gene-3	LOC100132891	0.3489	7663	0.021557	1.98E-06
11954115	5	26901834	C/T	unknown	CDH9	0.0559	7663	-0.0307423	2.31E-06
7476844	10	54234734	G/T	unknown	none	0.225	7663	0.0175078	3.32E-06
12568235	1	145785795	C/T	unknown	GJA5/8	0.0092	7663	-0.203205	3.45E-06
12713007	2	48337971	C/T	unknown	FBXO11, FOXN2	0.4905	7663	-0.0137985	5.11E-06
1955377	3	165554969	C/T	unknown	none	0.0763	7663	0.0282072	6.20E-06
1793615	11	130593236	A/G	unknown	NTM	0.3954	7663	-0.0206864	8.33E-06
11708430	3	139015963	A/G	unknown	none	0.1965	7663	-0.017484	8.63E-06
11136787	8	4726407	G/T	intron	CSMD1(i)	0.3375	7663	0.0159343	9.19E-06
9386702	6	108486976	C/T	intron	OSTM1(i)	0.0479	7663	-0.0319514	9.88E-06

Table 8: Top Regions found to Interact with BMI (Method 2)

Index SNPs of the 11 regions found through a GWA analysis in the ARIC cohort to interact significantly (p-value < 10^{-5}) with BMI to influence hsCRP levels in this sample of European-Americans. SNPs with function listed as unknown have nearby gene(s) reported. Those with function listed as intron have the gene they are in indicated with an (i) and any nearby gene reported (if relevant).

3.4 Exploratory Aim 2

To identify genomic regions which interact with BMI to influence hsCRP levels in the subsample of European-Americans used in this thesis, the results from the two methods used in the first aim were compared. Two regions, defined by the index SNPs, on chromosome 2 and chromosome 1 were identified to be significant or suggestive by both methods. Table 9 contains information about these regions. The region containing the SNP on chromosome 2 was found near the genes FXO11 and FOXN2, and the region containing the SNP on chromosome 1 was found near the genes GJA5 and GJA8 (see figure 1 above to view regions).

The region on chromosome 2, found near the genes FXO11 and FOXN2, contained 18 SNPs (including the index SNP, rs12713007) which were all found to be moderately associated (p-values between $4.87*10^{-6}$ and $9.65*10^{-6}$) with hsCRP levels in lean individuals. Seven of these SNPs had negative β values falling around -0.1550. The other 11 had positive β values falling in a similar range. The β-values of these SNPs in obese individuals were not found to be significant, but were discovered to have opposite effects (negative instead of positive β-values and vice versa) with magnitudes between 0.04 and 0.05. The differences between the β-values in lean and obese individuals for each SNP were significantly different from zero (p-values between $1.57*10^{-6}$ and $2.00*10^{-6}$). The analysis performed on the un-stratified European-American subsample also showed each of these 18 SNPs to interact significantly (p-values between $7.88*10^{-6}$ and $5.11*10^{-6}$) with BMI. The β-values for each interaction were all found to have a magnitude of around 0.01 (7 SNPs with positive values and 11 with negative values). The 7 SNPs with positive β-values of interaction were the same SNPs with positive β-values in the association analysis using only

obese individuals. The index SNP, rs12568235, on chromosome 1 was in a region containing the genes GJA5 and GJA8.

Through comparison of the two methods it was the only SNP in this region found to interact with BMI. It was significantly associated with hsCRP levels in lean individuals with a β value of 2.28. The SNP's association with hsCRP levels in obese individuals was not significant, but had an opposite effect with a β value of -0.16. Comparison of these two β values was significant (p-value = $1.80*10^{-5}$). The un-stratified interaction analysis was also significant (p-value of $3.45*10^{-6}$) with a β value of -0.20.

name	chr	position	alleles	function	nearby genes	β_lean	p-value_lean	β_obese	p-value obese	t-stat	2-sided p-val	β_interaction	p-value_interaction
12713007	2	48337971	C/T	unknown	FBXO11, FOXN2	0.1589	4.8700E-06	-0.0418	0.1095	-4.999	5.779E-07	-0.0138	5.110E-06
12568235	1	145785795	C/T	unknown	GJA5, GJA8	2.2839	6.9100E-06	-0.1563	0.6526	-4.285	1.800E-05	-0.2032	3.450E-06

Table 9: Two Regions which Possibly Interact with BMI (Exploratory Aim 2)

Information about the 2 regions found to interact with BMI to influence hsCRP levels in European-Americans. They were identified through the comparison of two separate methods used to investigate the interaction. This table contains the β values and corresponding p-values from the stratified association analyses (in lean and obese individuals), the t-statistics and p-values from the comparison of the results from these association analyses, and the β values and corresponding p-values from the non-stratified interaction analysis.

CHAPTER 4: DISCUSSION

4.1 Identifying Genes that Possibly Interact with BMI to Influence hsCRP Levels

C-Reactive Protein is a biomarker for inflammation whose increased plasma levels are observed to accompany atherosclerotic coronary heart disease (16, 17, 18, 19, 20, 21, 23, 24, 25). It is produced in hepatocytes and secreted by the liver and CRP production is regulated by cytokines such as IL-6 and TNF-α (6, 8). In addition to being a biomarker of CHD, increased levels of CRP may also play a key role in the development of CHD (14, 15, 20). Furthermore, CRP levels have been observed to be correlated with body mass index (20, 34, 36), a common risk factor known to increase CHD risk (20). CRP also has a recognized genetic component. Multiple GWAS have identified multiple genetic polymorphisms which influence CRP levels. SNPs implicated in such studies have been found in or near genes of interest including: CRP, APOE, APOC, IL-6, HNF1A, LEPR, and GCKR (39, 40, 41, 42, 43). Because of the genetic influence on CRP levels in conjunction with the link between BMI and CRP levels, we undertook and presented the results of a series of analyses designed to identify loci which interact with BMI to influence CRP levels in a subsample of European-Americans in the ARIC cohort. By using and then synthesizing the results from two statistical methods used to investigate this interaction, we identified three genes, GJA5, GJA8 and FBXO11, which interact with BMI to influence hsCRP levels. Because there have been previous GWA studies of CRP with much larger sample sizes and because those genes with main effects on CRP have been adequately discussed, this discussion will limit itself to only those genes with suggestive or significant evidence for an interaction with BMI as they combine to influence of hsCRP levels.

Method 1: Stratified analysis

By comparing our stratified analyses, we found 16 regions (see table 7) to have significantly distinct effects (p-value < $2.00*10^{-3}$) on hsCRP levels between the two obesity strata (lean and obese). Fifteen of them were not previously reported in CRP GWA studies, and only one region, containing the GCKR gene, was previously reported as being associated with CRP levels in non-stratified CRP GWA studies (39, 43).

In addition to the region containing GCKR, regions on chromosome 1 containing the TGF-β2 gene, on chromosome 10 containing ITGB1 and NRP1, on chromosome 2 containing FBXO11, on chromosome 1 containing GJA5 and GJA8, and on chromosome 13 containing FLT1 were of particular interest to us, because of their known affiliation with CRP or the inflammatory process. TGF-β has been identified as a cytokine having a regulatory effect on CRP production (8). ITGB1 is an integrin which is a protein known to participate in the immune response (52). NRP1 has been suggested as playing a role in inflammation via the vascular endothelial growth factor receptor (VEGF) signaling pathway (53). The VEGF-A protein and TGF-β were both found to increase during atherogenesis (54). It has been reported that the vascular endothelial growth factor receptor 1, a tyrosine-kinase found in monocytes and encoded by FLT1, in its soluble form (sFLT1) is also positively correlated with VEGF. After binding to VEGF, it inhibits VEGF from binding to endothelial cells and, in turn, prevents cell migration. In the same study, VEGF was also found to be positively correlated with CRP levels (55). We will discuss FBXO11, GJA5, and GJA8 in more detail in the following sections of the discussion. The loci containing

TGF-β2, ITGB1 and NRP1, FBXO11, and GJA5 and GJA8 were all associated significantly with hsCRP levels specifically in the lean subsample, this suggests their effect is greatest in individuals with a lower BMI. FLT1 was associated was significantly associated with hsCRP levels in only in the obese subsample.

