The Genetics of Blood Pressure and Body Mass Index: The Tale of two Major Risk Factors

by

Chun-Yi Wu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Epidemiological Science)
in The University of Michigan
2013

Doctoral Committee:

Professor Sharon Reilly Kardia, Chair
Professor Ana V. Diez Roux
Associate Professor Bhramar Mukherjee
Professor Patricia A. Peyser
Assistant Research Scientist Jennifer Smith

Acknowledgements

First of all, I would like to thank my mentor, Sharon Kardia, for her constant guidance and support since my masters program at the University of Michigan. When I had difficulty finding a student internship, Sharon introduced me to Dr. Jacquelyn Taylor who helped me get experience conducting epidemiological research. After I graduated from the MPH program, Sharon hired me as a research assistant and then helped me to get admitted to the Ph.D. program in epidemiology. The process of Ph.D. training was not easy and sometimes when I felt lost and frustrated, Sharon always encouraged me and backed me up.

I would also like to graciously thank my committee members, Ana Diex Roux, Bhramar Mukherjee, Jennifer Smith, and Pat Peyser for their helpful suggestions, guidance and support through the dissertation process. In addition, I would like to gratefully thank Ana Diex Roux for providing me funding during my first year of the Ph.D. program (2008-2009) from the grant titled "Michigan Center for Integrative Approaches to Health Disparities" (#MD002249, funded by NIH and NIMHD).

I am very lucky to be one of the members in the Kardia Research Group and have met many wonderful people who have helped me through the Ph.D. process. I would like to thank Alicia Lazarus, Erin Payne, Erin Ware, Stephanie Stenzel, Wei Zho, Tracy Fuller, Jacquelyn Taylor, Kristin Meyers, Todd Greene, Reagan Kelly, Jian Chu, Koji Yanagisawa, Linda Feldcamp, and

Yan Sun. Because of you, my Ph.D. experience was full of good memories and I will never forget the birthday party you held for me - that birthday party was one of the best parties I have ever had. In addition, I would particularly like to thank Alicia Lazarus, Erin Payne, and Stephanie Stenzel for providing useful suggestions, and helping me to edit my prospectus and this dissertation.

I would also like to thank my best Taiwanese friends, Hsing-Fang Hsieh, Yen-Ling Lin, Chi-shan Lee, Jenny Lee, Hui-Wen Chuang, Chunag-Chun Hu, and Cherry Chang. You all helped me through the stressful times, and sharing the joyful moments of my life in Michigan. Because of you, Ann Arbor is like my second home and I am grateful that I met you all here.

Last, I would like to thank my dearest parents, brother and sister for their undying love and support. Without them, I would never have had this chance to come to the U.S. and complete this dissertation at the University of Michigan. I have not seen you for almost 2 years, and I miss you all very much.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

High blood pressure and obesity are two major public health problems and important contributors to multiple chronic disorders including cardiovascular disease, stroke, and type II diabetes. Heritability studies have shown that there is a substantial fraction of the interindividual variation in blood pressure and body mass index (BMI) that is due to additive genetic variation. Recent genome-wide association studies (GWAS) have identified many novel chromosomal regions associated with these traits that are highly significant ($p<5x10^{-8}$) and replicate across multiple studies. Because most of the single nucleotide polymorphisms (SNPs) in these chromosomal regions fall in non-coding sequences, their functional mechanisms are not well understood. To develop a deeper knowledge of the biological mechanisms that may be underlying the genetic associations for blood pressure and BMI, we examined the relationship between genetic variants and gene expression levels among non-Hispanic white sibships in the Genetic Epidemiology Network of Arteriopathy (GENOA). In addition, we estimated the heritability of the gene expression levels from the 27 blood pressure-related genes and the 30 BMI-related genes.

The heritability study found that approximately 67% of the 27 BP-related genes and 47% of the 30 BMI-related genes had gene expression levels that were significantly heritable ($p<0.05$). The genetic association study revealed that many genetic variants within a gene region contribute to variation in expression levels. However, only a few of these gene expression levels were also associated with the phenotypic traits of blood pressure and BMI. For example, *ULK4* was the

only gene that was found to have genetic variants explaining significant variation in *ULK4* expression and its expression was also significantly associated with blood pressure levels.

Lastly, we investigated whether the relationship between gene expression levels and the outcomes of blood pressure and BMI were influenced by genetic variants. Our results indicated that the relationship between *CACNB2* expression and systolic blood pressure could be modified by genetic variants. This dissertation took an important step in following up the recently identified GWAS genetic loci associated with blood pressure and BMI to better understand the molecular epidemiology of these major risk factors.

Chapter 1

Introduction

Introduction

High blood pressure and obesity are two major risk factors for a wide range of common chronic diseases associated with target-organ damage of the heart, brain, kidneys, and peripheral arteries.[1,2] One major underlying cause of this collection of disease risks is the role of hypertension and obesity in arteriosclerosis (i.e., atherosclerosis and arteriolosclerosis) of the cardiac, cerebral, renal, and peripheral arteries. The clinical sequelae of arteriosclerosis are well known and include heart attack, heart failure, stroke, dementia, chronic kidney disease and claudication.[2,3]

Hypertension and obesity (body mass index ≥ 30 kg/m^2) affect approximately one-third of U.S adults.[4] Although many traditional epidemiological risk factors have been found to be associated with hypertension and obesity (e.g. lower physical activity, psychosocial stressors, dietary factors such as dietary fats, higher sodium intake, lower potassium intake, and excessive alcohol intake), there is also evidence from family studies that genetic factors play a considerable role in determining interindividual variation in both blood pressure and body mass index (BMI).[5-7]

Recently, several genome wide association studies (GWAS) have identified multiple new genomic regions where single nucleotide polymorphisms (SNPs) associated with blood pressure

and BMI are highly significant and replicate across populations.[8,9] The majority of these SNPs fall in non-coding regions and do not have any apparent functional effect. As a result, we have only limited information and insights into the biological mechanisms that may be underlying the genetic associations recently identified for blood pressure and BMI. Investigating the genetic influence on the gene expression levels in these genomic regions may allow us to better understand the potential functional mechanism at the molecular genetic level. In addition, examining whether the gene expression levels in these key genomic regions are associated with blood pressure or BMI could build our knowledge about the combined effects of multiple DNA variation within an individual's regional haplotype that cannot be detected with a single SNP.

In this dissertation, I focused on the study of the contribution of genetic variants and gene expression levels at previously identified genetic loci for blood pressure and BMI. In addition, this work investigated the potential modifying impact of SNPs on the relationship between gene expression levels and the three phenotypic traits (i.e. systolic blood pressure, diastolic blood pressure, and BMI).

Background and Public Health Significance

Blood Pressure and Hypertension

Blood pressure is the pressure exerted by circulating blood upon the walls of blood vessels. Systolic blood pressure (SBP) is the blood pressure when the heart is contracting to force blood to move from the left ventricle of the heart into the pulmonary artery and the aorta. Diastolic blood pressure (DBP) is the pressure while the heart is relaxing and the heart is filled with blood. In other words, SBP is the maximum arterial pressure during the cardiac cycle when the heart contracts and DBP is the minimum arterial pressure when the heart dilates.[10] The American Heart Association classifies the relationship between blood pressure and hypertension

as shown in Table 1.1.[11]

Table 1.1. Classification of blood pressure for adults ≥18 yeas old.[11]

Blood Pressure Category	Systolic Blood Pressure mm Hg		Diastolic Blood Pressure mm Hg
Normal	<120	and	<80
Prehypertension	120 – 139	or	80 – 89
Hypertension Stage 1	140 – 159	or	90 – 99
Hypertension Stage 2	≥160	or	≥100
Hypertensive Crisis	>180	or	>110

In epidemiological studies, hypertension is often defined as (1) having a SBP >= 140 mm Hg or a DBP >= 90 mm Hg or taking antihypertensive medication, or (2) having been told at least twice by a physician or other health professional that one has hypertension.[1] There are two types of hypertension: 1) essential (primary) hypertension and 2) secondary hypertension. Unlike secondary hypertension that results secondarily from other diseases (i.e. kidney disease, tumors, endocrine disease, etc.), essential hypertension has no known identifiable cause. Approximately 95% of patients diagnosed with hypertension have essential hypertension.[12]

Hypertension Epidemiology

According to the National Health and Nutrition Examination Survey (NHANES), approximately 1 in 3 American adults (73 million people) have hypertension. From 1988-1994 through 1999-2002, the age-adjusted prevalence increased from 24.3% to 28.1% among whites.[4] From 1995 to 2005, the death rate from hypertension rose 25.2% and the actual number of deaths increased 56.4%.[4] The estimated yearly direct and indirect costs for hypertension increased from $63.5 billion in 2006 to $73.4 billion in 2009. Among adults with hypertension, 22% were not aware of their condition, 32% of those hypertensive adults were not using antihypertensive medication, and 36% of those treated did not have their blood pressure controlled.[1]

Blood Pressure Physiology

Maintenance of a normal blood pressure is determined by the balance between cardiac output and peripheral vascular resistance.[12] Hypertension results from either or both a sustained raised cardiac output or increased peripheral resistance, or both. Cardiac output is determined by heart rate and stroke volume (the volume of blood pumped from one ventricular of the heart with each beat). Peripheral resistance is dependent on the functional and anatomic changes in small arteries and arterioles.[12]

Genetic Studies of Blood Pressure

Heritability Studies

Heritability is defined as the proportion of inter-individual variability in a trait explained by additive genetic variability. Previous family studies have shown that the heritability of blood pressure is ~ 30%-40%.[5,13] In the Framingham Heart Study, heritability was estimated to be 0.42 for SBP and 0.39 for DBP after adjustment for age, gender, BMI, and antihypertensive medication usage.[5] In another white population, heritability was estimated to be 0.34 for SBP and 0.37 for DBP after adjustment for age, gender, antihypertensive medication usage, BMI, cholesterol, high-density lipoprotein cholesterol, alcohol intake and glucose levels.[13] These consistent heritability estimates for blood pressure indicate that this phenotype has a significant genetic component even after adjustment for the traditional risk factors mentioned above.

Genome-Wide Association Studies (GWAS)

GWAS is an approach that involves scanning a dense collection of genetic markers across the genome for associations with a particular trait. It has become popular in the past few years due to the introduction of cost-effective genotyping technologies where hundreds of thousands of SNPs can be genotyped on a single "chip". This approach is particularly useful in finding new

chromosomal regions that contribute to common, complex diseases, such as coronary artery disease (CAD), cancer, and diabetes.

Compared to other complex disease like diabetes, it has been a challenge to identify the genomic regions associated with blood pressure. In order to increase the power to find significant, replicated genetic variants associated with measures of blood pressure, three large GWAS consortia formed to combine the GWAS results from multiple studies in non-Hispanic whites for meta-analysis.[6-8] Each of these consortia performed meta-analyses that identified SNPs that achieved genome-wide significant ($p < 5 \times 10^{-8}$) associations with systolic or diastolic blood pressure. The Global BP Genetics (Global BPgen) consortium comprises 34,433 individuals of European descent from 17 cohorts.[14] At stage 1, they tested 2.5 million genotyped and imputed SNPs for association with SBP and DBP in the 34,433 subjects. At stage 2, they followed up findings with direct genotyping in other independent samples (N=71,225 of European descent from 13 studies, and N=12,889 of Indian Asian descent), and *in silico* comparison with another GWAS consortium (Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE),[6] N=29,136). Finally, they conducted the joint meta-analysis of stage 1 and stage 2 using inverse variance weighting.

In the joint meta-analysis of stage 1 and stage 2, SNPs that achieved genome-wide significance ($p < 5 \times 10^{-8}$, a Bonferroni correction for the estimated 1 million independent common variant tests)[15] were considered to be strong signals. They found association between common variants and SBP in three regions in or near the *CYP17A1* ($p = 7 \times 10^{-24}$), *MTHFR* ($p = 2 \times 10^{-13}$), and *PLCD3* ($p = 1 \times 10^{-8}$) genes. Also, five loci associated with DBP were identified, and those regions are in or near the *CYP1A2* ($p = 1 \times 10^{-23}$), *FGF5* ($p = 1 \times 10^{-21}$), *SH2B3* ($p = 3 \times 10^{-18}$), *c10orf107* ($p = 1 \times 10^{-9}$), and *ZNF652* ($p = 5 \times 10^{-9}$) genes.

CHARGE, a consortium including 29,136 people of European descent from 6 population-based cohort studies, also identified 4 loci associated with SBP (*ATP2B1*, p=4x10^{-11}; *CYP17A1*, p=1x10^{-10}; *PLEKHA7*, p=2x10^{-9}; *SH2B3*, p=5x10^{-9}) and 6 loci associated with DBP (*SH2B3*, p=2x10^{-14}; *CSK-ULK3*, p=2x10^{-10}; *ATP2B1*, p=1x10^{-9}; *ULK4*, p=2x10^{-9}; *CACNB2*, p=1x10^{-8}; *TBX3-TBX5*, p =4x10^{-8}).[16] At stage 1, they tested genotyped and imputed SNPs (~2.5 million) for association with SBP and DBP in 29,136 individuals. At stage 2, they followed up findings with *in silico* comparison with Global BPgen and then conducted the joint meta-analysis of stage 1 and stage 2.

The International Consortium for Blood Pressure (ICBP), which used a multi-stage design in 200,000 individuals of European descent, identified genetic variants associated with blood pressure at 16 novel genetic loci and replicated all 12 genetic loci reported from CHARGE and Global BPgen except *PLCD3*.[17] All significant associations attained p<5x10^{-9}, an order of magnitude beyond the standard GWAS significance (p<5x10^{-8}). These significant associations collectively explain around 0.9 % of the phenotypic variance for SBP and DBP.

Some of the reported genes have been shown in previous studies to be associated with blood pressure regulation. For example, *NPR3* encodes the natriuretic peptide clearance receptor (NPR-C). *NPR3* knockout mice exhibit reduce clearance of circulating natriuretic peptides and lower blood pressure.[18] *ADM* encodes adrenomedullin and has natriurettic, vasodilatory and blood pressure lowering properties.[19] *CYP17A1* encodes the cytochrome P450 enzyme that mediates steroid 17 α-hydroxylase which is a key step in the biosynthesis of mineralocorticoids and glucocorticoids that affect sodium handling in the kidney.[20,21] In addition, *CYP17A1* is known to harbor rare variants that have large effects on blood pressure.[22]

Body Mass Index

Body Mass Index and Obesity

BMI is the most commonly used quantitative measure to diagnose obesity. It is defined as weight in kilograms divided by height in meters squared. According to World Health Organization, BMI is classified as follows: less than 18.5 (underweight), 18.50- 24.99 (normal weight), 25.00-29.99 (overweight), and 30.00-more (obese).