The genes in regions we found to have significantly different effects on hsCRP levels and which have no *a priori* obvious relationship to CRP include: FOXN2, KCNAB1, PTGR1, ZNF483, NRXN3, ZNF536, EPHA8, PCDH7, IFT172, MANEA, SLCO1B1, PDZRN4, and KLK5. Additional stratified analyses should be done to replicate or validate these findings in well-powered independent samples.

Method 2: Interaction analysis

Using the second method, a formal interaction modeling strategy, to accomplish the overall research objective, we observed 11 genomic regions (see table 8) to significantly interact (p-value < 10^{-5}) with BMI to influence hsCRP levels in the combined European-American sample. No SNP or gene reported as being associated with CRP levels in any published CRP GWAS was found to have an interactive effect with BMI on hsCRP levels in this thesis. Through literature searches we concluded that only two regions identified through this interaction analysis were biologically implicated in CRP production or regulation. These are the regions containing GJA5 and GJA8 (chromosome 1), and FBXO11 (chromosome 2). The genes in regions found to interact with BMI to influence hsCRP levels in this analysis which have no *a priori* obvious link to CRP are BDP1, LOC100132891, CDH9, FOXN2, NTM, CSMD1, and OSTM1.

Comparison of Methods

The two regions containing GJA5/GJA8 and FBXO11 provided consistent and statistically significant results using both methods employed in this thesis. Using the criterion of consistency between the two methods, these are the only two regions found to interact with BMI to influence hsCRP levels in this thesis. These statistically significant results do not necessarily mean that there is biologic interaction between the index SNPs (rs12713007 and rs12568235) and BMI causing a change in hsCRP levels. The results may simply imply that these SNPs are near or in linkage disequilibrium (LD) with the unmeasured or unknown causal variant or gene in their respective regions that interacts with BMI to influence hsCRP levels. For this reason, genes located near these two index SNPs were carefully considered and researched for functional relevance to CRP.

4.2 GJA8 and GJA5

In analytic strategies presented above, rs12568235 on chromosome 1 was the only SNP in a 800 kb region found to be statistically significant. The closest two genes, GJA5 and GJA8, located on either side of the SNP (see figure 1, region 13) were considered as loci having possible interactive effects with BMI on hsCRP levels.

GJA5 codes for the cardiac gap junction protein connexin 40 whose composition of intercellular channel arrays allows low molecular weight materials to move from cell to cell. It is expressed specifically in atrial myocytes, and its function is to mediate "the coordinated electrical activation of the atria" (56, 57). One study investigating heterozygous missense mutations in GJA5 concluded that such mutations may inhibit assembly of the gap junctions

or damage electrical coupling. These impairments of function could predispose individuals with such mutations to idiopathic atrial fibrillation (AF) (57). AF, which affects 2.3 million people in the US, is associated with increased risks for clot formation and stroke ((58). It has also been reported that traditional risk factors for CHD, such as CRP, are predictive of AF (59, 60). It is not known whether AF directly activates inflammation or if pre-existing inflammation and its accompanying elevated CRP levels promote AF (59, 61-63). Future studies should investigate the association between CRP and AF to speculate on which process is biologically most probable.

Although GJA8, a gene encoding for connexin 50 (64), has not been previously reported to be associated with CRP levels, it is known that variations in GJA8 have been associated with a number of cataract syndromes (65, 66) including congenital inherited cataracts (64). GJA8 is critical in the transport of "second messengers, metabolites, ions and water between the lens epithelium and the central lens nucleus" (67). Several studies have suggested and found that genetic polymorphisms involved in congenital cataracts, such as those in GJA8, also play a role in age-related cataracts (64, 68, 69, 70). Age-related cataract, a multifactorial disease making up 80% of all cataracts world-wide (71), is the principal source of visual impairment and blindness (72). In addition to having a genetic component, age-related cataracts are also environmentally influenced by high BMI (73) and inflammation (74, 75). Because of its association with inflammation, several studies investigated the relationship between age-related cataract and CRP levels (74, 76). Schaumberg et al studied CRP's relationship with the risk for future age-related cataract in 834 men and found an association in apparently healthy men (those with CRP levels ≥ 6.17 mg/l had a three-fold elevated risk of age-related cataract) (74).

The results of this thesis indicate that GJA5 and GJA8 possibly interact with BMI to influence hsCRP levels in European-Americans. In conjunction with the preceding information the results suggest that GJA5 and GJA8 may interact with BMI to influence hsCRP levels through the inflammatory processes involved in AF and age-related cataracts respectively.

4.3 FBXO11

rs12713007 and 18 other significantly or suggestive SNPs on chromosome 2 are found in a ~400 kb region located upstream of FBXO11 and downstream of FOXN2. FOXN2, is a transcriptional regulator of the T-cell leukemia virus long terminal repeat in humans (77), was found to have no obvious connection to CRP. In two separate studies, FBXO11 has been associated with a common childhood disease called otitis media (OM). OM is characterized by inflammation of the middle ear arising with infection (78, 79). Through use of *Jeff*, a mouse mutant model for OM, FBXO11 was identified as an underlying susceptibility gene (79, 80). In humans, the association between FBX011 and OM has been replicated in two separate studies of Western Australian children (78). In functional studies, FBXO11 was found to play a role in the regulation of the TGF-β pathway (79), possibly through neddylation and stabilization of p53, a cofactor of pospho-Smad2 (pSmad2)(81). Stabilization of p53 is mandatory to reduce accumulation of pSmad2 in the nucleus of epithelial cells (79, 82). The partnering of p53 and pSmad2 activates TGF- β target genes (82). The TGF-β pathway and release of the TGF- β protein during the acute stage of inflammation (83, 84) is known to play a regulatory role in cytokine-driven CRP

production (8). This information in addition to the results of this thesis suggests that FBXO11 interacts with BMI to influence hsCRP levels through the TGF-β pathway.

4.4 Conclusions and Future Studies

The primary objective of this thesis was to identify genes which interact with BMI to influence hsCRP levels in European-Americans in the ARIC cohort. Evidence for such an interaction with the genes GJA5, GJA8 and FBXO11 include: discovery through two separate statistical analysis strategies of rs12568235 and rs12713007 to interact significantly with BMI to affect hsCRP levels ; the close proximity of GJA5 and GJA8 to rs12568235 and of FBXO11 to rs12713007; and the biological link between mutations in these two genes and CRP levels. We speculate that atrial fibrillation (AF), age-related cataract and the TGF-β pathway may be biological processes influenced by the interaction of GJA5, GJA8 and FBXO11, respectively, with BMI to cause changes in hsCRP levels. If these results can be reproduced in other studies and populations, the next step could be to investigate in further detail how CRP levels and these biological processes affected each other. Future studies should focus on the influence of gene x bmi interaction on AF, age-related cataracts and the TGF-β pathway.

APPENDIX

Region	name (rs)	chr	position	alleles	function	nearby genes	MAF	n	β	p-value
	2272417	2	27560144	C/T	intron	IFT172(i), GCKR	0.4134	2083	-0.20259	2.67E-08
	2280737	2	27443314	C/T	intron	EIF2B4	0.3884	2083	-0.19276	7.43E-08
	13472	2	27453743	A/G	utr-3	ZNF513	0.3884	2083	-0.19235	7.71E-08
	12476704	2	27466535	A/C	intron	PPM1G	0.3882	2083	-0.19161	8.26E-08
	11689803	2	27420024	A/T	intron	GTF3C2	0.2866	2083	-0.18895	3.57E-06
	4425043	2	27586956	A/G	intron	GCKR	0.3877	2083	-0.18742	1.55E-07
	2293571	2	27582984	A/G	intron	GCKR	0.3863	2083	-0.18553	1.88E-07
	8395	2	27568711	A/T	utr-3	FNDC4	0.3877	2083	-0.18348	2.88E-07
	1528533	2	27449260	C/G	intron	SNX17	0.3922	2083	-0.18339	2.67E-07
	11681351	2	27596927	A/G	intron	GCKR	0.3966	2083	-0.18213	2.81E-07
	704791	2	27510671	C/T	intron	NRBP1	0.3911	2083	-0.1779	6.08E-07
	780102	2	27512995	C/T	intron	NRBP1	0.3911	2083	-0.17783	6.15E-07
	1260341	2	27516719	A/T	intron	NRBP1	0.3911	2083	-0.17775	6.22E-07
	780104	2	27531195	A/G	intron	IFT172	0.3911	2083	-0.17755	6.41E-07
	6760828	2	27432735	C/T	intron	GTF3C2	0.3931	2083	-0.17747	6.04E-07
	10205219	2	27422069	C/T	intron	GTF3C2	0.3931	2083	-0.17745	6.19E-07
	4665969	2	27428457	C/T	intron	GTF3C2	0.3931	2083	-0.17744	6.04E-07
	1647266	2	27546989	C/T	intron	IFT172	0.3912	2083	-0.17681	7.15E-07
	4665978	2	27502230	C/T	unknown	GCKR	0.4445	2083	-0.1741	8.66E-07
	704795	2	27569998	A/G	intron	FNDC4	0.3914	2083	-0.1727	1.24E-06
	813592	2	27575475	C/T	intron	GCKR	0.3911	2083	-0.17071	1.60E-06
	1260320	2	27575920	A/G	intron	GCKR	0.3911	2083	-0.17058	1.63E-06
	3739095	2	27410225	A/G	intron	GTF3C2	0.4311	2083	-0.17048	1.81E-06
	780110	2	27538892	A/G	intron	IFT172	0.4436	2083	-0.16945	1.68E-06
	780094	2	27594741	C/T	intron	GCKR	0.3995	2083	-0.1617	3.75E-06
	1260326	2	27584444	C/T	missense	GCKR	0.4144	2083	-0.16164	4.15E-06
	780093	2	27596107	C/T	intron	GCKR	0.3995	2083	-0.1615	3.85E-06
Region 1	2911711	2	27604050	A/T	unknown	GCKR	0.4514	2083	-0.15605	5.46E-06
	1260333	2	27602128	A/G	unknown	GCKR	0.4514	2083	0.156021	5.44E-06
	780117	2	27551847	C/G	intron	IFT172	0.3912	2083	0.176672	7.29E-07
	7586601	2	27438170	A/G	unknown	GCKR	0.444	2083	0.176832	6.20E-07
	1049817	2	27404471	A/G	coding-synonyn	GTF3C2	0.3929	2083	0.176879	7.33E-07
	11684134	2	27411756	A/G	intron	GTF3C2	0.4431	2083	0.176926	1.22E-06
	1647276	2	27542105	C/T	intron	IFT172	0.3912	2083	0.176929	7.03E-07
	780107	2	27538238	A/G	intron	IFT172	0.3912	2083	0.177222	6.73E-07
	780106	2	27535102	A/C	intron	IFT172	0.3912	2083	0.177305	6.65E-07
	6743819	2	27420911	G/T	intron	GTF3C2	0.393	2083	0.177431	6.29E-07
	4803	2	27520801	A/G	utr-3	IFT172	0.3911	2083	0.177625	6.34E-07
	1260342	2	27516920	G/T	intron	NRBP1	0.3911	2083	0.177709	6.26E-07
	780100	2	27505657	G/T	intron	NRBP1	0.3911	2083	0.178047	5.94E-07
	8179252	2	27600336	A/C	near-gene-3	GCKR	0.3967	2083	0.181606	3.04E-07
	6547626	2	27500274	C/T	unknown	GCKR	0.3921	2083	0.181718	3.29E-07
	1728922	2	27497968	A/C	unknown	GCKR	0.3921	2083	0.18177	3.26E-07
	4665976	2	27493829	A/G	unknown	GCKR	0.3921	2083	0.181875	3.19E-07
	1060525	2	27489086	A/G	unknown	GCKR	0.392	2083	0.182087	3.05E-07
	7563162	2	27484695	C/T	intron	PPM1G	0.392	2083	0.182102	3.04E-07
	2911712	2	27480449	A/T	intron	PPM1G	0.392	2083	0.182198	3.01E-07
	7594812	2	27464973	A/G	intron	PPM1G	0.3921	2083	0.18255	2.91E-07
	1647284	2	27461619	C/T	intron	PPM1G	0.3921	2083	0.182637	2.88E-07
	4582	2	27457783	A/G	utr-3	PPM1G	0.3921	2083	0.182972	2.79E-07
	2303369	2	27568920	C/T	intron	FNDC4	0.3877	2083	0.183346	2.93E-07
	2293572	2	27582281	C/G	intron	GCKR	0.3864	2083	0.185558	1.87E-07
	2010087	2	27490739	C/T	unknown	GCKR	0.3883	2083	0.191204	8.72E-08
	7602534	2	27445927	C/T	intron	EIF2B4	0.3884	2083	0.192667	7.49E-08
	1260345	2	27556999	A/G	intron	IFT172	0.4168	2083	0.195044	7.75E-08

Supplementary Table 1: Top 145 SNPs associated with hsCRP levels in European-American lean individuals. SNPs with function listed as unknown have nearby gene(s) reported. Those with function listed as intron have the gene they are in indicated with an (i) and any nearby gene reported (if relevant). The SNP highlighted in pink is the index SNP in its respective region.