Obesity Epidemiology

Approximately 32% and 34% of American adults are overweight and obese, respectively.[4] Data from NHANES 2005-2006 showed the prevalence of overweight and obesity is higher in non-Hispanic males (72.4%) than non-Hispanic females (57.5%). In addition, obesity rates have more than doubled among American children (6-11 years old) and more than tripled among those 12-19 years of age since the late 1970s.[23]

Obesity is a major public health problem which causes increased morbidity, mortality and economic burden worldwide. Obesity is a main risk factor for hypertension, cardiovascular disease, type II diabetes, kidney disease and some cancers. Epidemiological risk factors for obesity include excess energy intake and sedentary lifestyle. [24]

Genetic Studies of Body Mass Index

Heritability Studies

In family and twins studies, genetic factors contribute to between 25-60% of the population variation in BMI.[6,7,25-27] In the Framingham Heart Study, heritability was estimated to be 0.37 after adjustment for age, age^2, and gender.[27] Previous study reported the heritability of BMI among Genetic Epidemiology Network of Arteriopathy (GENOA) sample was 0.43 for whites and 0.45 for African Americans adjusted for age, age^2 and gender.[28]

Genome-Wide Association Studies

During 2009-2010, three meta-analysis GWAS identified multiple novel genomic regions associated with BMI in individuals of European descent.[9,29,30] First, the Genetic Investigation of Anthropometric Traits (GIANT) consortium comprises 32,387 individuals from 15 cohorts. At stage 1, they tested 2.5 million genotyped and imputed SNPs for association with BMI in their 32,387 subjects.[29] At stage 2, they followed up the 35 most significantly associated SNPs with direct genotyping in other independent samples (N=45,801 of European descent from 9 studies) and *in silico* comparison with 5 GWAS results (N=14,064). Finally, they conducted the joint meta-analysis of stage 1 and stage 2. In the joint meta-analysis of stage 1 and stage 2, they identified eight genes including *FTO* (p=4.3x10^{-51}), *MC4R* (p=4.6x10^{-18}), *TMEM18* (p=1.4x10^{-18}), *GNPDA2* (p=3.4x10^{-16}), *SH2B1* (p=5.1x10^{-11}), *MTCH2* (p=4.6x10^{-9}), *KCTD15* (p=2.3x10^{-8}), and *NEGR1* (p=6.0x10^{-8}). Since the p-value for *NEGR1* was close to genome-wide significance and BMI-increasing allele near *NEGR1* also showed a highly significant independent association with severe obesity in a pediatric cohort (p=2.2x 10^{-7}), it strongly suggested that this variant represents a newly discovered locus influencing BMI.

The deCODE Genetics consortium including 73,759 people of European descent from five GWAS studies also identified ten loci associated with BMI that reached genome wide significance.[30] At stage 1, they tested 305,846 genotyped SNPs for association with BMI in a sample of 73,759 people. At stage 2, they followed up the 23 most significantly associated SNPs with direct genotyping in another Denmark sample (N=11,036 of European descent) and *in silico* comparison with the GIANT consortium. Later, they conducted the joint meta-analysis of stage 1 and stage 2 and found 10 chromosomal regions passed genome wide significance (p= 1.6×10^{-7} [0.05/305,846 genotyped SNPs]). Four loci were new findings (*BDNF*, p=3.2x10^{-11}; *ETV5*,

p=7.2x10^{-11}; *SEC16B*, p=6.2x10^{-8}; *FAIM2*, p=1.2x10^{-7}). The other six loci were also reported in the GIANT consortium (*FTO*, p=1.1x10^{-47}; *TMEM18*, p=4.2x10^{-17}; *MC4R*, p=1.2x10^{-12}; *KCTD15*, p=7.3x10^{-12}; *NEGR1*, p=1.2x10^{-11}; *SH2B1*, p=3.2x10^{-10}).[30] Many of the loci are near genes that are highly expressed in the brain which underscores the importance of genes that regulate food intake over those in metabolism. This is not unexpected because a considerable part of the regulation of energy balance is controlled by eating behavior, which has been shown in monogenic obesity.

Most recently, the GIANT consortium expanded to include a total of 249,796 individuals of European descent and reported associations between SNPs and BMI in 12 known BMI susceptibility loci and 2 known waist circumference loci.[31,32] In addition, they identified 18 new loci associated with BMI.[9] Together, the 32 confirmed BMI loci explained 1.45% of the interindividual variation in BMI. *FTO* accounted for the largest proportion of the variance (0.34%), followed by *TMEM18* (0.15%). The proportion of the variance for the 18 newly discovered loci ranged from 0.01% (*NUDT3*) to 0.06% (*DNAJC27*).

Based on evidence from animal and human studies, *FTO* affects BMI in three ways: (1) it is widely expressed in the brain with a particularly high level of expression in the hypothalamic nuclei which are involved in regulating energy balance in mice.[33] (2) The risk allele is associated with decreased lipolytic activity in adipocytes, suggesting a role in fat cell lipolysis.[34] (3) The risk allele has a stronger effect in people who are less active.[35] Many of the remaining genes are expressed or known to act in the central nervous system, suggesting a likely neuronal component in the predisposition to obesity. For example, *SH2B1* is implicated in leptin signaling and Sh2b1-null mice are obese. Notably, the obesity in Sh2b1-null mice can be reversed by targeted Sh2b1 expression in neurons, suggesting that the effects of this gene on obesity are

mediated through the central nervous system (CNS).[36] *MTCH2* encodes a putative mitochondrial carrier protein that may function in cellular apoptosis.[37] *NEGR1* has a role in neuronal outgrowth.[38]

From Gene Variants to Gene Expression and then to Disease

The functional mechanisms that link genetic variation to complex traits is a longstanding puzzle in biology. Owing to the emergence of advanced microarray technology, many studies are now able to study the role of genetic variation in interindividual variation in gene expression levels as a first step in the mechanistic pathway. Gene expression is a highly heritable trait that is influenced both by *cis*-effects (i.e. variants in the gene's own regulatory regions) and *trans*-effects (i.e. genetic loci elsewhere in the genome). Non-genetic elements such as epigenetic factors also have an influence on gene expression levels. A genome wide association study of global gene expression has shown that 28% of all transcripts had heritability >0.3 in Epstein-Barr virus-transformed lymphoblastoid cell lines.[39] Recent publications have reported the potential causal impact of differential gene expression on complex disease risk.[40] For example, Emilsson *et al.* found significant correlation between gene expression in adipose tissue and obesity related traits. The evidence from these studies suggests that integrating the genetic variants, gene expression levels, and phenotypic traits will help us to begin to elucidate the functional mechanism of complex traits.

Study Population

The study population for this dissertation consisted of non-Hispanic whites from Rochester, Minnesota (MN) that were initially enrolled in the GENOA study. GENOA is a multicenter, community-based study of hypertensive sibships collected to identify genes influencing blood pressure levels and development of target organ damage due to hypertension.[41,42] In Rochester,

MN, the Mayo Clinic diagnostic index and medical record linkage system of the Rochester Epidemiology Project were used to identify non-Hispanic white residents of Olmsted County.[43] In the first phase (Phase I:1995-2000) of GENOA (N=1583 non-Hispanic whites), sibships containing at least two individuals with clinically diagnosed essential hypertension before age 60 were invited to participate, including both hypertensive and normotensive siblings. Participants were considered to have hypertension if they had either 1) a previous clinical diagnosis of hypertension by a physician with current anti-hypertensive treatment, or 2) an average SBP ≥140 mm Hg or DBP ≥90 mm Hg on the second and third blood pressure readings. Exclusion criteria were secondary hypertension, pregnancy, insulin-dependent diabetes mellitus, or active malignancy. Phase I (1995-2000) data consisted of demographic information, medical history, clinical characteristics, lifestyle factors, and blood samples for genotyping and biomarker assays. The GENOA study was approved by the institutional review board of the Mayo Clinic, Rochester, MN. Written informed consent was obtained from all participants.

Dissertation Research Objectives and Specific Aims

The goal of this dissertation is to investigate the influence of genetic variants on gene expression levels at genetic loci that have been previously identified to contribute to interindividual variation in measures of blood pressure and BMI. In order to better understand the potential functional mechanisms at the molecular genetic level of these complex traits, I investigated the contribution of genetic variants to gene expression levels and examined the influence of gene expression levels on these traits among non-Hispanic whites in GENOA. Furthermore, I explored the potential modifying effects of SNPs on the relationship between gene expression levels and these traits. The specific aims for this dissertation are as follows:

Aim 1: To investigate the functional genetic factors of blood pressure variation by (1) evaluating the associations between genetic variations in the significant and replicated genomic regions and blood pressure traits, (2) investigating the association between these genetic variations and proximal gene expression, and (3) examining the relationship between proximal gene expression and blood pressure traits in the non-Hispanic whites of the GENOA.

Aim 2: To investigate the functional genetic factors underlying body mass index variation by (1) evaluating the associations between genetic variations in the significant and replicated genomic regions and body mass index, (2) investigating the association between these genetic variations and proximal gene expression, and (3) examining the relationship between proximal gene expression and body mass index in the non-Hispanic whites of the GENOA.

Aim 3: To examine whether SNPs modify the association between gene expression levels and three complex traits (i.e. systolic blood pressure, diastolic blood pressure, and BMI) in the non-Hispanic whites of the GENOA.

There are a total of six chapters in this dissertation. Chapter 1 contains the epidemiological descriptions, genetic background, and research objectives. In chapter 2, we examine the conventional predictors and heritability of blood pressure, BMI, and gene expression levels. In addition, the phenotypic correlations are explored within blood pressure measures and within gene expression levels. In chapter 3, through integrating the genetic variant and gene expression data, the functional genetic factors underlying blood pressure variation are investigated as described in Aim 1. Similarly, in chapter 4, we examine the effect of genetic factors on the

expression of the BMI-related genes as described in Aim 2. In chapter 5, we explore the contribution of SNP by gene expression to the variation in blood pressure and BMI as described in Aim 3. Lastly, chapter 6 contains the integration of the results we found in this dissertation. Insights from the findings, limitations, and future directions are also discussed in this chapter.

Chapter 2

Heritability, Correlation and Predictors of Gene Expression Levels and three Phenotypic Traits among non-Hispanic Whites in Genetic Epidemiology Network of Arteriopathy (GENOA)

Introduction

Complex diseases are attributable to multiple genetic and environmental risk factors and the complex interactions between them. A prerequisite for most genetic epidemiological research is to first demonstrate that the trait in question is influenced by inherited factors and to estimate how much variation is due to genetic factors. Heritability, defined as the proportion of total phenotypic variation due to genetic variation in a population,[44] is commonly estimated using related individuals to test the null hypothesis that no genes are involved in determining the variation in a trait. Many previous studies have estimated the heritability of blood pressure and body mass index (BMI) which ranges from 0.25 to 0.6.[5-7] In addition, a recent genome wide association study of global gene expression has shown substantial evidence of heritability in gene expression levels.[39] Dixon *et al.* found that among 54,675 transcripts that were expressed, 15,084 transcripts (28%) of them had substantial heritability (h^2>0.3).

Heritability is a population-specific parameter reflecting the range of additive and non-additive genetic factors and the environmental factors in each population. Genetic variance is dependent on the segregation the alleles that influence the trait in a population, the allele frequencies, the effect sizes of the variants, and the mode of gene inheritance. Other factors such

as environmental variance, pedigree size and structure, measurement error of the trait, and phenotypic variability can also affect the estimate of a trait's heritability.[45]

In this chapter, we estimate the heritability of gene expression levels and three phenotypic traits (systolic blood pressure, diastolic blood pressure, and BMI) before conducting the following genetic epidemiological research in this dissertation. The epidemiological predictors of gene expression levels and three phenotypic traits were assessed before estimating their heritability. These predictors were then included in the heritability models as covariates.

Methods

Study Population

The study population for this dissertation consists of non-Hispanic, white individuals from Rochester, Minnesota (MN) that were initially enrolled in the GENOA study. GENOA is a multicenter, community-based study of hypertensive sibships collected to identify genes influencing blood pressure levels and development of target organ damage due to hypertension.[41] In Rochester, MN, the Mayo Clinic diagnostic index and medical record linkage system of the Rochester Epidemiology Project were used to identify non-Hispanic white residents of Olmsted County.[43] In the first phase (Phase I:1995-2000) of GENOA (N=1583 non-Hispanic whites), sibships containing at least two individuals with clinically diagnosed essential hypertension before age 60 were invited to participate, including both hypertensive and normotensive siblings. Participants were considered to have hypertension if they had either 1) a previous clinical diagnosis of hypertension by a physician with current anti-hypertensive treatment, or 2) an average systolic blood pressure (SBP) ≥140 mm Hg or diastolic blood pressure (DBP) ≥90 mm Hg on the second and third blood pressure readings. Exclusion criteria were secondary hypertension, pregnancy, and insulin-dependent diabetes mellitus. There were 1583 people with

phenotype data, and we removed people who had secondary hypertension (N=9), had gender reassignment (N=1), or lacked SNP data (N=138). Participants who did not self report as non-Hispanic whites (N=5) or were outliers (> mean + 4 standard deviation or < mean - 4 standard deviation)) for the outcome (SBP, DBP, and BMI) (N=7) were also excluded. There were 789 people with phenotype data and gene expression data. The GENOA study was approved by the institutional review board of the Mayo Clinic, Rochester, MN. Written informed consent was obtained from all participants.

Phenotype Measurement

Blood was drawn by venipuncture after an overnight fast of at least eight hours. Resting SBP and DBP were measured with random zero sphygmomanometers (Hawksley and Sons, West Sussex, England) and cuffs appropriate for arm size. Three readings were taken in the right arm after the participant rested in the sitting position for at least five minutes according to standards set by The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High BP (JNC-7) guidelines.[46] Of the three readings, the last two readings were averaged for the analyses. Height was measured by wall stadiometer, weight by electronic balance, and BMI was calculated by definition as weight in kilograms divided by height in meters squared.

Selected Gene Expression Levels for Blood Pressure-related Genes and BMI-related Genes

The International Consortium for Blood Pressure (ICBP)[8] is the largest blood pressure consortium which contains all of the study samples from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE)[16] and Global Blood Pressure Genetics (Global BPgen).[14] Similarly, the expanded Genetic Investigation of Anthropometric Traits (GIANT)[9] is the largest BMI consortium and it includes all of the study sample from deCODE genetics.[30] Top

SNPs that were identified from these two consortia were used as index SNPs to select for proximal genes and their gene expression levels for study.

Because only some of the index SNPs identified from previous GWAS fall within gene boundaries, we used four strategies to choose the genes for the expression studies based on the location of index SNPs. First, if the index SNP is inside of a gene, then that specific gene was selected. Second, if the index SNP is outside of a gene and the distance between the index SNP and the closest gene is within 100 kb, then the closest gene upstream and the closest gene downstream of the index SNP were selected. Third, in gene sparse regions, there is typically only one gene either upstream or downstream within ± 100kb of the index SNP. In this case, the single gene was chosen. Finally, in gene deserts where there is no gene within ± 100kb of the index SNP, we excluded the index SNP from our analysis. The detailed information of selected genes for blood pressure and BMI are shown in Appendix 2.1 and Appendix 2.2, respectively.

Gene Expression Assessment and Quality Control

Blood samples for beta-lymphocyte extraction were collected during GENOA Phase I (1995-2000) study and GENOA Phase II (2000-2005) study. Beta-lymphocytes were isolated and transformed using Epstein-Barr virus at the time of collection. For the present study, individuals who participated in an ancillary GENOA study: Genetics of Microangiopathic Brain Injury Study (GMBI) had their lymphocytes re-initiated for gene expression assessment. Specifically, cell lines of 237 subjects and 552 subjects were collected and transformed during 1995-2000 and 2001-2005, respectively. Immortalized lymphocytes were stored at -180°C in a freezing medium containing 10% dimethyl sulfoxide (DMSO) as a cryoprotectant. To re-initiate the lymphocytes, a vial of cells is retrieved from cryostorage, thawed at 37°C, rinsed twice in Roswell Park Memorial Institute (RPMI) 1640 culture medium (Irvine Scientific, Santa Ana CA), and

transferred to a culture flask along with fresh culture medium supplemented with 15% fetal bovine serum (Irvine Scientific, Santa Ana CA). All cultured lymphocytes were maintained in the standardized culture medium, supplemented with 1% penicillin/streptomycin/L-glutamine (Sigma-Aldrich, St. Louis MO), at 37°C and 90% humidity in 5% CO2 for two to three weeks with regular changes of culture medium until a cell density of 5 x 10^5 cells/mL was achieved.

RNA samples from cell lines were extracted using standard protocols. RNA quality was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Foster City, CA) and quantified by spectrophotometry using the Nanodrop ND-1000 (Nanodrop Inc., Wilmington, DE). All RNA samples used in the present study yielded both an A260/A280 absorbance ratio greater than 2.0 and a RNA Integrity Number (RIN) ≥ 8. One μg of RNA was labeled using the WT Expression labeling assay (Applied Biosystems/ Ambion, Foster City, CA) including the labeling controls from the GeneChip Eukaryotic Poly-A RNA Control Kit (Affymetrix, Santa Clara, CA). Each step of the labeling protocol was monitored using the Agilent 2100 Bioanalyzer or the Nanodrop spectrophotometer, as recommended by the manufacturer. Hybridization buffer, Eukaryotic Hybridization Controls, and OligoB2 controls were added to the cDNA fragments just prior to hybridization to the Affymetrix Human Exon 1.0 ST Array. Hybridization was performed at 45°C for 17 hours. Following hybridization, the chips were washed and stained with a phycoerythrin-strepavidin conjugate and were scanned at an excitation wavelength of 488 nm.

Array quality control was performed using Affymetrix Expression Console™ (v 1.1) at the transcript level using core-level probe sets. All array images passed visual inspection. Hybridization controls were all present with signal increases following concentration. Labeling control signal strengths followed the order Lys < Phe < Thr < Dap. Signal intensity plots were

examined for raw and processed data to identify outliers.

Raw intensity data were processed using the Affymetrix Power Tool software. Probe summarization and probe set normalization were performed using Robust Multi-array Analysis (RMA), which included background correction, quantile normalization, log2-transformation, and probe set summarization.[47] Only core probesets were used to assess exon-level expression. Gene-level expression was assessed by averaging all the core probe sets for that gene.

Quality control of exon array data were done by examining the QC matrix to identify possible outliers or non-performing samples using the Partek Genomic Suite software (V.6.6). Analyses were restricted to core probe sets. Probe sets which are known to cross-hybridize and those with undetectable expression were also excluded.

Demographic Predictors

Age, age^2, gender, and BMI were examined as potential predictors of blood pressure. Age, age^2, gender were examined as potential predictors of BMI. For gene expression levels, the contributions of age and gender were estimated for gene expression levels. These predictors have been shown to have significant association with these phenotypic traits thus were selected to be included in the statistical model.[8,9] Linear mixed models were used to control for family structure while examining the conventional demographic predictors that were associated with systolic blood pressure (SBP), diastolic blood pressure (DBP), BMI, and gene expression levels. The multivariable statistical models for each phenotype are listed as follows:

$$\text{Blood Pressure}_{ij} = \beta_{oi} + \beta_1(\text{Age}_{ij}) + \beta_2(\text{Age}^2_{ij}) + \beta_3(\text{Gender}_{ij}) + \beta_4(\text{BMI}_{ij}) + \varepsilon_{ij}$$

$$\text{BMI}_{ij} = \beta_{oi} + \beta_1(\text{Age}_{ij}) + \beta_2(\text{Age}^2_{ij}) + \beta_3(\text{Gender}_{ij}) + \varepsilon_{ij}$$

$$\text{Gene Expression Level}_{ij} = \beta_{oi} + \beta_1(\text{Age}_{ij}) + \beta_2(\text{Gender}_{ij}) + \varepsilon_{ij}$$

In the linear mixed effect model, i represents family and j represents individual. β_{oi} is a random intercept for the i^{th} family, and it is normally distributed with mean β_o and variance σ_b^2. ε_{ij} is the residual variation within the j^{th} individual from the i^{th} family, and it is normally distributed with mean 0 and variance σ^2. ε_{ij} is assumed to be independent from β_{oi}.

Pearson's Correlation

Pearson's correlation was used to explore the relationship between each pair of traits. These traits were approximately normally distributed and were adjusted for known risk factors before examing their correlations. Although Pearson's correlation does not take the family correlation into account, it provides a robust estimate of the general relationship or covariability among the outcomes of interests. The correlation, r, between trait X and trait Y is estimated as follows:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Where $\bar{X}$ is the sample mean of X, $\bar{Y}$ is the sample mean of Y, and r is an estimate of the true correlation, ρ.

H_o: ρ=0 H_a: ρ≠0.

When the two traits are approximately normally distributed, the null distribution of the test statistic follows a Student's t-distribution with n-2 degrees of freedom.

Heritability

In quantitative genetic models, the total variance of a phenotype ($\sigma^2_{\mathbf{P}}$), can be modeled as the sum of genetic variance ($\sigma^2_{\mathbf{G}}$) and environmental variance ($\sigma^2_{\mathbf{E}}$): $\sigma^2_{\mathbf{P}} = \sigma^2_{\mathbf{G}} + \sigma^2_{\mathbf{E}}$, assuming no gene-environment interaction.[45] There are two types of heritability: broad sense heritability (H^2) and narrow sense heritability (h^2). The broad sense heritability is defined as the proportion of total phenotypic variation due to the variation in "all genetic effects" (i.e. additive, dominant, gene-gene interaction, and gene-environment effects): $H^2 = (\sigma^2_A+\sigma^2_D+\sigma^2_I) /\sigma^2_P$. On the other

hand, the narrow sense heritability is defined as the proportion of total phenotypic variation due to additive genetic variation: $h^2 = \sigma^2_A/\sigma^2_P$.[44] Most research studies are interested in narrow sense heritability because except for full siblings and twins, most relatives only share one or no copies that are identical-by-decent (IBD). Therefore, dominance and other non-additive genetic effects that require two copies do not contribute to their phenotypic resemblance.[45]

Heritability can be estimated by using observed phenotypic data from families with known genetic relationships between the family members. Using analysis of variance (ANOVA), the variance of an observed trait can be partitioned into the variance between families (σ^2_B) and the variance within families (σ^2_W).[48] The proportion of total phenotypic variance due to the between family variance represents the intra-class correlation coefficient: $r=\sigma^2_B/(\sigma^2_B+\sigma^2_W)$. This intra-class correlation coefficient multiplied by the reciprocal of the expected amount of genetic information shared by family members can provide an estimate of heritability.[44] Full siblings share approximately ½ of their genetic information, and thus, the estimate is $h^2=2r$.

When the family structure is complicated, a more robust heritability can be estimated by variance component estimation. By using this approach, additive genetic effects can be modeled using the variance covariance matrix of the trait which is expressed as a function of identity-by-decent as follows :

$$\mathrm{Cov}(P_x,P_y) = 2\,\Phi_{xy}\sigma^2_g + I_{xy}\sigma^2_e$$

Φ = kinship coefficient for two individuals, *x* and *y*. This coefficient represents the probability that a randomly selected allele from each individual at the same locus is identical-by-decent.

σ^2_g = additive genetic variance

I = identity matrix

σ^2_e = random environment variance including measurement error

In this study, heritability was estimated by using SOLAR (Sequential Oligogenic Linkage Analysis Routines)[49] which implements a variance component regression based on maximum likelihood estimation to partition phenotypic variance according to the following model:

$$y_i = \mu + \Sigma\beta_j X_{ij} + g_i + \varepsilon_i$$

Where i represents individual and j represents covariates, y_i is the outcome measure for the i^{th} individual, μ is the mean of y in the population, β_j is the regression coefficient associated with j^{th} covariate, X_{ij} is the j-th covariate with associated regression coefficient β_j, g_i is the additive genetic effect (normally distributed, mean=0, variance =σ^2_g), and ε_i is a random residual effect (normally distributed, mean=0, variance=σ^2_e). The model is constrained so that the sum of variances $\sigma^2_g+\sigma^2_e$ equal 1. σ^2_e includes variance due to any non-additive or unmeasured environmental component of variance and measurement error. Based on this equation, σ^2_g provides the heritability of the residual variance of the trait after adjusting for covariates. Statistical significance of the heritability is estimated by comparing the log-likelihood of the full model above to the log-likelihood of a model with σ^2_g constrained to 0. The hypothesis test is as follows:

$$H_0: \sigma^2_g = 0 \text{ vs. } H_a: \sigma^2_g \neq 0.$$

Under the null hypothesis, the distribution of the likelihood ratio test statistic is a 50:50 mixture of a chi-squared distribution with one degree of freedom and a point mass at zero.

Results

Descriptive Statistics

There are 789 subjects from 391 families with both phenotype and gene expression data in the non-Hispanic whites of GENOA. The distribution of sibship sizes is shown in Table 2.1. The most frequent sibship size was two (181/391 families=46.29%). Descriptive statistics of the

covariates and outcome variables can be found in Table 2.2. The mean age of the subjects was 55.65 years old, 58.05% were female, 41.44% were obese, around three-fourth (74.14%) had hypertension, and more than half (66.16%) were taking antihypertensive medication. Participants had a mean SBP of 132.80 mm Hg and DBP of 78.52 mm Hg. For people who were taking antihypertensive medication, we adjusted antihypertensive drug treatment by adding 15 mm Hg and 10 mm Hg to SBP and DBP, respectively. This adjustment is consistent with the International Consortium for Blood Pressure Genome-Wide Association Studies (ICBP) approach to adjusting for medication use .[8]

The histograms of the drug adjusted SBP and DBP were approximately normally distributed (Figure 2.1 and Figure 2.2, respectively). Figure 2.3 shows that the distribution of BMI was slightly right skewed with modest positive kurtosis (0.7691). The mean, standard deviation, minimum, and maximum statistics of normalized gene expression data for the 27 blood pressure-related genes (BP-related genes) and the 30 BMI-related genes are shown in Table 2.3 and Table 2.4, respectively.

Conventional Predictors for Phenotypes and Gene Expression Levels

Linear mixed models were used to explore the predictors associated with SBP, DBP, BMI and gene expression levels in GENOA. In the multivariable analysis, age, age^2 and BMI were significantly associated with SBP (p<0.05). For DBP, age, age^2, gender, and BMI were all strong predictors (p<1x10^{-4}). Age was the only predictor significantly associated with BMI (p=0.009) (Table 2.5).

We then examined the association between conventional risk factors (i.e. age and gender) and gene expression levels. Before investigating the association, the relationship between age and batch were found to be highly correlated (Appendix 2.3, ANOVA : p=2x10^{-6}). This

correlation was probably due to the GENOA re-sampling strategy. In Phase II, older individuals were re-recruited first to participate in this phase. In addition, we also found an association between sibship and batch (Fisher's test: $p<0.0001$). Siblings tended to visit the clinic at the same time thus many of the cell lines were established for sibships in close time proximity. Re-initiation of cell lines followed the same time clustering, resulting in association between sibship and gene expression batch.

Gender was found to be moderately associated with batch effect (Appendix 2.4, χ^2 test: $p=0.0286$) but not with age (data not shown, $p=0.12$). Age and gender were then used to predict gene expression level before and after adjusting for batch effects. Before adjusting for batch effect, gene expression levels were found to be associated with age for 9 BP-related genes and 9 BMI-related genes. Gender was a significant predictor for only 3 BP-related genes and 3 BMI-related genes (Appendix 2.5 and Appendix 2.6, respectively). After removing the variation in gene expression due to batch effects, gene expression levels were not associated with age for BP-related genes but were significantly associated for 5 BMI-related genes (*ETV5*, $p=0.019$; *KCTD15*, $p=0.044$; *MAP2K5*, $p=0.002$; *TFAP2B*, $p=0.028$) (Table 2.5 and Table 2.6, respectively). Age explained from 0.003% (*NEGR1*) to 1.02% (*MAP2K5*) variation in these BMI-related gene expression levels. Because age and batch were highly correlated, it is possible that when we adjusted for batch effect, we also removed the variation of gene expression level due to age. Thus, the associations between age and gene expression levels were examined in each batch for the 27 BP- related genes. The Q-Q plot of model p-values indicated that no association was found between age and gene expression levels for these BP-related genes (Appendix 2.7).

Gender was significantly associated with gene expression levels for 3 BP-related genes (*MTHFR*, p=0.013; *SH2B3*, p=0.039; *SLC39A8*, p=0.026) and 2 BMI-related genes (*FTO*, p=0.009; *SLC39A8*, p=0.026) (Table 2.6 and Table 2.7, respectively). Gender explained 0.004% (*GOSR2*) to 0.685% (*MTHFR*) and 0.001% (*NEGR1*) to 0.685% (*FTO*) variation of gene expression levels for BP-related genes and BMI-related genes, respectively (Table 2.6 and Table 2.7, respectively).

Pearson Correlation Coefficient for Phenotypes and Gene Expression Levels

Systolic blood pressure and diastolic blood pressure are biologically and physiologically related traits. The Pearson correlation coefficient between these two traits was 0.728 after adjusting for age, age^2, gender, and BMI. The scatter plot in Figure 2.4 illustrates the covariation between these two phenotypes for their residuals after adjusting for known risk factors.

Gene expression levels are regulated both by *cis* and *trans* effects, and proximal SNPs may influence the expression of multiple genes nearby. Therefore, exploring correlation among the gene expression levels will allow us to better understand their general relationship. Table 2.8a-Table 2.8c and Table 2.9a-Table 2.9c illustrate the Pearson correlation coefficients among expression levels of BP-related genes and BMI-related genes, respectively, after adjusting for age, gender, and batch effect. For the 27 BP-related genes with 351 pairwise correlations, there were 262 pairs (75%) of gene expression levels that have significant correlation coefficients with p-value < 0.05. Among these significant correlation coefficients, 54 pairs (15%) of gene expression levels had correlation coefficients > 0.3 or < -0.3, and 8 pairs (2%) of gene expression levels had correlation coefficients >0.5 or < -0.5 (Table 2.8a - Table 2.8c). The largest correlation coefficient was between *PLCD3* expression (chr17) and *PLCE1* expression (chr10) (r=0.6338), followed by *PLCD1* expression (chr17) and *FES* expression (chr15) (r=0.6318), and

ATP2B1 expression (chr12) and *NT5C2* expression (chr10) (r=0.5189) (Table 2.8c). For the 30 BMI-related genes, there were more pairs of gene expression levels that were moderately correlated. Out of 435 pairwise correlations, 345 pairs (80%) of gene expression levels had significant correlation coefficients with p-value < 0.05. Among these significant correlation coefficients,154 pairs (35%) of them had correlation coefficients >0.3 or < -0.3 and 50 pairs (11%) of them had correlation coefficients >0.5 or < -0.5 (Table 2.9a - Table 2.9c). The largest correlation coefficient was between *CHST8* expression (chr19) and *FAIM2* expression (chr12) (r=0.8151) (Table 2.9c), followed by *LRP1B* expression (chr2) and *NEGR1* expression (chr1) (r=0.7758) (Table 2.9a), and *KCTD15* expression (chr19) with *CHST8* expression (chr19) (r=0.7525) (Table 2.9c).

Heritability of Phenotypes and Gene Expression Levels

Heritability was estimated using SOLAR[49]. The estimates of h^2 for unadjusted SBP, DBP, and BMI were 0.223, 0.247 and 0.464, respectively (Table 2.10). After including age, age^2, gender, and BMI in the statistical model in the SOLAR analysis, the estimate of heritability for SBP increased to 0.352, and thus the additive genetic factors accounted for ~0.281 [0.352 x (1.0-0.2014)] of the total variance in SBP. Similarly, after adjusting for covariates, the heritability for DBP was 0.312, and thus the additive genetic factors accounted for ~0.284 [0.312 x (1.0-0.0885)] of the total variance in DBP. After adjusting for age and gender, the estimate of heritability for BMI was 0.472 . Thus, the additive genetic factors accounted for ~0.471 [0.472 x (1.0-0.0021)] of the total variance in BMI.