Region	name (rs)	chr	position	alleles	function	nearby genes	MAF	n	β	p-value
	11265260	1	157966663	A/G	unknown	CRP	0.0601	2083	-0.36391	1.45E-07
	12081252	1	157973137	C/T	unknown	CRP	0.06	2083	0.364024	1.45E-07
	12081264	1	157973184	C/T	unknown	CRP	0.06	2083	0.364106	1.45E-07
	16842599	1	157964099	C/T	unknown	CRP	0.0601	2083	0.363607	1.47E-07
	12068753	1	157959161	A/T	unknown	CRP	0.0623	2083	0.359797	1.73E-07
	12755606	1	157936960	C/G	unknown	CRP	0.3316	2083	-0.18584	4.48E-07
	7553007	1	157965173	A/G	unknown	CRP	0.3286	2083	-0.18223	5.21E-07
	3093075	1	157946537	G/T	unknown	CRP	0.0621	2083	-0.34655	5.82E-07
	3093077	1	157946260	A/C	unknown	CRP	0.0621	2083	-0.34499	6.48E-07
	16842568	1	157942844	A/G	unknown	CRP	0.062	2083	-0.34497	6.57E-07
	1341665	1	157958183	A/G	unknown	CRP	0.3301	2083	-0.18076	6.57E-07
	16842559	1	157942795	C/T	unknown	CRP	0.062	2083	0.344904	6.64E-07
Region 2	1205	1	157948857	C/T	utr-3	CRP	0.3302	2083	0.180356	6.99E-07
	2808629	1	157943420	A/G	unknown	CRP	0.3304	2083	-0.17704	1.06E-06
	2794520	1	157945440	C/T	unknown	CRP	0.3305	2083	0.177009	1.07E-06
	2027471	1	157956012	A/T	unknown	CRP	0.3395	2083	-0.17798	1.08E-06
	2808628	1	157942635	A/G	unknown	CRP	0.3301	2083	-0.17764	1.12E-06
	16842502	1	157920487	A/C	unknown	CRP	0.0616	2083	0.344749	1.48E-06
	876537	1	157941557	C/T	unknown	CRP	0.3796	2083	0.164229	4.76E-06
	11265257	1	157935608	C/T	unknown	CRP	0.379	2083	0.164396	4.87E-06
	2808624	1	157932545	C/G	unknown	CRP	0.3789	2083	0.164392	4.89E-06
	12081480	1	157973614	G/T	unknown	CRP	0.0416	2083	0.427218	5.99E-06
	11588887	1	157983786	A/G	unknown	CRP	0.1269	2083	-0.26201	6.53E-06
	1470515	1	157920223	C/T	unknown	CRP	0.3778	2083	0.16381	7.02E-06
	2592887	1	157919563	C/T	unknown	CRP	0.3824	2083	0.164128	9.76E-06
	12463210	19	35788102	A/G	unknown	ZNF536	0.0075	2083	-0.99342	4.28E-07
	726644	19	35784264	C/T	unknown	ZNF536	0.0074	2083	-1.02787	4.29E-07
	12459020	19	35784794	C/G	unknown	ZNF536	0.0075	2083	-1.01997	4.29E-07
	12461477	19	35789462	A/G	unknown	ZNF536	0.0075	2083	0.990881	4.33E-07
	12461628	19	35789579	C/T	unknown	ZNF546	0.0075	2083	-0.94941	4.36E-07
Region 3	4805589	19	35809795	C/T	unknown	ZNF546	0.0079	2083	-0.86626	2.39E-06
	1529722	19	35808730	C/T	unknown	ZNF546	0.0079	2083	-0.86214	2.56E-06
	2043313	19	35805590	C/T	unknown	ZNF546	0.008	2083	-0.85144	2.99E-06
	7359955	19	35821299	A/G	unknown	ZNF546	0.0079	2083	-0.85747	3.17E-06
	7359965	19	35821309	A/G	unknown	ZNF546	0.0079	2083	0.856199	3.29E-06
	12461634	19	35821632	C/T	unknown	ZNF546	0.0079	2083	0.855365	3.37E-06
	2043311	19	35824238	C/T	unknown	ZNF546	0.0079	2083	0.85335	3.57E-06
	4448101	6	96095326	C/T	unknown	MANEA	0.3587	2083	-0.17635	6.90E-07
Region 4	2613562	6	95995138	C/G	unknown	MANEA	0.3441	2083	0.168897	3.39E-06
	2613565	6	96001120	A/C	unknown	MANEA	0.345	2083	0.166692	3.50E-06
	157582	19	50088059	C/T	intron	TOMM40(i), APOE/C	0.2291	2083	0.309804	7.59E-07
Region 5	4420638	19	50114786	A/G	intron	LOC100129500(i), APOE/C	0.1736	2083	0.228672	1.06E-06
	6857	19	50084094	C/T	utr-3	PVRL2, APOE/C	0.153	2083	0.30722	3.46E-06
Region 6	663002	1	22798200	A/C	intron	EPHA8	0.2174	2083	-0.20348	1.85E-06
Region 7	12260720	10	33415966	C/T	unknown	ITGB1, NRP1	0.2135	2083	-0.20125	4.21E-06

Supplementary Table 1 continued: Top 145 SNPs associated with hsCRP levels in European-American lean individuals. SNPs with function listed as unknown have nearby gene(s) reported. Those with function listed as intron have the gene they are in indicated with an (i) and any nearby gene reported (if relevant). The SNP highlighted in pink is the index SNP in its respective region.

Region	name (rs)	chr	position	alleles	function	nearby genes	MAF	n	β	p-value
	12713007	2	48337971	C/T	unknown	FBXO11, FOXN2	0.4905	2083	0.158899	4.87E-06
	963134	2	48162112	C/T	unknown	FBXO11, FOXN2	0.4583	2083	0.158877	5.47E-06
	6545018	2	48157885	C/T	unknown	FBXO11, FOXN2	0.4583	2083	0.158881	5.50E-06
	1483213	2	48168771	A/G	unknown	FBXO11, FOXN2	0.4593	2083	0.158553	5.82E-06
	10865228	2	48145389	C/T	unknown	FBXO11, FOXN2	0.4583	2083	-0.15751	7.11E-06
	2128720	2	48105539	G/T	unknown	FBXO11, FOXN2	0.4579	2083	0.155644	8.52E-06
	10193482	2	48105182	G/T	unknown	FBXO11, FOXN2	0.4579	2083	-0.15562	8.53E-06
	2128722	2	48106434	A/C	unknown	FBXO11, FOXN2	0.4581	2083	-0.15571	8.63E-06
Region 9	4525741	2	48106759	C/T	unknown	FBXO11, FOXN2	0.4581	2083	0.155699	8.69E-06
	4324359	2	48106772	A/G	unknown	FBXO11, FOXN2	0.4581	2083	0.155689	8.72E-06
	6545014	2	48108695	A/G	unknown	FBXO11, FOXN2	0.4581	2083	0.155662	8.78E-06
	10865227	2	48109258	A/T	unknown	FBXO11, FOXN2	0.4581	2083	-0.15564	8.83E-06
	6707641	2	48109337	A/G	unknown	FBXO11, FOXN2	0.4582	2083	-0.15552	9.02E-06
	10176365	2	48116260	A/G	unknown	FBXO11, FOXN2	0.4582	2083	0.155343	9.25E-06
	2056138	2	48120680	G/T	unknown	FBXO11, FOXN2	0.4582	2083	0.15531	9.30E-06
	6749276	2	48120803	C/T	unknown	FBXO11, FOXN2	0.4582	2083	-0.15529	9.32E-06
	7584493	2	48121249	A/G	unknown	FBXO11, FOXN2	0.4582	2083	-0.15518	9.45E-06
	7584401	2	48121348	C/T	unknown	FBXO11, FOXN2	0.4583	2083	0.155005	9.65E-06
Region 10	10841753	12	21212637	C/T	intron	SLCO1B1	0.188	2083	0.199965	5.02E-06
	4335431	1	216624232	C/T	intron	TGFB2	0.0842	2083	-0.27746	5.87E-06
Region 11	11466381	1	216625068	A/G	intron	TGFB2	0.0843	2083	0.277399	6.06E-06
	12058490	1	216625903	A/G	intron	TGFB2	0.0843	2083	0.278127	6.28E-06
	17052680	6	117407601	A/C	unknown	RFX6	0.2335	2083	-0.1814	6.87E-06
	13194910	6	117414024	C/G	unknown	RFX6	0.2938	2083	-0.16642	8.46E-06
	13195177	6	117428076	A/T	unknown	RFX6	0.294	2083	-0.1665	8.69E-06
	874403	6	117418448	C/T	unknown	RFX6	0.2938	2083	-0.16613	8.74E-06
	13197353	6	117419910	A/T	unknown	RFX6	0.2938	2083	0.166078	8.78E-06
	2883129	6	117420799	A/G	unknown	RFX6	0.2938	2083	0.166042	8.81E-06
	2353281	6	117420808	G/T	unknown	RFX6	0.2938	2083	-0.16603	8.82E-06
	11961055	6	117421413	C/T	unknown	RFX6	0.2938	2083	-0.16599	8.85E-06
Region 12	13197405	6	117419723	G/T	unknown	RFX6	0.2898	2083	0.167749	8.92E-06
	1334682	6	117422761	A/G	unknown	RFX6	0.2938	2083	0.165883	8.98E-06
	10485192	6	117424688	C/T	unknown	RFX6	0.2938	2083	0.165677	9.22E-06
	4946207	6	117396458	C/G	unknown	RFX6	0.2944	2083	-0.16866	9.27E-06
	1321363	6	117409336	A/G	unknown	RFX6	0.2905	2083	-0.16642	9.66E-06
	11961334	6	117425962	C/G	unknown	RFX6	0.2939	2083	0.165285	9.66E-06
	10485193	6	117426440	G/T	unknown	RFX6	0.2939	2083	-0.16528	9.67E-06
	17078735	6	117430029	C/T	unknown	RFX6	0.2352	2083	0.178277	9.84E-06
	13213573	6	117427156	G/T	unknown	RFX6	0.2942	2083	-0.16522	9.89E-06
	11968834	6	117427854	A/G	unknown	RFX6	0.2943	2083	-0.1652	9.94E-06
Region 13	12568235	1	145785795	C/T	unknown	GJA5, GJA8	0.0092	2083	2.28392	6.91E-06
Region 14	7026971	9	113371319	A/T	intron	PTGR1,ZNF483	0.0427	2083	-0.36723	7.13E-06
Region 15	13093806	3	157680330	A/C	intron	KCNAB1	0.3751	2083	0.16719	7.86E-06