For expression data of the 27 BP-related genes, after including age, gender, and batch effects in the statistical model in the SOLAR analysis, the estimate of h^2 ranged from 0.006 (*GOSR2, p*=0.47) to 0.713 (*ULK4, p*=3.93×10^{-14}) (Table 2.11). Around 67% (=18/27 genes) of these genes

had expression profiles that have statistically significant heritability (p<0.05) and 40% (=11/27) of them have h^2>0.25. *ULK4* had the highest h^2 with an estimate of 0.713. The heritabilities of gene expression levels for the 30 BMI- related genes ranged from 0.0048 (*NEGR1, p*=0.494) to 0.5092 (*STK33, p*=1.00x10^{-7}) after adjusting for age, gender, and batch effects (Table 2.12). Out of the 30 genes, 14 genes (46.67%) had statistically significant heritability (p<0.05) and 6 genes (20%) had h^2>0.25. *STK33* had the most significant heritability (0.509) followed by *ETV5* (0.428).

Discussion

In this study, we found that estimates of heritability for SBP, DBP and BMI were 0.35, 0.31, and 0.47, respectively, after including known risk factors in the statistical model in the SOLAR analysis. Our results were consistent with previous genetic epidemiological studies which indicated that the heritability of these three traits ranged from 0.25-0.60 depending on the population, study design, and adjustment risk factors.[6]

Both age and gender have been shown to be associated with gene expression levels in peripheral blood,[50] and many studies have adjusted for these two covariates when conducting gene expression association analysis.[51,52] In our study, we found that age and gender had moderate associations with some of the gene expression levels. The Pearson correlation analysis of gene expression levels suggested that 75% and 80% of the correlation coefficients were significant (p<0.05) for BP-related genes and BMI-related genes, respectively.

Many earlier studies have demonstrated that gene expression levels were influenced by inherited genetic factors and were highly heritable in mice[53], yeast[54] and humans.[39,55] Specifically, a genome-wide gene expression study has shown that around 28% transcripts of those expressed had significant h^2>0.3 in beta-transformed lymphocytes in humans.[39] Our results

revealed that around 67% and 47% of the expression in the 27 BP-related genes and the 30 BMI-related genes were significantly heritable (p>0.05). In addition, there were more BP-related genes (11 genes, 40%) with significant heritability >0.25 than BMI-related genes (6 genes, 20%). These results suggested that there was more variation in BP-related gene expression due to additive genetic variation than in BMI-related gene expression in beta-transformed lymphocyte among non-Hispanic whites in GENOA.

One notable feature of the gene expression heritability results is that some genes had high heritability (e.g. *ULK4*, h^2=0.713, p=3.93x10^{-14}) while some others had low heritability close to 0 (e.g. *GOSR2*, h^2=0.006, p=0.47) (Table 2.11). This observation could be attributable to the measurement of narrow sense heritability used in this study, which assumes no gene-environment correlation, no gene-environment or gene-gene interaction, and no epigenetic effects.[56] Complex trait studies have shown that part of their variation is due to gene-environment and gene-gene interaction, and violations of these assumptions may lead to invalid estimates of heritability.[57] Zeller et al. reported that the some of the variation in gene expression levels was explained by the age*SNP interactions and gender*SNP interactions.[52] Other factors such as tissue type, amount of environmental variability, measurement error, and epigenetic factors can influence the estimate of heritability.

In conclusion, our results showed that there were significant correlations among expression levels of some genes we investigate. The heritability estimates of the three phenotypes and gene expression levels indicated that there was substantial variation in the traits that was due to additive genetic factors. Therefore, the results from this chapter warrant the remaining genetic epidemiological studies in this dissertation.

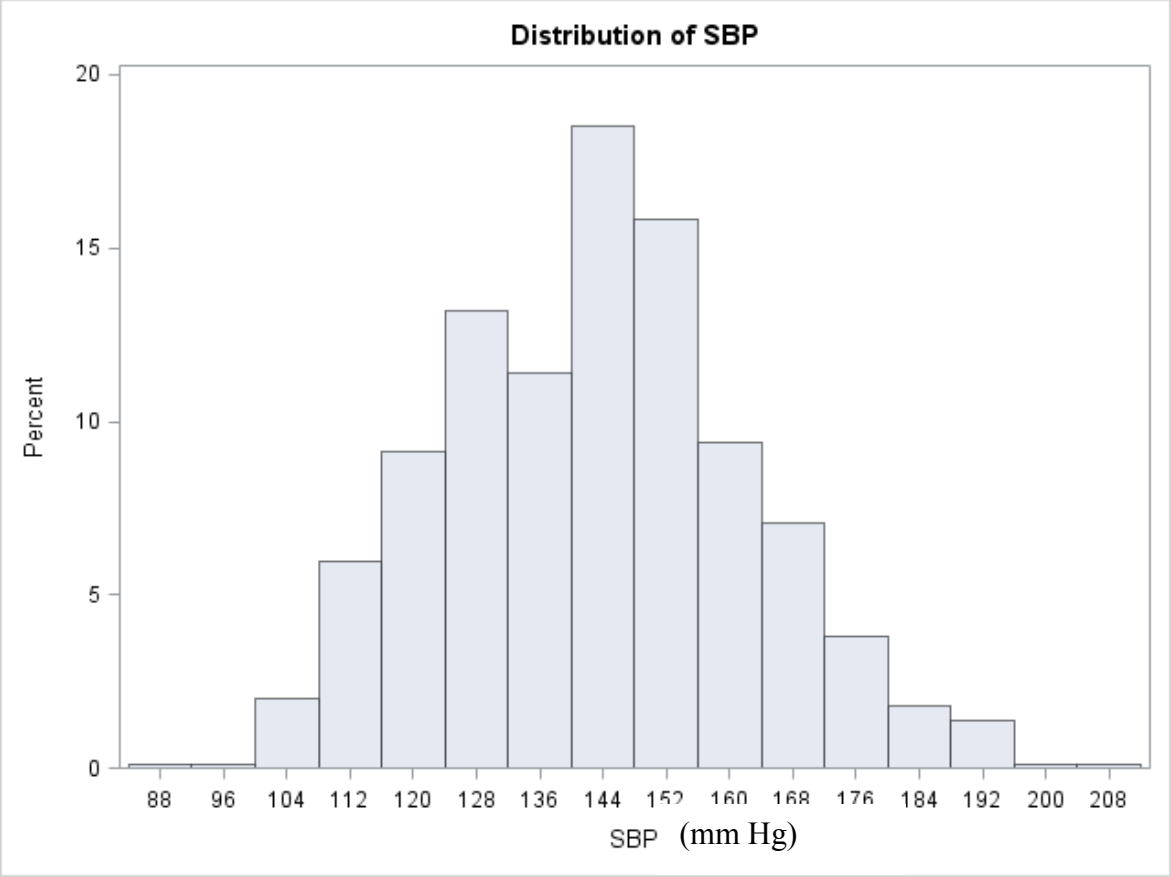
Table 2.1. Distribution of sibship structure in the non-Hispanic whites of GENOA.

No. of Siblings in Sibship	No. of Sibships	Total No. of Individuals	% of Total No. of Individuals
1	131	131	16.60%
2	181	362	45.88%
3	47	141	17.87%
4	16	64	8.11%
5	10	50	6.34%
6	4	24	3.04%
7	0	0	0.00%
8	1	8	1.01%
9	1	9	1.14%
Total	391 sibships	789 individuals	100%

Table 2.2 Descriptive statistics of covariates and outcome variables.

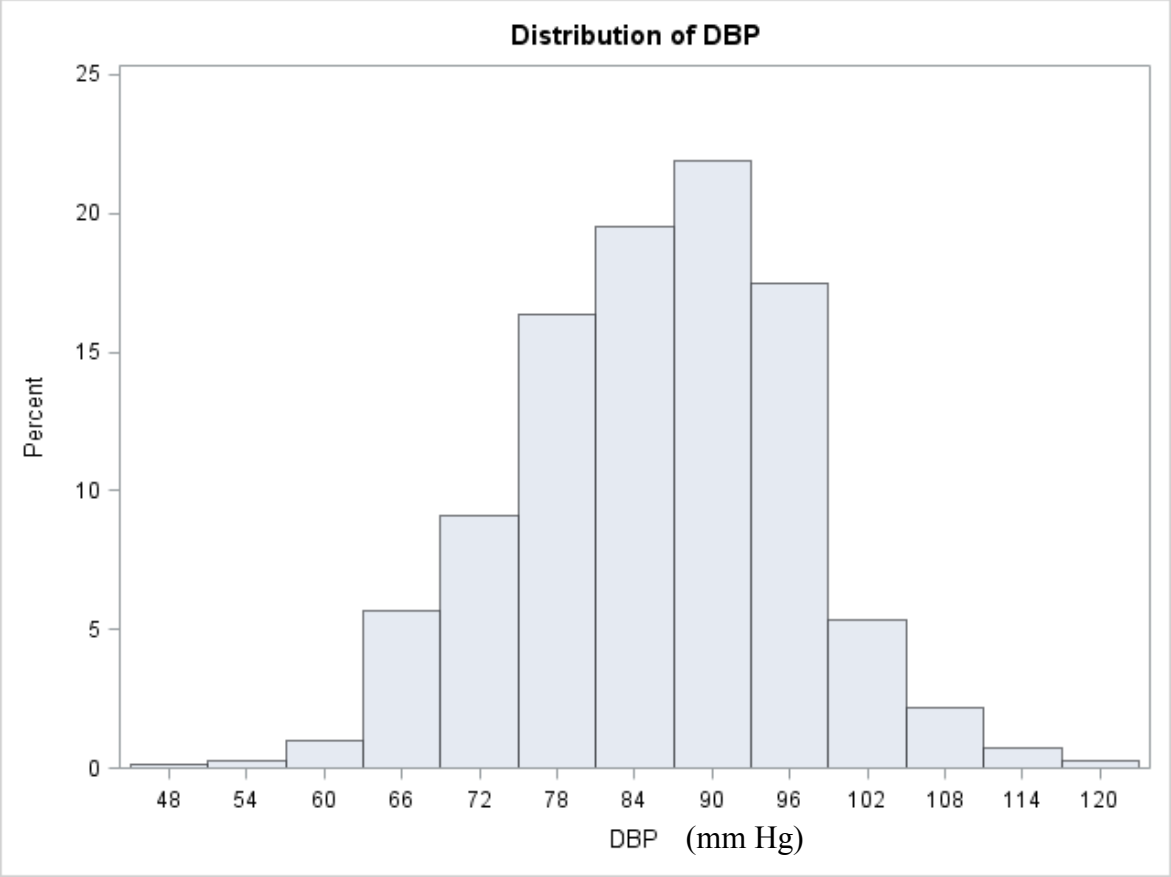
Continuous Variables	Mean	SD
Age, years	55.65	10.36
Height, cm	168.41	9.09
Weight, kg	84.53	18.64
BMI, kg/m^2	29.73	5.82
SBP, mm Hg	132.80	16.66
SBP+15mm Hg for Hypertension Drug	142.73	19.27
DBP, mm Hg	78.52	9.43
DBP+10mm Hg for Hypertension Drug	85.14	10.69
Categorical Variables	**Number**	**%**
Female	458	58.05
Obesity	327	41.44
Hypertension	585	74.14
Antihypertensive Medication	522	66.16

Figure 2.1. Distribution of systolic blood pressure with adjustment for antihypertensive drug treatment.*



*15 mm Hg was added to raw systolic blood pressure for those people who were taking antihypertensive medication.

Figure 2.2. Distribution of diastolic blood pressure with adjustment for antihypertensive drug treatment.*



*10 mm Hg was added to raw diastolic blood pressure for those people who were taking antihypertensive medication.

Figure 2.3. Distribution of raw body mass index (BMI) among non-Hispanic whites in GENOA.

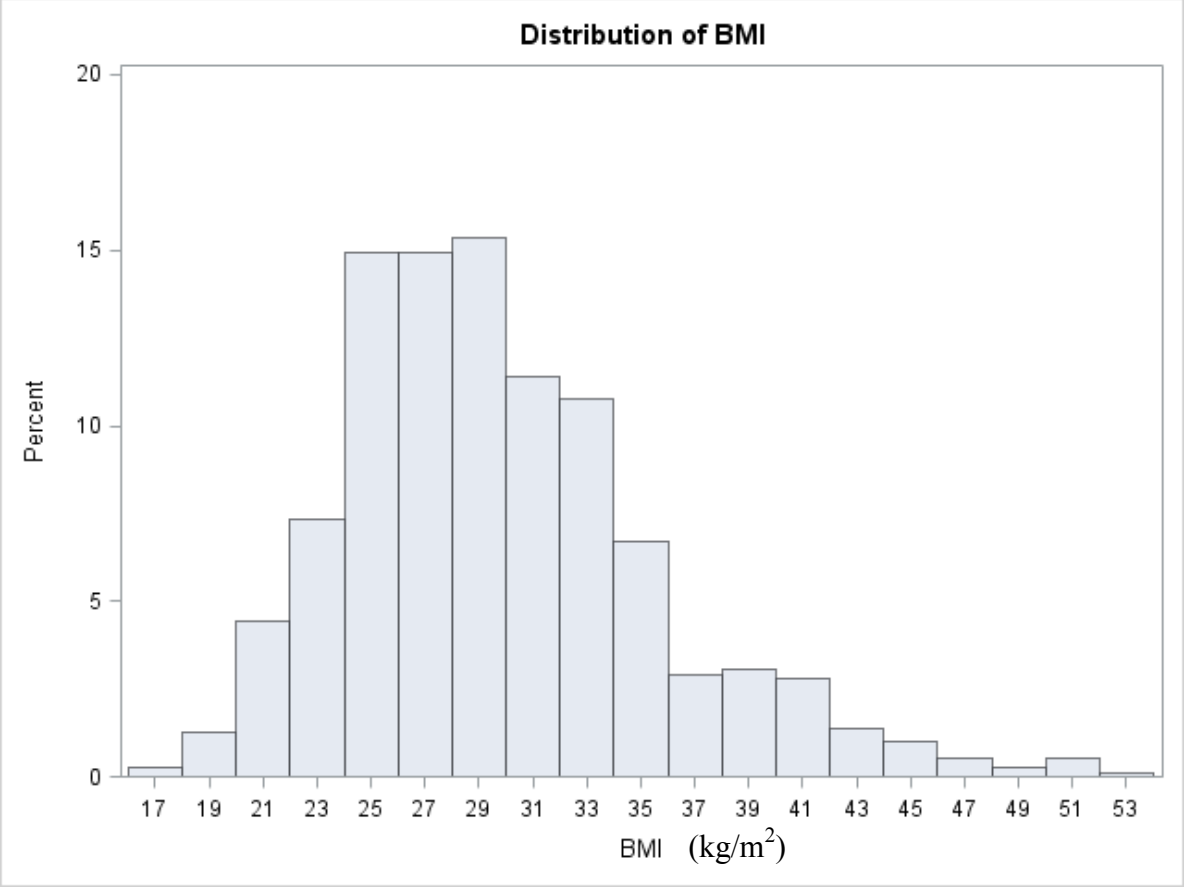


Table 2.3. Descriptive statistics of gene expression data for the 27 BP-related genes.