Supplementary Table 1 continued: Top 145 SNPs associated with hsCRP levels in European-American lean individuals. SNPs with function listed as unknown have nearby gene(s) reported. Those with function listed as intron have the gene they are in indicated with an (i) and any nearby gene reported (if relevant). The SNP highlighted in pink is the index SNP in its respective region.

Region	name (rs)	chr	position	alleles	function	nearby genes	MAF	n	β	p-value
Region 1	4420638	19	50114786	A/G	intron	LOC100129500 (i), APOE	0.1736	2411	0.215498	1.86E-09
	6857	19	50084094	C/T	utr-3	PVRL2, APOE/C	0.153	2411	0.290624	1.31E-08
	2075650	19	50087459	A/G	intron	TOMM40 (i), APOE/C	0.1389	2411	0.28007	3.84E-07
	439401	19	50106291	C/T	intron	LOC100129500 (i), APOE	0.3867	2411	-0.260541	4.71E-07
Region 2	12068753	1	157959161	A/T	unknown	CRP	0.0623	2411	0.271131	8.02E-07
	2808628	1	157942635	A/G	unknown	CRP	0.3301	2411	-0.135287	1.18E-06
	2794520	1	157945440	C/T	unknown	CRP	0.3305	2411	0.134296	1.23E-06
	2808629	1	157943420	A/G	unknown	CRP	0.3304	2411	-0.134262	1.24E-06
	1205	1	157948857	C/T	utr-3	CRP	0.3302	2411	0.134366	1.27E-06
	1341665	1	157958183	A/G	unknown	CRP	0.3301	2411	-0.134052	1.32E-06
	2027471	1	157956012	A/T	unknown	CRP	0.3395	2411	-0.13339	1.65E-06
	3093075	1	157946537	G/T	unknown	CRP	0.0621	2411	-0.262416	1.96E-06
	16842502	1	157920487	A/C	unknown	CRP	0.0616	2411	0.270112	2.01E-06
	7553007	1	157965173	A/G	unknown	CRP	0.3286	2411	-0.131546	2.04E-06
	16842559	1	157942795	C/T	unknown	CRP	0.062	2411	0.261416	2.13E-06
	16842568	1	157942844	A/G	unknown	CRP	0.062	2411	-0.261345	2.13E-06
	3093077	1	157946260	A/C	unknown	CRP	0.0621	2411	-0.261219	2.14E-06
	2592887	1	157919563	C/T	unknown	CRP	0.3824	2411	0.134288	3.02E-06
	12081480	1	157973614	G/T	unknown	CRP	0.0416	2411	0.340329	5.27E-06
	1470515	1	157920223	C/T	unknown	CRP	0.3778	2411	0.127779	5.90E-06
	12081264	1	157973184	C/T	unknown	CRP	0.06	2411	0.250863	7.02E-06
	12081252	1	157973137	C/T	unknown	CRP	0.06	2411	0.250694	7.06E-06
	11265260	1	157966663	A/G	unknown	CRP	0.0601	2411	-0.250438	7.15E-06
	16842599	1	157964099	C/T	unknown	CRP	0.0601	2411	0.250396	7.17E-06
	2808624	1	157932545	C/G	unknown	CRP	0.3789	2411	0.124762	7.54E-06
	11265257	1	157935608	C/T	unknown	CRP	0.379	2411	0.124538	7.79E-06
	876537	1	157941557	C/T	unknown	CRP	0.3796	2411	0.122747	9.95E-06
Region3	13330968	16	79898765	G/T	unknown	BCMO1, GAN	0.0141	2411	-1.57314	1.49E-06
Region 4	4278077	7	33356392	C/G	intron	BBS9 (i)	0.0158	2411	0.442459	6.19E-06
Region 5	2808898	9	43432012	C/T	unknown	FAM75A6	0.1265	2411	0.976443	6.70E-06
Region 6	5008004	12	40283747	C/T	unknown	PDZRN4	0.3251	2411	0.128582	8.49E-06
Region 7	17086609	13	27827711	A/G	intron	FLT1 (i)	0.2992	2411	-0.153033	8.56E-06
Region 8	629882	11	125747907	C/T	intron	ST3GAL4 (i)	0.1278	2411	-0.17358	8.89E-06
Region 9	4903731	14	77765959	A/G	unknown	NRXN3	0.0743	2411	0.220079	9.16E-06
Region 10	10408037	19	56133595	A/G	unknown	KLK5	0.009	2411	-1.08014	9.94E-06
Region 11	7019284	9	28384358	G/T	intron	LINGO2 (i)	0.3167	2411	0.126009	9.95E-06

Supplementary Table 2: Top 36 SNPs associated with hsCRP levels in European-American lean individuals. SNPs with function listed as unknown have nearby gene(s) reported. Those with function listed as intron have the gene they are in indicated with an (i) and any nearby gene reported (if relevant). The SNP highlighted in pink is the index SNP in its respective region.

REFERENCES

1. Pepys MB, and Baltz ML. Acute phase proteins with special reference to C-reactive protein and related proteins (pentaxins) and serum amyloid A protein. *Advances in Immunology.* 1983;34:141-212.
2. Pepys, Mark B, and Hirschfield, Gideon M. C-reactive protein: a critical update. *The Journal of Clinical Investigation.* 2003;111(12):1805-1812.
3. Clyne, Brian; Jonathan S. Olshaker. The C-reactive protein. *Journal of Emergency Medicine.* 1999;17(6):1019–25.
4. hs-CRP. CRPhealth.com [online]. Available: http://www.crphealth.com/information/hs-crp.html. [May 15, 2011].
5. Vigushin DM, Pepys MB, Hawkins PN. Metabolic and scintigraphic studies of reioiodinated human C-reactive protein in health and disease. *J. Clin. Invest.* 2003;91:1351-1357.
6. Black S, Kishner I, Samols D. C-Reactive Protein. *The Journal of Biological Chemistry.* 2004;279(47):48487-48490.
7. Cai G, Cole SA, Butte NF, Smith CW, Mehta NR, Voruganti VS, Proffitt JM, and Comuzzie AG. A genetic contribution to circulating cytokines and obesity in children. *Cytokine.* 2008;44:242-247.
8. Taylor AW, KU N, Mortensen RF. Regulation of cytokine-induced human C-reactive protein production by transforming growth factor-β. *Journal of Immunology.* 1990;145(8):2507-2513.