Gene	Chr	Mean	SD	Minimum	Maximum
MTHFR	1	8.16	0.20	7.54	8.77
CAPZA1	1	11.90	0.18	11.22	12.51
MOV10	1	9.89	0.22	8.88	10.52
SLC4A7	3	8.22	0.28	7.30	8.96
ULK4	3	6.39	0.39	5.35	7.49
MECOM	3	5.62	0.17	5.15	6.14
PRDM8	4	7.33	0.22	6.85	8.31
FGF5	4	6.37	0.37	5.56	8.47
SLC39A8	4	9.95	0.26	8.98	10.81
GUCY1A3	4	7.57	0.94	5.72	10.02
NPR3	5	5.70	0.28	5.03	7.09
HFE	6	6.05	0.19	5.42	6.64
BAG6	6	9.75	0.14	9.19	10.29
CACNB2	10	7.15	0.82	5.69	9.42
PLCE1	10	5.50	0.15	5.07	6.03
CNNM2	10	7.85	0.18	7.35	8.36
NT5C2	10	9.93	0.29	9.16	10.77
ADM	11	7.76	0.37	7.00	9.54
PLEKHA7	11	7.29	0.36	6.51	8.81
ARHGAP42	11	5.31	0.49	3.76	7.52
ATP2B1	12	9.38	0.30	8.38	10.38
SH2B3	12	9.16	0.20	8.31	9.76
CSK	15	10.19	0.17	9.52	10.84
FES	15	7.14	0.20	6.60	7.91
PLCD3	17	7.93	0.17	7.43	8.55
GOSR2	17	9.13	0.17	8.56	9.68
ZNF652	17	8.60	0.28	7.67	9.53

Table 2.4. Descriptive statistics of gene expression data for the 30 BMI-related genes.

Gene	Chr	Mean	SD	Minimum	Maximum
NEGR1	1	5.86	0.20	5.19	6.42
TNNI3K	1	5.36	0.18	4.90	5.96
SEC16B	1	6.40	0.20	5.83	7.39
TMEM18	2	8.97	0.18	8.36	9.75
ADCY3	2	9.93	0.31	8.96	10.85
DNAJC27	2	6.94	0.21	6.26	7.67
LRP1B	2	4.82	0.16	4.35	5.33
CADM2	3	7.88	0.25	6.96	8.67
ETV5	3	7.65	0.62	6.12	9.68
DGKG	3	7.04	0.26	6.32	8.44
SLC39A8	4	9.95	0.26	8.98	10.81
POC5	5	8.95	0.25	8.22	9.65
NUDT3	6	9.77	0.20	9.12	10.22
TFAP2B	6	6.88	0.52	5.99	9.61
LINGO2	9	5.15	0.21	4.63	5.80
STK33	11	6.96	0.38	5.35	8.16
BDNF	11	5.31	0.17	4.80	5.88
MTCH2	11	10.22	0.25	9.48	10.94
BCDIN3D	12	7.29	0.16	6.74	7.79
FAIM2	12	7.54	0.22	6.96	8.19
MTIF3	13	7.30	0.18	6.73	7.90
NRXN3	14	7.14	0.64	6.03	9.63
MAP2K5	15	7.73	0.25	6.90	8.36
GPRC5B	16	6.84	0.24	6.21	7.74
SH2B1	16	8.10	0.15	7.63	8.67
ATP2A1	16	6.94	0.12	6.58	7.39
FTO	16	11.17	0.13	10.73	11.59
CHST8	19	7.49	0.18	7.03	8.08
KCTD15	19	7.39	0.19	6.81	8.06
QPCTL	19	8.91	0.18	8.30	9.40

Table 2.5. Multivariable linear mixed model analysis of conventional predictors associated with SBP, DBP, and BMI.

Predictor	Systolic Blood Pressure		Diastolic Blood Pressure		BMI	
	Beta	P-value	Beta	P-value	Beta	P-value
Age, years	1.880	4.11×10^{-4}	1.680	3.54×10^{-7}	-0.039	0.009
Age^2, $years^2$	-0.010	0.045	-0.015	2.57×10^{-7}	NA*	NA*
Gender, (Female=1, Male=0)	-0.452	0.713	-3.898	1.82×10^{-7}	0.203	0.536
BMI, kg/m^2	0.593	1.93×10^{-8}	0.262	3.54×10^{-5}	-----	-------

* Age^2 was not significantly associated with BMI and it was not included in the final staitiscal model.
NA=Not applicable.

Table 2.6. Multivariable linear mixed model analysis of the effects of age and gender on expression levels in the 27 BP-related genes after adjusting for batch effect.

Gene	Chr	Age			Gender (Female=1, Male=0)		
		Beta	P-value	R^2 %	Beta	P-value	R^2 %
ADM	11	-0.0024	0.126	0.015	0.0394	0.125	0.041
ARHGAP42	11	0.0036	0.091	0.273	0.0144	0.685	0.275
ATP2B1	12	-0.0004	0.739	0.349	-0.0276	0.169	0.020
BAG6	6	-0.0001	0.814	0.012	0.0021	0.815	0.204
CACNB2	10	-0.0010	0.783	0.005	0.0305	0.603	0.005
CAPZA1	1	0.0003	0.681	0.009	0.0077	0.489	0.033
CNNM2	10	0.0001	0.929	0.001	0.0156	0.211	0.170
CSK	15	0.0009	0.186	0.194	0.0085	0.459	0.061
FES	15	0.0004	0.657	0.023	-0.0070	0.611	0.030
FGF5	4	0.0006	0.719	0.016	0.0124	0.647	0.027
GOSR2	17	0.0008	0.265	0.141	-0.0021	0.857	0.004
GUCY1A3	4	-0.0009	0.818	0.006	0.0265	0.691	0.018
HFE	6	0.0010	0.195	0.198	-0.0070	0.603	0.032
MECOM	3	0.0007	0.325	0.113	-0.0149	0.214	0.181
MOV10	1	0.0007	0.456	0.063	0.0045	0.769	0.010
MTHFR	1	-0.0008	0.333	0.103	-0.0350	**0.013**	0.685
NPR3	5	-0.0006	0.625	0.029	-0.0219	0.303	0.128
NT5C2	10	-0.0011	0.268	0.092	0.0288	0.084	0.225
PLCD3	17	0.0010	0.136	0.243	0.0041	0.726	0.013
PLCE1	10	0.0004	0.466	0.060	0.0122	0.244	0.153
PLEKHA7	11	0.0015	0.324	0.108	-0.0464	0.062	0.392
PRDM8	4	-0.0009	0.337	0.108	-0.0160	0.310	0.121
SH2B3	12	0.0011	0.139	0.225	0.0272	**0.039**	0.437
SLC39A8	4	-0.0010	0.347	0.106	-0.0422	**0.026**	0.594
SLC4A7	3	-0.0021	0.072	0.356	-0.0199	0.292	0.122
ULK4	3	0.0011	0.437	0.052	-0.0288	0.212	0.133
ZNF652	17	-0.0009	0.433	0.064	0.0157	0.402	0.072

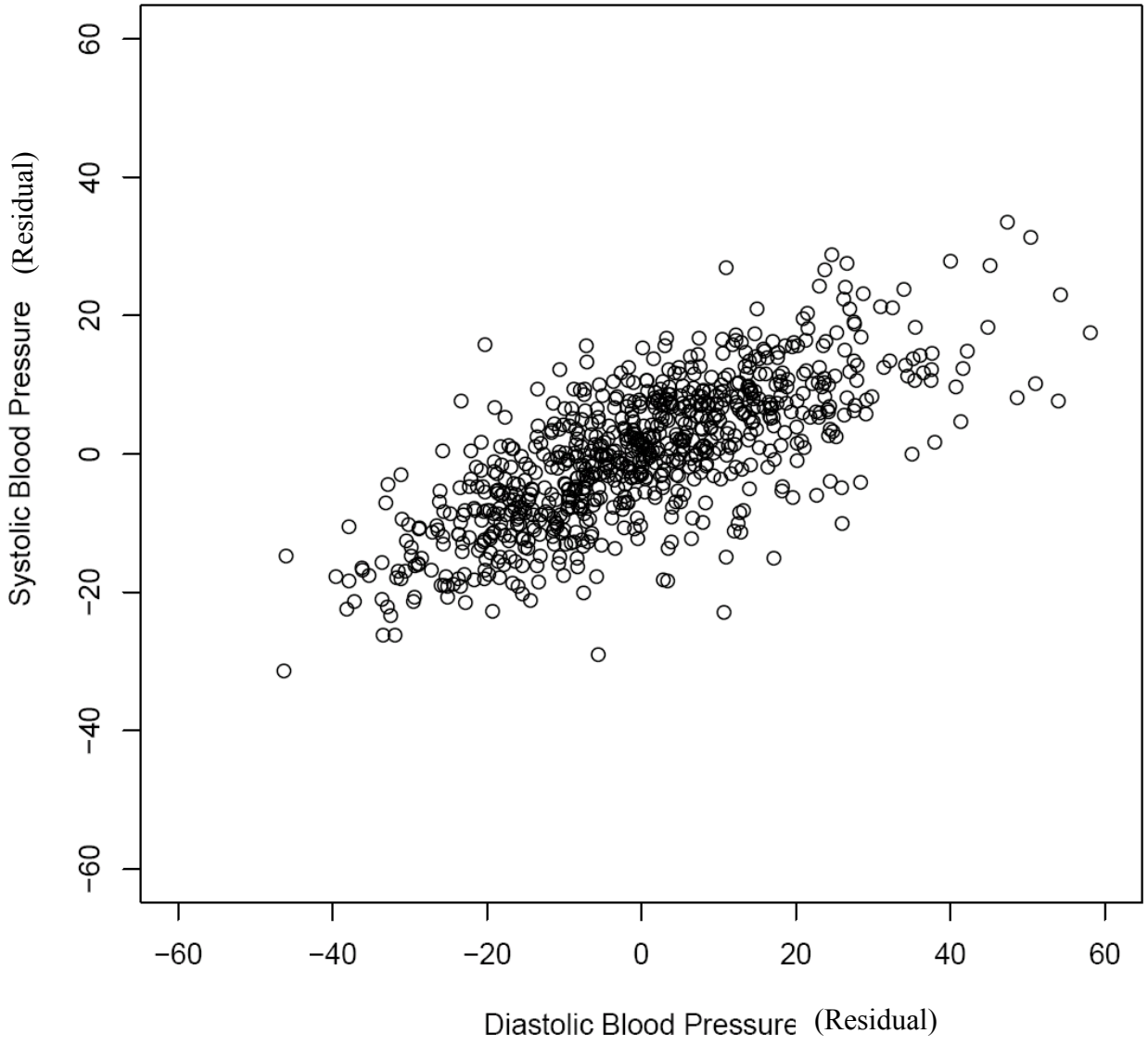
Bold:$p<0.05$

Table 2.7. Multivariable linear mixed model analysis of the effects of age and gender on expression level in the 30 BMI-related genes after adjusting for batch effect.

Gene	Chr	Age			Gender (Female=1, Male=0)		
		Beta	P-value	R^2 %	Beta	P-value	R^2 %
ADCY3	2	-0.0013	0.250	0.113	-0.02	0.308	0.089
ATP2A1	16	0.0008	0.107	0.289	0.00	0.773	0.009
BCDIN3D	12	-0.0004	0.549	0.043	0.01	0.198	0.201
BDNF	11	0.0013	0.064	0.407	-0.01	0.225	0.174
CADM2	3	0.0011	0.285	0.130	-0.01	0.579	0.035
CHST8	19	0.0006	0.412	0.076	0.00	0.914	0.001
DGKG	3	0.0013	0.234	0.165	-0.03	0.134	0.261
DNAJC27	2	-0.0007	0.427	0.067	0.02	0.223	0.159
ETV5	3	0.0062	**0.019**	0.642	-0.03	0.508	0.051
FAIM2	12	0.0005	0.596	0.034	0.00	0.866	0.003
FTO	16	-0.0007	0.182	0.197	0.02	**0.009**	0.762
GPRC5B	16	0.0005	0.587	0.032	-0.02	0.188	0.189
KCTD15	19	0.0017	**0.044**	0.490	-0.01	0.505	0.053
LINGO2	9	-0.0009	0.316	0.121	0.00	0.769	0.010
LRP1B	2	-0.0004	0.556	0.040	-0.01	0.390	0.085
MAP2K5	15	-0.0031	**0.002**	1.023	-0.02	0.325	0.101
MTCH2	11	0.0011	0.173	0.136	0.00	0.856	0.002
MTIF3	13	-0.0011	0.160	0.235	0.01	0.584	0.036
NEGR1	1	0.0001	0.876	0.003	0.00	0.967	0.000
NRXN3	14	0.0006	0.826	0.006	-0.03	0.533	0.050
NUDT3	6	-0.0017	**0.020**	0.472	0.00	0.958	0.000
POC5	5	0.0002	0.797	0.006	0.01	0.450	0.050
QPCTL	19	-0.0007	0.370	0.098	-0.02	0.144	0.260
SEC16B	1	0.0013	0.126	0.283	-0.01	0.525	0.049
SH2B1	16	0.0002	0.757	0.009	0.00	0.925	0.001
SLC39A8	4	-0.0010	0.347	0.106	-0.04	**0.026**	0.594
STK33	11	0.0026	0.105	0.289	0.02	0.431	0.068
TFAP2B	6	0.0050	**0.028**	0.580	0.01	0.859	0.004
TMEM18	2	-0.0010	0.208	0.175	-0.02	0.160	0.219
TNNI3K	1	-0.0006	0.455	0.064	0.00	0.771	0.010

Bold:$p<0.05$

Figure 2.4. Correlation of systolic blood pressure and diastolic blood pressure after adjusting for age, age^2, gender, and BMI.



* Person correlation coefficient=0.728.

Table 2.8a. Pearson correlation coefficients among expression levels of selected BP-related genes after adjusting for age, gender and batch.