9. Chen M. (June 21, 2010). Coronary heart disease. PubMed Health. [online]. Available: http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0004449/. [May 15, 2011].
10. Libby P, Ridker PM, Maseri A. Inflammation and Atherosclerosis. *Circulation.* 2002;105:1135-1143.
11. Kovanen PT and Pentikäinen. Pharmacologic prevention of coronary plaque rupture – the major cause of acute coronary syndromes. *Heart Metab.* 2007;36:9-14.
12. Virmani R, Burke AP, Farb A, Kolodgie FD. Pathology of the vulnerable plaque. J Am Coll Cardiol. 2006;47 (suppl 8):C13–C18.
13. (May 2011). Atherosclerosis. National Heart Lung and Blood Institute: Diseases and Conditions Index. [Online]. Available: http://www.nhlbi.nih.gov/health/dci/Diseases/Atherosclerosis/Atherosclerosis_WhatIs.html. [May 15, 2011].
14. Paffen E., de Maat MPM. C-reactive protein in atherosclerosis: A causal factor? *Cardiovascular Research.* 2006;71:30-39.
15. Calabrô P, Golia E, Yeh ETH. CRP and the risk of atherosclerotic events. *Seminars in Immunopathology.* 2009;31:79-94.
16. Kuller LH, Tracy RP, Shaten J, and Meilahn EN. Relation of C-reactive protein and coronary heart-disease in the MRFIT nested case control study. *Am. J. Epidemiology.* 1996;144:537-547.
17. Tracy RP, Lemaitre RN, Psaty BM, Ives DG, Evans RW, Cushman M, Meilahn EN, and Kuller LH. Relationship of C-Reactive protein to risk of cardiovascular disease

in the elderly. Results from the Cardiovascular Health Study and the Rural Health Promotion Project. *Arterioscler Thromb Vasc Biol.* 1997;17(6):1121-1127.

18. Ridker PM, Cushman M, Stampfer MJ, Tracy RP, and Hennekens CH. Inflammation, aspirin, and the risk of cardiovascular disease in apparently healthy men. *N. Engl. J. Med.* 1997;336:973-979.
19. Koenig W, Sund M, Fröhlich M, Fischer HG, Löwel H, Döring A, Hutchinson WL, Pepys MB. C-reactive protein, a sensitive marker of inflammation, predicts future risk of coronary heart disease in initially healthy middle-aged men: results for the MONICA (Monitoring Trends and Determinants in Cardiovascular Disease) Augsburg Cohort Study, 1984 to 1992. *Circulation.* 1999;99:237-242.
20. Danesh J, Whincup P, Walker M, Lennon L, Thomson A, Appleby P, Gallimore JR, and Pepys MB. Low grade inflammation and coronary heart disease: prospective study and updated meta-analyses. *BMJ.* 2000;321:199-204.
21. Ridker PM, Hennekens CH, Buring JE, and Rifai N. C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. *N. Engl. J. Med.* 2000;836-843.
22. Makover M. (Feb. 10, 2010). C-Reactive Protein. US National Library of Medicine: National Institutes of Health. [Online]. Available: http://www.nlm.nih.gov/medlineplus /ency/article/003356.htm. [May 15, 2011].
23. Ridker PM, Buring JE, Shih J, Matias M, and Hennekens CH. Prospective study of C-reactive protein and the risk of future cardiovascular events among apparently healthy women. *Circulation.* 1998;98:731-733.

24. Ridker PM, Cushman M, Stampfer MJ, Tracy RP, and Hennekens CH. Plasma concentration to C-reactive protein and risk of developing peripheral vascular disease. *Circulation.* 1998;97:425-428.
25. Liuzzo G, Biasucci LM, Gallimore JR, Grillo RL, Rebuzzi AG, Pepys MB, and Maseri A. The prognostic value of C-reactive protein and serum amyloid A protein in severe unstable angina. *N. Enlg. J. Med.* 1994;331:417-424.
26. Anzai T, Yoshikawa T, Shiraki H, Asakura Y, Akaishi M, Mitamura H, and Ogawa S. C-reactive protein as a predictor of infarct expansion and cardiac rupture after a first Q-wave acute myocardial infarction. *Circulation.* 1997;96:778-784.
27. Puska P, Nishida C., Porter D. Obesity and Overweight. *WHO*. 2003.
28. Flegal KM, Carroll MD, Ogden CL. Prevalence and Trends in Obesity Among US Adults, 1999-2008. *JAMA*. 2010;303(3):235-241.
29. Trasande L, Cronk C, Durkin M, Weiss M, Schoeller D, Gall E, Hewitt J, Carrel A, Landrigan P, and Gillman M. Environment and Obesity in the National Children's Study. *Ciencia & Saude Coletiva*. 2010;15(1):195-210.
30. Van Cleave J, Gortmaker SL, Perrin JM. Dynamics of Obesity and Chronic Health Conditions Among Children and Youth. *JAMA*. 2010;303(7):623-630.
31. Bouchard C. Childhood obesity: are genetic differences involved? *Am J Clin Nutr*. 2009; 89(suppl):1494S-501S.
32. Farhat T, Iannotti RJ, Simons-Morton BG. Overweight, Obesity, Youth, and Health-Risk Behaviors. *Am J Prev Med*. 2010;38(3):258-267.

33. Knox S, Barnes A, Kiefe C, Lewis C, Irabarren C, Matthews K, Wong N, Whooley M. History of Depression, Race and Cardiovascular Risk in CARDIA. *International Journal of Behavioral Medicine*. 2006;13(1):44-50.

34. Visser M, Bouter LM, McQuillan GM, Wener MH, and Harris TB. Elevated C-Reactive Protein Levels in Overweight and Obese Adults. *JAMA.* December 8, 2009;282(22):2131-2135.

35. Bochud M, Marquant F, Marques-Vidal PM, Vollenweider P, Beckmann JS, Mooser V, Paccaud F, and Rousson V. Association between C-Reactive Protein and Adiposity in Women. *J. Clin. Endocrinol. Metab.* October 2009;94(10):3969-3977.

36. Eiriksdottier G, Smith AV, Aspelund T, Hafteinsdottir SH, Olafsdottir E, Launer LJ, Harris TB and Gudnason V. The interaction of adiposity with the CRP gene affects CRP levels: Age, Gene/Environment Susceptibility-Reykjavik Study. *International Journal of Obesity.* 2009;33:267-272.

37. Lange LA, Burdon K, Langefeld CD, Liu Y, Beck SR, Rich SS, Freedman BI, Brosnihan KB, Herrington DM, Wagenknecht LE, and Bowden DW. Heritability and Expression of C-Reactive Protein in Type 2 Diabetes in the Diabetes Heart Study. *Annals of Human Genetics.* 2006;70:717-725.

38. (March 28, 2011). A Catalog of Published Genome-Wide Association Studies: C-Reactive Protein. National Human Genome Research Institute: National Institutes of Health. [Online]. Available: http://www.genome.gov/GWAStudies/index.cfm?pageid=26525384#searchForm. [May 15, 2011].

39. Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Lu C, Pellikka N, Wallaschofski K, Kettunen J, Henneman P, Baumert J, Strachan DP, Fuchsberger C, Vitart V, Wilson JF, Paré G, Naitza S, Rudock ME, Surakka I, de Geus EJ, Alizadeh BZ, Guralnik K, Shuldiner A, Tanaka T, Zee RY, Schnabel RB, Nambi V, Kavousi M, Ripatti S, Nauck M, Smith NL, Smith AV, Sundvall J, Scheet P, Liu Y, Ruokonen A, Rose LM, Larson MG, Hoogeveen RC, Freimer NB, Teumer A, Tracy RP, Launer LJ, Buring JE, Yamamoto JF, Folsom AR, Sijbrands EJ, Pankow J, Elliott P, Keaney JF, Sun W, Sarin AP, Fontes JD, Badola S, Astor BC, Hofman A, Pouta A, Werdan K, Greiser KH, Kuss O, Meyer ZU, Schwabedissen HE, Theiry J, Jamshidi Y, Nolte IM, Soranzo N, Spector TD, Völzke H, Parker AN, Aspelund T, Bates D, Young L, Tsui K, Siscovick DS, Guo X, Rotter JI, Uda M, Schlessinger D, Rudan I, Hicks AA, Penninx BW, Thorand B, Gieger C, Coresh J, Willemsem G, Harris TB, Uitterlinden AG, Järvelin MR, Rice K, Radke D, Salomaa V, Willems can Dijk K, Boerwinkle E, Vasan RS, Ferrucci L, Gibson QD, Bandinelli S, Snieder H, Boomsma DI, Xiao X, Campbell H, Hayward C, Pramstaller PP, van Duijn CM, Peltonen L, Psaty BM, Gudnason V, Ridker PM, Homuth G, Koenig W, Ballantyne CM, Witteman JC, Benjamin EJ, Perola M, and Chasman DI. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation.* 2011 Feb 22;123(7):731-8.

40. Okada Y, Takahashi A, Ohmiya H, Kumasaka N, Kamatani Y, Hosono N, Tsunoda T, Matsuda K, Tanaka T, Kubo M, Nakamura Y, Yamamoto K, Kamatani N. Genome-wide association study for C-reactive protein levels identified pleiotropic

associations in the IL6 locus. *Human Molecular Genetics.* 2011 Mar 15;20(6):1224-31.

41. Elliott P, Chambers JC, Zhang W, Clarke R, Hopewell JC, Peden JF, Erdmann J, Braund P, Engert JC, Bennett D, Coin L, Ashby D, Tzoulaki I, Brown IJ, Mt-Isa S, McCarthy MI, Peltonen L, Freimer NB, Farrall M, Ruokonen A, Hamsten A, Lim N, Froquel P, Waterworth DM, Vollenweider P, Waeber G, Jarvelin MR, Mooser V, Scott J, Hall AS, Schunkert H, Anand SS, Collins R, Samani NJ, Watkins H, and Kooner JS. Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA.* 2009 Jul 1;302(1):37-48.
42. Reiner AP, Barber MJ, Guan Y, Ridker PM, Lange LA, Chasman DI, Walston JD, Cooper GM, Jenny NS, Rieder MJ, Durda JP, Smith JD, Novembre J, Tracy RP, Rotter JI, Stephens M, Nickerson DA, Krauss RM. Polymorphisms of the HNF1A gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. *American Journal of Human Genetics.* 2008 May;82(5):1193-201.
43. Ridker PM, Pare G, Parker A, See RY, Danik JS, Buring JE, Kwiatkowski D, Cook NR, Miletich JP, Chasman DI. Loci related to metabolic-syndrome pathways including LEPR, HNF1A, IL6R, and GCKR associate with plasma C-reactive protein: The Women's Genome Health Study. *Am. J. Hum. Genet.* 2008;82:1185-1192.
44. Mendoza-Carrera F, Ramírez-López G, Ayala-Martínez NA, García-Zapién AG, Flores-Martínez SE, and Sánchez-Corona J. Influence of CRP, IL6, and TNFA Gene Polymorphisms on Circulating Levels of C-Reactive Protein in Mexican Adolescents. *Archives of Medical Research.* 2010;41:472-477.

45. Benjamin EJ, Dupuis J, Larson MG, Lunetta KL, Booth SL, Govindaraju DR, Kathiresan S, Keaney JF Jr, Keyes MJ, Lin LP, Meigs JB, Robins SJ, Rong J, Schnabel R, Vita JA, Wang TJ, Wilson PW, Wolf PA, and Vasan RS. Genome-wide association with select biomarker traits in the Framingham Heart Study. *BMC Medical Genetics.* 2007;8(Suppl 1):S11.

46. Davignon J, Cohn JS, Mabile L, Bernier L. Apolipoprotein E and atherosclerosis: insight from animal and human studies. *Clin Chim Acta.* 1999;286:115-143.

47. Moghadasian MG, McManus BM, Nguyen LB, Shefer S, Nadji M, Godin DV, Green TJ, Hill J, Yang Y, Scudamore CH, and Frohlich JJ. Pathophysiology of apolipoprotein E deficiency in mice: relevance to apo E-related disorders in humans. *FASEB J.* 2001;15:2623-2630.

48. Stannard AK, Riddell DR, Sacre SM, Tagalakis AD, Langer C, von Eckardstein A, Cullen P, Athanasopoulos T, Dickson G, and Owen JS. Cell-derived apolipoprotein E (ApoE) particles inhibit vascular call adhesion molecule-1 (VCAM-1) expression in human endothelial cells. *J Biol Chem.* 2001;276:46011-46016.

49. Zhang YY, Gottardo L, Mlynarski W, Frazier W, Nolan D, Duffy J, Marescotti MC, Gervino EV, Johnstone MT, Mantzoros CS, Avogaro A, Doria A. Genetic variability at the leptin receptor (LEPR) locus is a determinant of plasma fibrinogen and C-reactive protein levels. *Atherosclerosis.* 2007;191:121-127.

50. Chen K, Li F, Li J, Cai H, Strom S, Bisello A, Kelley DE, Friedman-Einat M, Skibinski GA, McCrory MA, Szalai AJ, and Zhao AZ. Induction of leptin resistance through direct interaction of C-reactive protein with leptin. *Nat Med.* 2006;12:425-432.

51. de Bakker PIW, Ferreira MAR, Jia X, Neale BM, Raychaudhuri S, and Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Human Molecular Genetics.* 2008;17(R2):R122-E128.

52. (July 3, 2011). ITGB1 integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12) [*Homo sapiens*]. NCBI: Gene. [Online]. Available: http://www.ncbi.nlm.nih.gov/gene/3688. [July 12, 2011].

53. Saban MR, Sferra TJ, Davis CA, Simpson C, Allen A, Maier J, Fowler B, Knowlton N, Birder L, Wu X, and Saban R. Neuropilin – VEGF signaling pathway acts as a key modulator of vascular, lymphatic, and inflammatory cell responses of the bladder to intravesical BCG treatment. Am J Physiol Renal Physiol. 2010 Dec;299(6):F1245-56. Epub 2010 Sep 22.

54. Panutsopulos D, Papalambros E, Sigala F, Zafiropoulos A, Arvanitis DL, and Spandidos DA. Protein and mRNA expression levels of VEGF-A and TGF-β1 in different types of human coronary atherosclerotic lesions. *Int J Mol Med.* 2005 Apr;15(4):603-610.

55. Sandhofer A, Tatarczyk T, Kirchmair R, Iglseder B, Paulweber B, Patsch JR, and Schratzberger P. Are plasma VEGF and its soluble receptor sFlt-1 atherogenic risk factors? Cross-sectional data from the SAPHIR study. *Atherosclerosis.* 2009;206:265-269.

56. (June 5, 2011). GJA5 gap junction protein, alpha 5, 40kDa [*Homo sapiens*]. NCBI: Gene. [Online]. Available: http://www.ncbi.nlm.nih.gov/gene/2702. [July 12, 2011].

57. Gollob MH, Jones DL, Krahn AD, Danic L, Gong X, Shao Q, Liu X, Veinot JP, Tang ASL, Stewart AFR, Tesson F, Klein GJ, Yee R, Skanes AC, Guiraudon GM, Ebihara L, and Bai D. Somatic Mutations in the Connexin 40 Gene (GJA5) in Atrial Fibrillation. *The New England Journal of Medicine.* 2006 June;354(25):2677-2688.