		MTHFR	*CAPZA1*	*MOV10*	*SLC4A7*	*ULK4*	*MECOM*	*PRDM8*	*FGF5*	*SLC39A8*	*GUCY1A3*	*NPR3*	*HFE*	*BAG6*	*CACNB2*
MTHFR	**r**	1													
	p	NA													
CAPZA1	**r**	0.1234	1												
	p	0.0005	NA												
MOV10	**r**	0.2931	0.1481	1											
	p	<.0001	<.0001	NA											
SLC4A7	**r**	0.1441	0.3580	0.2582	1										
	p	<.0001	<.0001	<.0001	NA										
ULK4	**r**	0.0233	0.2622	0.0184	0.0506	1									
	p	0.5137	<.0001	0.6056	0.1556	NA									
MECOM	**r**	-0.0974	-0.3388	-0.2373	-0.4749	-0.1207	1								
	p	0.0062	<.0001	<.0001	<.0001	0.0007	NA								
PRDM8	**r**	-0.1823	0.0000	0.1031	-0.2640	-0.1052	0.4295	1							
	p	<.0001	<.0001	0.0038	<.0001	0.0031	<.0001	NA							
FGF5	**r**	-0.0285	-0.0572	-0.0538	-0.1939	0.0303	0.2724	0.2386	1						
	p	0.4247	0.1082	0.1314	<.0001	0.3960	<.0001	<.0001	NA						
SLC39A8	**r**	0.1755	0.2215	0.1277	0.1578	0.1544	-0.1137	0.1805	-0.0542	1					
	p	<.0001	<.0001	0.0003	<.0001	<.0001	0.0014	<.0001	0.1280	NA					
GUCY1A3	**r**	0.0309	-0.0506	0.2450	0.0366	-0.0824	-0.0299	0.0994	0.0530	-0.1213	1				
	p	0.3863	0.1557	<.0001	0.3050	0.0206	0.4020	0.0052	0.1368	0.0006	NA				
NPR3	**r**	-0.0582	-0.2210	-0.1549	-0.3408	-0.1398	0.4237	0.2934	0.1646	-0.1198	0.2779	1			
	p	0.1026	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0007	<.0001	NA			
HFE	**r**	0.0090	-0.1892	0.1255	-0.1279	-0.1000	0.3466	0.2921	0.1052	-0.0067	0.0160	0.0956	1		
	p	0.8012	<.0001	0.0004	0.0003	0.0049	<.0001	<.0001	0.0031	0.8501	0.6536	0.0072	NA		
BAG6	**r**	0.2000	0.1185	0.2763	0.1081	0.0914	-0.0078	0.1236	0.0362	0.1862	-0.0346	-0.0199	0.0247	1	
	p	<.0001	0.0009	<.0001	0.0024	0.0102	0.8277	0.0005	0.3100	<.0001	0.3319	0.5763	0.4883	NA	
CACNB2	**r**	-0.1616	-0.0903	0.1734	0.2660	-0.3778	-0.0751	0.2340	-0.0501	-0.0614	0.1032	-0.1136	0.2034	-0.0337	1
	p	<.0001	0.0112	<.0001	<.0001	<.0001	0.0349	<.0001	0.1597	0.0846	0.0037	0.0014	<.0001	0.3440	NA

r = Pearson correlation coefficient. p=p-value. Red color= r > 0.3 or r < -0.3.

Table 2.8b. Pearson correlation coefficients among expression levels of selected BP-related genes after adjusting for age, gender and batch.

		MTHFR	*CAPZA1*	*MOV10*	*SLC4A7*	*ULK4*	*MECOM*	*PRDM8*	*FGF5*	*SLC39A8*	*GUCY1A3*	*NPR3*	*HFE*	*BAG6*	*CACNB2*
PLCE1	**r**	-0.2124	-0.3372	-0.3924	**-0.5036**	-0.1871	**0.5153**	0.1968	0.1814	-0.1989	-0.1468	0.2613	0.0498	-0.1787	-0.0926
	p	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.1625	<.0001	0.0093
CNNM2	**r**	0.1477	-0.0626	-0.1521	-0.1239	-0.1707	0.0512	-0.1058	0.0063	-0.0808	-0.0976	0.1154	-0.0841	-0.1341	-0.0403
	p	<.0001	0.0788	<.0001	0.0005	<.0001	0.1509	0.0029	0.8603	0.0233	0.0061	0.0012	0.0181	0.0002	0.2586
NT5C2	**r**	-0.0584	0.3688	0.1240	0.2270	0.1478	-0.3055	-0.1403	-0.0521	0.0004	-0.0121	-0.2273	-0.1785	-0.1334	0.0750
	p	0.1011	<.0001	0.0005	<.0001	<.0001	<.0001	<.0001	0.1437	0.9912	0.7348	<.0001	<.0001	0.0002	0.0353
ADM	**r**	0.0226	-0.1687	-0.2751	-0.1747	-0.1770	0.1960	0.0252	-0.0100	0.1145	-0.3714	-0.0040	-0.0418	-0.0025	-0.0643
	p	0.5255	<.0001	<.0001	<.0001	<.0001	<.0001	0.4803	0.7784	0.0013	<.0001	0.9114	0.2408	0.9445	0.0710
PLEKHA7	**r**	0.1066	-0.0788	-0.2169	-0.1491	-0.1052	0.0764	-0.0808	-0.0715	0.1238	-0.3803	-0.1467	-0.0261	0.0138	-0.0281
	p	0.0027	0.0269	<.0001	<.0001	0.0031	0.0318	0.0233	0.0446	0.0005	<.0001	<.0001	0.4636	0.6993	0.4307
ARHGAP42	**r**	-0.0819	0.1021	-0.1503	-0.0877	0.1627	-0.0609	-0.2061	-0.0128	-0.0150	-0.1447	-0.1638	-0.0502	-0.0559	-0.1156
	p	0.0215	0.0041	<.0001	0.0137	<.0001	0.0875	<.0001	0.7190	0.6733	<.0001	<.0001	0.1591	0.1165	0.0011
ATP2B1	**r**	0.1031	0.2682	0.0735	0.2538	0.1030	-0.3668	-0.3383	-0.1025	0.0582	-0.2749	-0.3656	-0.1219	-0.1046	-0.0208
	p	0.0037	<.0001	0.0392	<.0001	0.0038	<.0001	<.0001	0.0039	0.1022	<.0001	<.0001	0.0006	0.0033	0.5604
SH2B3	**r**	-0.0996	-0.0570	0.1299	-0.2681	0.0701	0.0727	0.1958	0.0565	-0.1485	-0.0047	0.0478	-0.0168	0.1961	-0.0154
	p	0.0051	0.1099	0.0003	<.0001	0.0490	0.0411	<.0001	0.1131	<.0001	0.8954	0.1797	0.6373	<.0001	0.6661
CSK	**r**	0.0508	-0.2195	0.2599	0.0572	-0.2089	-0.0912	0.0792	-0.0559	-0.0370	0.3840	0.1028	0.0909	0.2305	0.2175
	p	0.1538	<.0001	<.0001	0.1084	<.0001	0.0104	0.0261	0.1169	0.2998	<.0001	0.0039	0.0107	<.0001	<.0001
FES	**r**	-0.1517	-0.3674	-0.0929	-0.3943	-0.1724	0.3759	0.3480	0.0766	-0.1236	0.0616	0.3132	0.0950	-0.0712	-0.0173
	p	<.0001	<.0001	0.0090	<.0001	<.0001	<.0001	<.0001	0.0314	0.0005	0.0839	<.0001	0.0076	0.0455	0.6274
PLCD3	**r**	-0.2620	-0.4722	-0.2309	-0.4312	-0.3745	0.4366	0.3712	0.1145	-0.2508	0.0694	0.3092	0.1257	-0.1595	0.1417
	p	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0013	<.0001	0.0515	<.0001	0.0004	<.0001	<.0001
GOSR2	**r**	0.0899	0.2505	0.0630	0.1516	0.2212	-0.3750	-0.2666	-0.0971	0.1907	-0.0829	-0.3248	-0.1650	0.0236	-0.0241
	p	0.0116	<.0001	0.0769	<.0001	<.0001	<.0001	<.0001	0.0063	<.0001	0.0198	<.0001	<.0001	0.5080	0.4994
ZNF652	**r**	0.1281	0.2120	0.0563	0.0680	0.0953	-0.1409	-0.1891	-0.0897	-0.1657	0.0344	-0.0950	-0.0572	-0.0112	-0.0051
	p	0.0003	<.0001	0.1143	0.0563	0.0074	<.0001	<.0001	0.0118	<.0001	0.3342	0.0076	0.1086	0.7542	0.8858

r = Pearson correlation coefficient. p=p-value. Red color= r > 0.3 or r < -0.3. Bold= r > 0.5 or r < -0.5.

Table 2.8c. Pearson correlation coefficients among expression levels of selected BP-related genes after adjusting for age, gender and batch.

		PLCE1	*CNNM2*	*NT5C2*	*ADM*	*PLEKHA7*	*ARHGAP42*	*ATP2B1*	*SH2B3*	*CSK*	*FES*	*PLCD3*	*GOSR2*	*ZNF652*
PLCE1	**r**	1												
	p	NA												
CNNM2	**r**	0.1920	1											
	p	<.0001	NA											
NT5C2	**r**	-0.3300	0.0265	1										
	p	<.0001	0.4567	NA										
ADM	**r**	0.3318	0.3016	-0.2106	1									
	p	<.0001	<.0001	<.0001	NA									
PLEKHA7	**r**	0.2184	0.1596	-0.2230	0.4780	1								
	p	<.0001	<.0001	<.0001	<.0001	NA								
ARHGAP42	**r**	0.0028	-0.0131	0.1575	0.0829	0.1929	1							
	p	0.9367	0.7125	<.0001	0.0199	<.0001	NA							
ATP2B1	**r**	-0.3453	-0.0086	**0.5189**	-0.0523	0.0637	0.1926	1						
	p	<.0001	0.8103	<.0001	0.1419	0.0736	<.0001	NA						
SH2B3	**r**	0.0551	-0.1656	0.1326	-0.3014	-0.1923	-0.1286	0.0347	1					
	p	0.1222	<.0001	0.0002	<.0001	<.0001	0.0003	0.3299	NA					
CSK	**r**	-0.1427	-0.0038	-0.1529	-0.1711	-0.1779	-0.1109	-0.2467	0.1071	1				
	p	<.0001	0.9164	<.0001	<.0001	<.0001	0.0018	<.0001	0.0026	NA				
FES	**r**	**0.5002**	-0.0705	-0.2564	0.1227	-0.0674	-0.1722	-0.4340	0.1944	0.0424	1			
	p	<.0001	0.0477	<.0001	0.0006	0.0583	<.0001	<.0001	<.0001	0.2343	NA			
PLCD3	**r**	**0.6338**	0.0319	-0.4307	0.2142	0.0834	-0.1911	**-0.5027**	0.1538	0.1112	**0.6318**	1		
	p	<.0001	0.3716	<.0001	<.0001	0.0191	<.0001	<.0001	<.0001	0.0018	<.0001	NA		
GOSR2	**r**	-0.3766	0.0794	0.3315	-0.0532	0.0881	0.2607	0.3438	-0.0676	0.0788	-0.4575	**-0.5126**	1	
	p	<.0001	0.0258	<.0001	0.1355	0.0133	<.0001	<.0001	0.0575	0.0270	<.0001	<.0001	NA	
ZNF652	**r**	-0.1692	0.1590	0.4483	-0.1545	-0.1771	0.1754	0.3628	0.0658	-0.0529	-0.2337	-0.3335	0.2165	1
	p	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0645	0.1376	<.0001	<.0001	<.0001	NA

r = Pearson correlation coefficient. p=p-value. Red color= r > 0.3 or r < -0.3. Bold= r > 0.5 or r < -0.5.

Table 2.9a. Pearson correlation coefficients among expression levels of selected BMI-related genes after adjusting for age, gender and batch.

		NEGR1	*TNNI3K*	*SEC16B*	*TMEM18*	*ADCY3*	*DNAJC27*	*LRP1B*	*CADM2*	*ETV5*	*DGKG*	*SLC39A8*	*POC5*	*NUDT3*	*TFAP2B*	*LINGO2*
NEGR1	**r**	1														
	p	NA														
TNNI3K	**r**	0.2188	1													
	p	<.0001	NA													
SEC16B	**r**	**0.6300**	0.2311	1												
	p	<.0001	<.0001	NA												
TMEM18	**r**	-0.0947	0.1008	-0.0644	1											
	p	0.0078	0.0046	0.0707	NA											
ADCY3	**r**	-0.0978	0.1027	-0.0930	0.0825	1										
	p	0.0060	0.0039	0.0090	0.0205	NA										
DNAJC27	**r**	-0.3639	0.1622	-0.3185	0.1547	0.1132	1									
	p	<.0001	<.0001	<.0001	<.0001	0.0014	NA									
LRP1B	**r**	**0.7758**	0.3472	**0.7094**	0.0047	-0.0853	-0.3555	1								
	p	<.0001	<.0001	<.0001	0.8946	0.0166	<.0001	NA								
CADM2	**r**	**0.6414**	0.1478	**0.5681**	-0.0631	-0.0574	-0.3838	**0.7041**	1							
	p	<.0001	<.0001	<.0001	0.0764	0.1075	<.0001	<.0001	NA							
ETV5	**r**	-0.1440	-0.1142	-0.1272	-0.0449	0.1120	0.1117	-0.1484	-0.1265	1						
	p	<.0001	0.0013	0.0003	0.2082	0.0016	0.0017	<.0001	0.0004	NA						
DGKG	**r**	0.3071	0.2644	0.2887	-0.1320	0.0214	-0.1040	0.3470	0.3063	0.1901	1					
	p	<.0001	<.0001	<.0001	0.0002	0.5480	0.0034	<.0001	<.0001	<.0001	NA					
SLC39A8	**r**	-0.1669	0.0549	-0.1516	0.2297	0.1040	0.2387	-0.1240	-0.1316	0.0644	0.0265	1				
	p	<.0001	0.1232	<.0001	<.0001	0.0035	<.0001	0.0005	0.0002	0.0704	0.4582	NA				
POC5	**r**	**-0.5134**	-0.1347	-0.4870	0.2628	0.0705	**0.5166**	**-0.5279**	**-0.5080**	0.1314	-0.3172	0.2269	1			
	p	<.0001	0.0001	<.0001	<.0001	0.0476	<.0001	<.0001	<.0001	0.0002	<.0001	<.0001	NA			
NUDT3	**r**	-0.3425	-0.1717	-0.3364	0.2253	-0.0320	0.2318	-0.3595	-0.2908	0.0561	-0.2167	0.3002	0.4329	1		
	p	<.0001	<.0001	<.0001	<.0001	0.3696	<.0001	<.0001	<.0001	0.1153	<.0001	<.0001	<.0001	NA		
TFAP2B	**r**	0.1695	-0.0418	0.1884	0.0072	-0.0725	-0.0614	0.1895	0.1206	0.2837	-0.1008	-0.0553	-0.0560	-0.0699	1	
	p	<.0001	0.2411	<.0001	0.8401	0.0417	0.0849	<.0001	0.0007	<.0001	0.0046	0.1207	0.1157	0.0496	NA	
LINGO2	**r**	0.4061	0.0904	0.3295	-0.0692	-0.0347	-0.3245	0.4375	0.3825	-0.1801	0.1857	-0.1654	-0.4553	-0.3545	-0.0223	1
	p	<.0001	0.0111	<.0001	0.0521	0.3305	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.5326	NA

r = Pearson correlation coefficient. p=p-value. Red color= r > 0.3 or r < -0.3. Bold= r > 0.5 or r < -0.5.

Table 2.9b. Pearson correlation coefficients among expression levels of selected BMI-related genes after adjusting for age, gender and batch.

		NEGR1	*TNNI3K*	*SEC16B*	*TMEM18*	*ADCY3*	*DNAJC27*	*LRP1B*	*CADM2*	*ETV5*	*DGKG*	*SLC39A8*	*POC5*	*NUDT3*	*TFAP2B*	*LINGO2*
STK33	**r**	-0.2775	-0.1931	-0.2720	-0.0780	0.0529	0.1717	-0.3419	-0.2925	0.1678	-0.1788	-0.0907	0.2616	0.2111	0.0282	-0.2374
	p	<.0001	<.0001	<.0001	0.0285	0.1379	<.0001	<.0001	<.0001	<.0001	<.0001	0.0108	<.0001	<.0001	0.4293	<.0001
BDNF	**r**	0.4595	0.0430	0.4814	-0.1067	-0.0224	-0.3522	0.4765	0.4638	-0.0339	0.1430	-0.2531	-0.4923	-0.3762	0.2190	0.4480
	p	<.0001	0.2280	<.0001	0.0027	0.5307	<.0001	<.0001	<.0001	0.3417	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
MTCH2	**r**	-0.4085	-0.2113	-0.3892	0.0636	0.0553	0.0640	-0.4485	-0.3630	0.0126	-0.1928	0.0522	0.2971	0.4530	-0.1319	-0.2852
	p	<.0001	<.0001	<.0001	0.0743	0.1206	0.0726	<.0001	<.0001	0.7245	<.0001	0.1432	<.0001	<.0001	0.0002	<.0001
BCDIN3D	**r**	0.1886	0.0022	0.1459	-0.0592	-0.1245	-0.2297	0.1877	0.1580	-0.2002	0.0314	-0.0617	-0.2559	-0.0975	-0.0525	0.2479
	p	<.0001	0.9516	<.0001	0.0965	0.0005	<.0001	<.0001	<.0001	<.0001	0.3779	0.0831	<.0001	0.0061	0.1408	<.0001
FAIM2	**r**	**0.6031**	0.0814	**0.5153**	-0.1422	-0.1174	-0.4867	**0.6137**	**0.6300**	-0.2047	0.2600	-0.1996	**-0.6006**	-0.4242	0.0779	**0.5569**
	p	<.0001	0.0223	<.0001	<.0001	0.0010	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.0286	<.0001
MTIF3	**r**	**-0.5560**	-0.1431	-0.4777	0.1228	-0.1558	0.2916	**-0.5749**	**-0.5288**	0.0047	-0.3317	0.1759	0.4199	0.2657	-0.0987	-0.3377
	p	<.0001	<.0001	<.0001	0.0005	<.0001	<.0001	<.0001	<.0001	0.8952	<.0001	<.0001	<.0001	<.0001	0.0055	<.0001
NRXN3	**r**	-0.0385	0.0654	-0.0265	-0.0620	0.1271	-0.0422	-0.0392	0.0218	0.0589	0.1161	0.1711	-0.1483	-0.0429	-0.1010	0.0420
	p	0.2806	0.0664	0.4575	0.0820	0.0003	0.2361	0.2715	0.5402	0.0983	0.0011	<.0001	<.0001	0.2292	0.0045	0.2381
MAP2K5	**r**	-0.3236	0.0350	-0.3352	-0.0010	-0.1764	0.1889	-0.3657	-0.3583	-0.1816	-0.0407	0.1025	0.1909	0.2282	-0.4156	-0.1650
	p	<.0001	0.3258	<.0001	0.9770	<.0001	<.0001	<.0001	<.0001	<.0001	0.2538	0.0040	<.0001	<.0001	<.0001	<.0001
GPRC5B	**r**	0.4920	0.1333	0.4752	-0.1302	0.0409	-0.3990	0.4766	**0.5274**	-0.1308	0.2641	-0.1977	**-0.5481**	-0.3965	0.1012	0.3783
	p	<.0001	0.0002	<.0001	0.0002	0.2508	<.0001	<.0001	<.0001	0.0002	<.0001	<.0001	<.0001	<.0001	0.0044	<.0001
SH2B1	**r**	0.4586	0.0313	0.3620	-0.2563	-0.0749	**-0.5020**	0.4358	0.4276	-0.2137	0.2124	-0.1942	**-0.6005**	-0.3657	0.0077	0.4316
	p	<.0001	0.3795	<.0001	<.0001	0.0354	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.8296	<.0001
ATP2A1	**r**	0.3989	0.1054	0.3315	-0.1439	-0.0780	-0.3861	0.4002	0.4379	-0.1672	0.2520	-0.1589	-0.4627	-0.3754	-0.0128	0.4424
	p	<.0001	0.0030	<.0001	<.0001	0.0285	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.7205	<.0001
FTO	**r**	-0.0871	0.0190	-0.0882	-0.2244	0.0310	0.0498	-0.1174	-0.1565	-0.0157	-0.0023	-0.0222	-0.0078	0.0171	-0.0137	-0.0692
	p	0.0144	0.5950	0.0132	<.0001	0.3841	0.1624	0.0010	<.0001	0.6598	0.9494	0.5333	0.8263	0.6308	0.7008	0.0520
CHST8	**r**	**0.5297**	0.0063	0.4658	-0.1382	-0.0959	-0.4953	**0.5246**	**0.5851**	-0.2235	0.2591	-0.2394	**-0.6078**	-0.3962	0.0487	**0.5094**
	p	<.0001	0.8607	<.0001	<.0001	0.0070	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.1721	<.0001
KCTD15	**r**	**0.5262**	-0.0424	0.4578	-0.1027	-0.0719	-0.4782	**0.5263**	**0.5475**	-0.1145	0.1829	-0.2279	**-0.5286**	-0.3646	0.1497	0.4482
	p	<.0001	0.2347	<.0001	0.0039	0.0436	<.0001	<.0001	<.0001	0.0013	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
QPCTL	**r**	0.1128	0.0645	0.0541	0.0303	-0.1685	-0.2893	0.1374	0.1085	-0.1345	0.1415	0.0255	-0.2152	-0.0532	-0.1182	0.2030
	p	0.0015	0.0702	0.1286	0.3948	<.0001	<.0001	0.0001	0.0023	0.0002	<.0001	0.4748	<.0001	0.1351	0.0009	<.0001

r = Pearson correlation coefficient. p=p-value. Red color= r > 0.3 or r < -0.3. Bold= r > 0.5 or r < -0.5.

Table 2.9c. Pearson correlation coefficients among expression levels of selected BMI-related genes after adjusting for age, gender and batch.

		STK33	*BDNF*	*MTCH2*	*BCDIN3D*	*FAIM2*	*MTIF3*	*NRXN3*	*MAP2K5*	*GPRC5B*	*SH2B1*	*ATP2A1*	*FTO*	*CHST8*	*KCTD15*	*QPCTL*
STK33	**r**	1														
	p	NA														
BDNF	**r**	-0.2047	1													
	p	<.0001	NA													
MTCH2	**r**	0.3475	-0.3591	1												
	p	<.0001	<.0001	NA												
BCDIN3D	**r**	-0.1376	0.2234	-0.1505	1											
	p	0.0001	<.0001	<.0001	NA											
FAIM2	**r**	-0.3406	**0.6463**	-0.4732	0.3362	1										
	p	<.0001	<.0001	<.0001	<.0001	NA										
MTIF3	**r**	0.1441	-0.4638	0.3070	-0.0831	-0.4986	1									
	p	<.0001	<.0001	<.0001	0.0196	<.0001	NA									
NRXN3	**r**	0.0154	-0.0454	0.0725	0.0015	0.0289	-0.0935	1								
	p	0.6660	0.2025	0.0418	0.9663	0.4174	0.0086	NA								
MAP2K5	**r**	-0.0084	-0.3909	0.1760	0.0275	-0.2972	0.3856	0.1719	1							
	p	0.8134	<.0001	<.0001	0.4399	<.0001	<.0001	<.0001	NA							
GPRC5B	**r**	-0.2695	**0.5869**	-0.3499	0.1855	**0.6810**	-0.4489	-0.0203	-0.3710	1						
	p	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.5685	<.0001	NA						
SH2B1	**r**	-0.2761	0.4318	-0.3112	0.3781	**0.6036**	-0.3783	0.0541	-0.1332	0.4412	1					
	p	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.1287	0.0002	<.0001	NA					
ATP2A1	**r**	-0.3365	0.4658	-0.4013	0.2746	**0.6496**	-0.3522	-0.0386	-0.1490	**0.5030**	**0.5594**	1				
	p	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.2794	<.0001	<.0001	<.0001	NA				
FTO	**r**	0.2219	-0.0813	0.2165	-0.1299	-0.1810	-0.0143	-0.0084	0.0404	-0.1332	0.0195	-0.1073	1			
	p	<.0001	0.0223	<.0001	0.0003	<.0001	0.6891	0.8138	0.2565	0.0002	0.5844	0.0025	NA			
CHST8	**r**	-0.3207	**0.6088**	-0.3677	0.3154	**0.8151**	-0.4632	0.0161	-0.2868	**0.6724**	**0.5740**	**0.6231**	-0.1760	1		
	p	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.6527	<.0001	<.0001	<.0001	<.0001	<.0001	NA		
KCTD15	**r**	-0.2772	**0.6320**	-0.3848	0.2971	**0.7525**	-0.4957	-0.0428	-0.3842	**0.6277**	**0.5726**	**0.5854**	-0.1851	**0.7519**	1	
	p	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	0.2303	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	NA	
QPCTL	**r**	-0.2357	0.0501	-0.0572	0.3139	0.1982	0.0327	0.0437	0.1544	0.0634	0.3648	0.2630	-0.1081	0.1912	0.1327	1
	p	<.0001	0.1596	0.1086	<.0001	<.0001	0.3595	0.2201	<.0001	0.0753	<.0001	<.0001	0.0024	<.0001	0.0002	NA

r = Pearson correlation coefficient. p=p-value. Red color= r > 0.3 or r < -0.3. Bold= r > 0.5 or r < -0.5.

Table 2.10. Heritability for systolic blood pressure, diastolic blood pressure, and BMI before and after adjusting for covariates.

	Systolic Blood Pressure (SBP)†		Diastolic Blood Pressure (DBP)†		BMI	
	h^2	P-value	h^2	P-value	h^2	P-value
Raw	0.223	4.78×10^{-3}	0.247	1.62×10^{-3}	0.464	2.96×10^{-17}
Adjusted‡	0.352	3.64×10^{-5}	0.312	1.20×10^{-4}	0.472	9.18×10^{-18}
Proportion of variance due to all covariates	20.14%		8.85%		0.21%	
Proportion of total variance due to additive genetic variance*	0.281		0.284		0.471	

†SBP=SBP+15mm Hg for taking hypertension drug.
DBP=DBP+10mm Hg for taking hypertension drug.

‡Adjusted heritability for SBP and DBP were estimated by including age, age^2, gender, and BMI in the statistical model in the SOLAR analysis.Adjusted heritability for BMI was estimated by including age and gender in the statistical model in the SOLAR analysis.

*Proportion of total variance due to additive genetic variance for each phenotype was estimated as follows: SBP=0.352x(1-0.2014); DBP=0.312x(1-0.085);BMI=0.472x(1-0.0021).

Table 2.11. Heritability of gene expression levels for the 27 BP-related genes after adjusting for age, gender and batch.

Gene	Chr	h^2	h^2 P-value
ULK4	3	0.713	**3.93E-14**
CNNM2	10	0.343	**2.17E-05**
CACNB2	10	0.315	**3.66E-05**
BAG6	6	0.330	**1.38E-04**
CAPZA1	1	0.294	**1.57E-04**
MOV10	1	0.302	**1.98E-04**
ZNF652	17	0.278	**2.61E-04**
NT5C2	10	0.272	**9.70E-04**
MTHFR	1	0.270	**1.86E-03**
SLC4A7	3	0.265	**1.88E-03**
PLEKHA7	11	0.291	**2.67E-03**
FES	15	0.227	**3.61E-03**
CSK	15	0.215	**6.98E-03**
ARHGAP42	11	0.215	**7.20E-03**
ADM	11	0.212	**0.012**
GUCY1A3	4	0.210	**0.013**
PRDM8	4	0.168	**0.013**
FGF5	4	0.137	**0.028**
NPR3	5	0.102	0.127
HFE	6	0.070	0.178
SH2B3	12	0.074	0.181
PLCD3	17	0.071	0.209
MECOM	3	0.051	0.278
ATP2B1	12	0.029	0.352
SLC39A8	4	0.034	0.361
PLCE1	10	0.017	0.422
GOSR2	17	0.006	0.470

*18/27 (66.67%) genes have expression profiles that have statistically significant heritability ($p<0.05$).
Bold:$p<0.05$.

Table 2.12. Heritability of gene expression levels for the 30 BMI-related genes after adjusting for age, gender and batch.

Gene	Chr	h^2	h^2 P-value
STK33	11	0.5092	**1.00E-07**
ETV5	3	0.4277	**1.30E-06**
MTCH2	11	0.3227	**9.21E-05**
MAP2K5	15	0.2332	**0.001**
NUDT3	6	0.2885	**0.001**
DGKG	3	0.3028	**0.002**
TMEM18	2	0.2551	**0.005**
FAIM2	12	0.2251	**0.006**
ATP2A1	16	0.2177	**0.006**
TNNI3K	1	0.1615	**0.023**
GPRC5B	16	0.1649	**0.027**
TFAP2B	6	0.1430	**0.035**
ADCY3	2	0.1753	**0.037**
CADM2	3	0.1528	**0.042**
QPCTL	19	0.1331	0.066
POC5	5	0.0921	0.137
CHST8	19	0.0752	0.181
BDNF	11	0.0711	0.209
SH2B1	16	0.0540	0.274
NRXN3	14	0.0507	0.277
SEC16B	1	0.0530	0.280
KCTD15	19	0.0429	0.302
SLC39A8	4	0.0384	0.349
MTIF3	13	0.0071	0.432
FTO	16	0.0067	0.468
LINGO2	9	0.0061	0.474
BCDIN3D	12	0.0058	0.481
DNAJC27	2	0.0054	0.489
LRP1B	2	0.0051	0.491
NEGR1	1	0.0048	0.494

*14/30 (46.67%) genes have expression profiles that have statistically significant heritability ($p<0.05$).
Bold:$p<0.05$.

Appendix 2.1. Summary of twenty-nine blood pressure associated Index SNPs and their closest Genes. †

Index SNP	Nearest Gene	Chr	SNP Position	Gene Start	Gene End	Gene Size (bp)
Gene Region with Index SNP inside of it						
rs1799945	*HFE*	6	26,091,179	26,087,509	26,095,469	7,960
rs2521501	*FES*	15	91,437,388	91,427,665	91,439,006	11,341
rs805303	*BAT3*	6	31,616,366	31,606,805	31,620,170	13,365
rs17608766	*GOSR2*	17	45,013,271	45,000,486	45,018,732	18,246
rs17367504	*MTHFR*	1	11,862,778	11,845,787	11,866,115	20,328
rs12946454	*PLCD3*	17	43,208,121	43,189,009	43,209,891	20,882
rs1378942	*CSK*	15	75,077,367	75,074,425	75,095,539	21,114
rs3184504	*SH2B3*	12	111,884,608	111,843,752	111,889,426	45,674
rs13139571	*GUCY1A3*	4	156,645,513	156,587,862	156,658,212	70,350
rs12940887	*ZNF652*	17	47,402,807	47,366,569	47,439,835	73,266
rs13107325	*SLC39A8*	4	103,188,709	103,182,821	103,266,655	83,834
rs381815	*PLEKHA7*	11	16,902,268	16,809,212	17,035,963	226,751
rs633185	*ARHGAP42*	11	100,593,538	100,558,407	100,861,656	303,249
rs932764	*PLCE1*	10	95,895,940	95,753,746	96,088,146	334,400
rs1813353	*CACNB2*	10	18,707,448	18,429,606	18,830,688	401,082
rs419076	*MECOM*	3	169,100,886	168,801,287	169,381,563	580,276
rs3774372	*ULK4*	3	41,877,414	41,288,091	42,003,660	715,569
Gene Dense Region (Multiple Genes within ±100 kb of Index SNP)						
rs1458038	*FGF5*	4	81,164,723	81,187,742	81,212,170	24,428
	PRDM8	4	81,164,723	81,118,657	81,125,482	6,825
rs2932538	*MOV10*	1	113,216,543	113,217,048	113,243,367	26,319
	CAPZA1	1	113,216,543	113,162,075	113,214,241	52,166
rs11191548	*NT5C2*	10	104,846,178	104,847,774	104,953,063	105,289
	CNNM2	10	104,846,178	104,678,075	104,838,344	160,269
rs7129220	*ADM**	11	10,350,538	10,326,642	10,328,923	2,281
Gene Sparse Region (Only One Gene within ±100 kb of Index SNP)						
rs17249754	*ATP2B1*	12	90,060,586	89,981,828	90,049,844	68,016
rs1173771	*NPR3*	5	32,815,028	32,711,665	32,787,252	75,587
rs13082711	*SLC4A7*	3	27,537,909	27,414,215	27,498,245	84,030
Index SNPs Excluded (No gene within ±100 kb of Index SNP)						
rs10850411	*TBX3*	12	115,387,796	115,108,059	115,121,969	13,910
rs6015450	*EDN3*	20	57,751,117	57,875,499	57,901,047	25,548
rs1327235	*JAG1*	20	10,969,030	10,618,332	10,654,694	36,362
rs11953630	*EBF1*	5	157,845,402	158,122,924	158,526,788	403,864
SNPs Excluded (No gene expression data available)						
rs4590817	*C10orf107*	10	63,467,553	63,422,719	63,526,089	103,370