58. Go AS, Hylek EM, Phillips KA, Chang Y, Henault LE, Selby JV, and Singer DE. Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. *JAMA.* 2001;285:2370-2375.

59. Aviles RJ, Martin DO, Apperson-Hansen C, Houghtaling PL, Rautaharju P, Kronmal RA, Tracy RP, Von Wagoner DR, Psaty BM, Lauer MS, Chung MK. Inflammation as a Risk Factor for Atrial Fibrillation. *Circulation.* 2003;108:3006-3010.

60. Smith GJ, Newton-Cheh C, Almgren P, Struck J, Morgenthaler NG, Bergmann A, Platonov PG, Hedblad B, Engstrom G, Wang TJ, Melander O. Assessment of Conventional Cardiovascular Risk Factors and Multiple Biomarkers for Prediction of Incident Heart Failure and Atrial Fibrillation. *JACC.* 2010 Nov.;56(21):1712-1719.

61. Mevorach D. Opsonization of apoptotic cells: implications for uptake and autoimmunity. *Ann N Y Acad Sci.* 2000;926:226-235.

62. Allessie M, Ausma J, Schotten U. Electrical, contractile and structural remodeling during atrial fibrillation. *Cardiovascular Res.* 2002;54:230-246.

63. Mihm MJ, Yu F, Carnes CA, Reiser PJ, McCarthy PM, Van Wagoner DR, and Bauer JA. Impaired myofibrillar energetic and oxidative injury during human atrial fibrillation. *Circulation.* 2001;104:174-180.

64. (July 2, 2011). GJA8 gap junction protein, alpha 8, 50kDa [*Homo sapiens*]. NCBI: Gene. [Online]. Available: http://www.ncbi.nlm.nih.gov/gene/2703. [July 12, 2011].

65. Liu Y, Ke M, Yan M, Guo S, Mothobi ME, Chen Q and Zheng F. Association between gap junction protein-alpha 8 polymorphisms and age-related cataract. *Mol Biol Rep.* 2011;38:1301-1307.

66. Arora A, Minogue PH, Liu X, Reddy MA, Ainsworth JR, Bhattacharya SS, Webster AR, Hunt DM, Ebihara L, Moore AT, Beyer EC, and Merthoud VM. A novel *GJA8* mutation is associated with autosomal dominant lamellar pulverulent cataract: further evidence for gap junction dysfunction in human cataract. *J Med Genet.* 2006;43:e2.

67. Gong X, Cheng C, and Xia CH. Connexins in lens development and cataractogenesis. *J Membr Bio.* 2007;218:9-12.

68. Shiels A, Hejtmancik JF. Genetic origins of cataract. *Arch Ophthalmol.* 2007;125:165-173.

69. Iyengar SK, Klein BE, Klein R., Jun G, Schick JH, Millard C, Liptak R, Russo K, Lee KE, and Elston RC. Identification of a major locus for age-related cortical cataract on chromosome 6p12-q12 in the Beaver Dam Eye Study. *Proc Natl Acad Sci USA.* 2004;101:14485-14490.

70. Zhou Z, Wang B, Hy S, Zhang C, Ma X, and Qi Y. Genetic variations in *GJA3, GJA8*, *LIM2*, and age-related cataract in the Chinese population: a mutation screening study. *Molecular Vision.* 2011;17:621-626.

71. Foster A, Resnikoff S. The impact of Vision 2020 on global blindness. *Eye.* 2005;19:1133-1135.

72. Thylefors B, Negrel AD, Pararajasegaram A, Dadzie KY. Global data on blindness. *Bull World Health Organ.* 1995;73:115-121.

73. Glynn RY, Christen WG, Manson JE, Bernheimer J, Hennekens CH. Body mass index: an independent predictor of cataract. *Arch Ophthalmol.* 1995;1131-1137.

74. Schaumberg DA, Ridker PM, Glynn RJ, Christen WG, Dana MR, Hennekens CH. High levels of plasma C-reactive protein and future risk of age-related cataract. *Ann Epidemiol.* 1999;9:166-171.

75. Klein BE, Klein R, Lee KE, Knudtson MD, Tsai MY. Markers of inflammation, vascular endothelial dysfunction, and age-related cataract. *Am J Ophthalmol.* 2006;141:116-122.

76. Boey PY, Tay WT, Lamoureux E, Tai ES, Mitchell P, Wang JJ, Saw SM, and Wong TY. C-Reactive Protein and Age-Related Macular Degeneration and Cataract: The Singapore Malay Eye Study. *Investigative Ophthalmology & Visual Science.* 2010 April;51(4):1880-1885.

77. (June 14, 2011). FOXN2 forkhead box N2 [*Homo sapiens*]. NCBI: Gene. [Online]. Available: http://www.ncbi.nlm.nih.gov/gene/3344. [July 12, 2011].

78. Rye MS, Wiertsema Sp, Scaman ESH, Oommen J, Sun W, Francis RW, Ang W, Pennell CE, Burgner D, Richmond P, Vijayasekaran S, Coates HL, Brown SD, Blackwell JM and Jamieson SE. FBXO11, a regulator of the TGFβ pathway, is associated with severe otitis media in Western Australian children. *Genes and Immunity.* 2011:1-8.

79. Tateossian K, Hardisty-Hughes RE, Morse S, Romero MR, Hilton H, Dean C and Brown SDM. Regulation of TGF-β signaling by *Fbxo11*, the gene mutated in the *Jeff* otitis media mouse mutant. *Pathogenetics.* 2009;2:5-19.

80. Hardisty-Hughes RS, Tateossian H, Morse SA, Romero MR, Middleton A, Tymowska-Lalanne Z, Hunter AJ, Cheeseman M, and Brown SD. A mutation in the F-Box gene, Fbxo11, causes otitis media in the Jeff mouse. *Hum Mol Genet.* 2006;15:3273-3279.

81. Abida MW, Nikolaev A, Zhao W, Zhang W, Gu W. FBXO11 promotes the neddylation of p53 and inhibits its transcriptional activity. *J Biol Chem.* 2007;282:1797-1804.

82. Cordenonsi M, Dupont S, Maretto S, Insinga A, Imbriano C, Piccolo S. Links between tumor suppressors: p53 is required for TGF-beta gene responses by cooperating with Smads. *Cell.* 2003;113:301-314.

83. Roberts AB, and Sporn MB. Transforming growth factor-β. *Adv. Cancer Res.* 1988;51:107.

84. Wahl SM, McCartney-Francis MN, and Mergenhagen SE. Inflammatory and immunomodulatory roles of TGF-β. *Immunol. Today.* 1989;10:258.

VITA

Sarah Elizabeth Tudor was born in Englewood, Colorado on the 3rd of January, 1986 to Beatrice Peek and Thomas Tudor. After completion of her work at Air Academy High School in Colorado Springs, Colorado, she attended the Colorado School of Mines in Golden, Colorado. She earned a degree of Bachelor of Science with a major in Chemistry and a minor in Bioengineering and Life Sciences from the Colorado School of Mines in May of 2009. In August of 2009, she entered the University of Texas Health Science Center at Houston Graduate School of Biomedical Sciences. She will graduate with a degree of Masters of Science in Biomedical Sciences with an emphasis in Human and Molecular Genetics in the summer of 2011. She will be attending the CU Denver Colorado School of Public Health in Aurora, Colorado this fall with plans of earning a Masters of Science degree in Epidemiology.

Permanent Address:

6363 Kremmling Circle

Colorado Springs, CO 80919