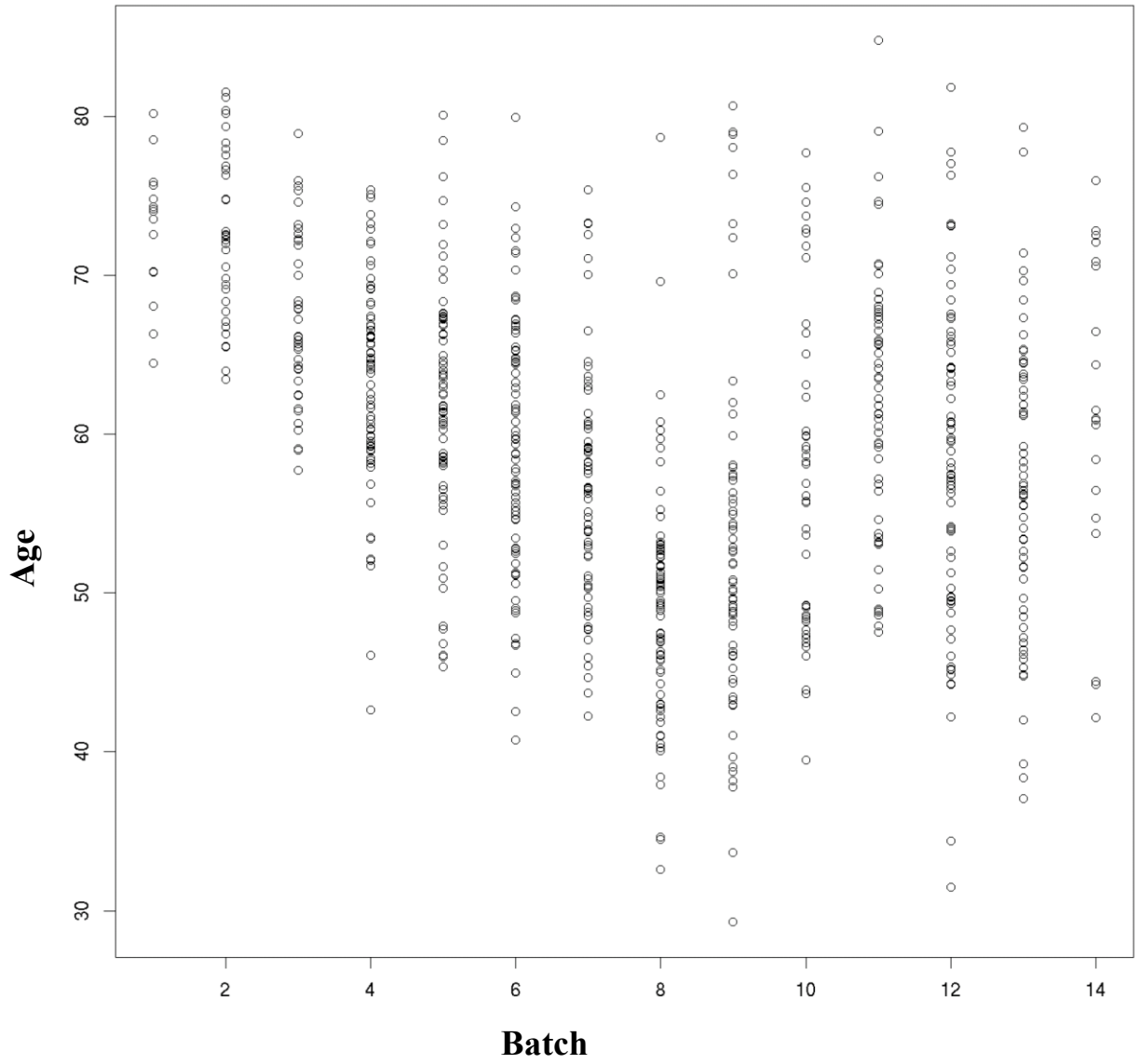
*There is the other one gene ,"*SBF2*", upstream of the index SNP but no genes downstream.
†Genes were sorted by gene size.

Appendix 2.2. Summary of thirty-two BMI associated Index SNPs and their closest Genes.

Index SNP	Nearest Gene	Chr	SNP Position	Gene Start	Gene End	Gene Size (bp)
Gene Region with Index SNP inside of it						
rs2287019	*QPCTL*	19	46,202,172	46,195,741	46,207,240	11,499
rs4771122	*MTIF3*	13	28,020,180	28,009,783	28,024,711	14,928
rs3817334	*MTCH2*	11	47,650,993	47,638,859	47,664,206	25,347
rs987237	*TFAP2B*	6	50,803,050	50,786,439	50,815,325	28,886
rs13107325	*SLC39A8*	4	103,188,709	103,182,821	103,266,655	83,834
rs206936	*NUDT3*	6	34,302,869	34,256,002	34,360,441	104,439
rs4929949	*STK33*	11	8,604,593	8,413,418	8,615,503	202,085
rs2241423	*MAP2K5*	15	68,086,838	67,835,021	68,099,451	264,430
rs13078807	*CADM2*	3	85,884,150	85,775,632	86,117,948	342,316
rs1514175	*TNNI3K*	1	74,991,644	74,663,926	75,010,108	346,182
rs1558902	*FTO*	16	53,803,574	53,737,875	54,148,378	410,503
rs10968576	*LINGO2*	9	28,414,339	27,948,528	28,719,303	770,775
rs10150332	*NRXN3*	14	79,936,964	78,870,093	80,330,758	1,460,665
Gene Dense Region (Multiple Genes within ±100 kb of Index SNP)						
rs7138803	*BCDIN3D*	12	50,247,468	50,229,826	50,236,912	7,086
	FAIM2	12	50,247,468	50,260,679	50,297,760	37,081
rs7359397	*SH2B1*	16	28,885,659	28,875,078	28,885,533	10,455
	ATP2A1	16	28,885,659	28,889,809	28,915,830	26,021
rs29941	*KCTD15*	19	34,309,532	34,287,751	34,306,665	18,914
	CHST8	19	34,309,532	34,112,861	34,264,414	151,553
rs713586	*DNAJC27*	2	25,158,008	25,166,505	25,194,824	28,319
	ADCY3	2	25,158,008	25,042,039	25,142,055	100,016
rs9816226	*ETV5*	3	185,834,499	185,764,108	185,826,901	62,793
	DGKG	3	185,834,499	185,864,990	186,080,023	215,033
rs12444979	*GPRC5B**	16	19,933,600	19,870,295	19,896,151	25,856
rs543874	*SEC16B***	1	177,889,480	177,897,489	177,939,050	41,561
Gene Sparse Region (Only One Gene within ±100 kb of Index SNP)						
rs2867125	*TMEM18*	2	622,827	667,975	677,439	9,464
rs2112347	*POC5*	5	75,015,242	74,970,024	75,013,313	43,289
rs10767664	*BDNF*	11	27,725,986	27,676,442	27,722,600	46,158
rs2815752	*NEGR1*	1	72,812,440	71,868,626	72,748,405	879,779
rs2890652	*LRP1B*	2	142,959,931	140,988,996	142,889,270	1,900,274
Index SNPs Excluded (No gene within ±100 kb of Index SNP)						
rs571312	*MC4R*	18	57,839,769	58,038,564	58,040,001	1,437
rs10938397	*GNPDA2*	4	45,182,527	44,704,169	44,728,612	24,443
rs1555543	*PTBP2*	1	96,944,797	97,187,175	97,280,599	93,424
rs4836133	*ZNF608*	5	124,332,103	123,972,610	124,080,865	108,255
rs11847697	*PRKD1*	14	30,515,112	30,045,689	30,396,899	351,210
SNPs Excluded (No gene expression data available)						
rs3810291	*ZC3H4*	19	47,569,003	47,567,447	47,617,009	49,562
rs887912	*FLJ30838*	2	59,302,877	58,747,888	59,290,901	543,013

*There is one gene ,"*IQCK*", upstream of the index SNP but no genes downstream.
**There is one gene ,"*LOC730102*", downstream of the index SNP but no genes upstream.

Appendix 2.3. Scatterplot of age and batch effect*.



*ANOVA test:*p*-value=$2x10^{-6}$

Appendix 2.4. Distribution of gender in each batch.

Batch	No. of Males	No of. Females	Total
1	9	6	15
2	14	19	33
3	19	19	38
4	20	53	73
5	31	46	77
6	34	44	78
7	20	46	66
8	36	44	80
9	26	39	65
10	17	28	45
11	30	30	60
12	45	33	78
13	23	39	62
14	7	12	19
Total	331	458	789

χ^2 test: p-value=0.0286

Appendix 2.5. Univariate linear mixed model analysis of the effects of age and gender on expression levels in the 27 BP-related genes.

Gene	Chr	Age		Gender (Female=1, Male=0)	
		Beta	P-value	Beta	P-value
ADM	11	-0.00413	**0.00249**	0.02868	0.27504
ARHGAP42	11	0.00455	**0.01348**	0.00869	0.80650
ATP2B1	12	0.00068	0.52284	-0.02529	0.23936
BAG6	6	0.00031	0.56477	-0.00099	0.92137
CACNB2	10	-0.00048	0.87752	0.02965	0.61099
CAPZA1	1	0.00285	**0.00005**	0.02037	0.12016
CNNM2	10	-0.00125	0.07218	0.01317	0.31408
CSK	15	0.00135	**0.03106**	0.00254	0.83328
FES	15	-0.00102	0.16613	-0.00748	0.59408
FGF5	4	0.00077	0.57303	0.01380	0.60519
GOSR2	17	0.00001	0.98794	-0.00044	0.97160
GUCY1A3	4	-0.00628	0.07493	0.06362	0.34912
HFE	6	0.00197	**0.00429**	-0.00874	0.52652
MECOM	3	0.00013	0.83674	-0.01659	0.17861
MOV10	1	0.00153	0.07039	0.00263	0.86946
MTHFR	1	-0.00280	**0.00031**	-0.03284	**0.02455**
NPR3	5	-0.00120	0.26789	-0.01746	0.41517
NT5C2	10	0.00371	**0.00089**	0.03668	0.07887
PLCD3	17	-0.00140	**0.02833**	0.00135	0.91351
PLCE1	10	-0.00102	0.06143	0.00615	0.57431
PLEKHA7	11	-0.00212	0.12198	-0.05609	**0.02947**
PRDM8	4	-0.00060	0.46882	-0.02486	0.12139
SH2B3	12	0.00078	0.28256	0.01565	0.27923
SLC39A8	4	0.00076	0.42678	-0.05471	**0.00442**
SLC4A7	3	0.00037	0.72566	-0.00506	0.79761
ULK4	3	0.00778	**1.88E-07**	-0.03599	0.15744
ZNF652	17	-0.00011	0.91519	0.00043	0.98298

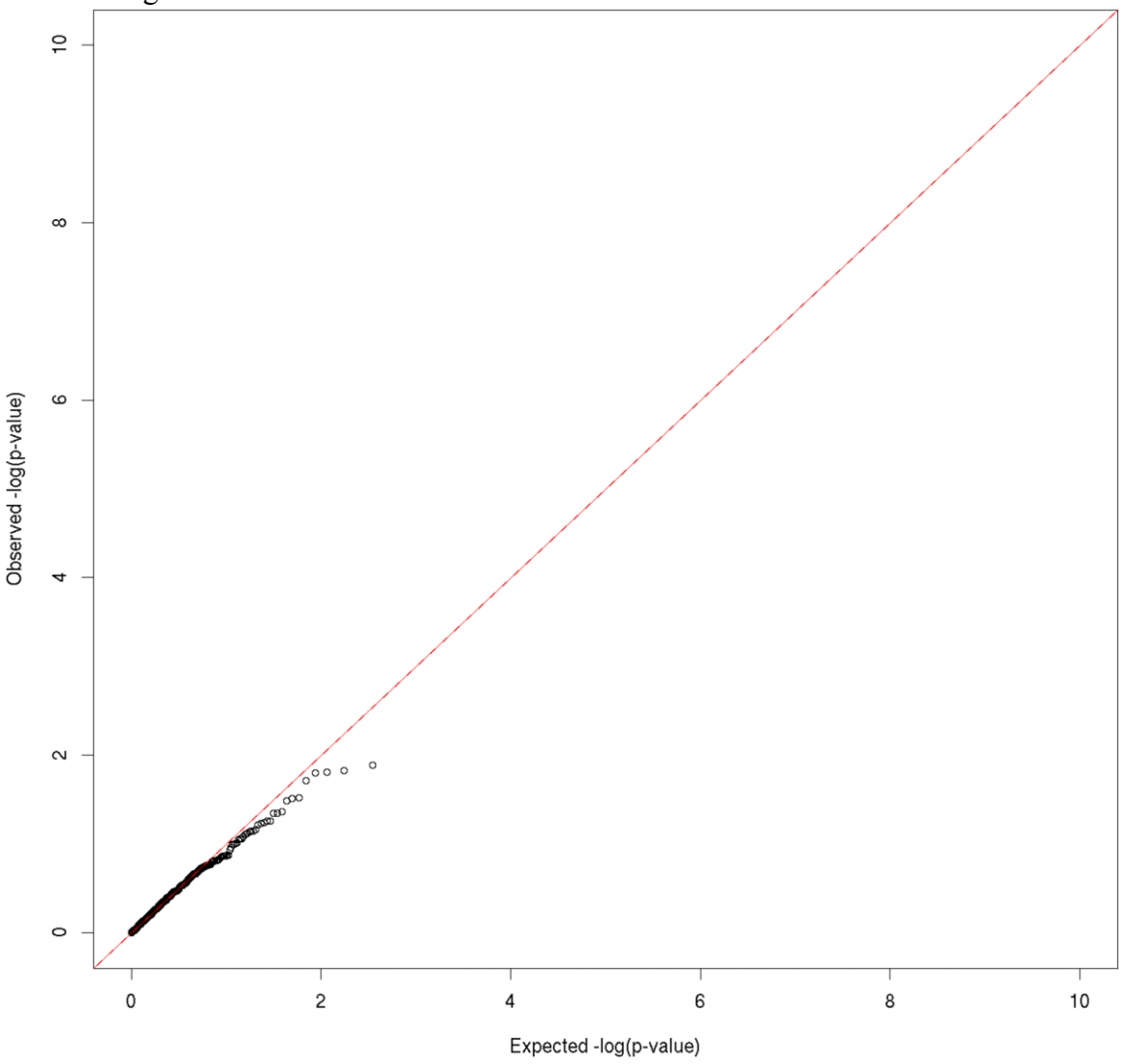
Bold:$p<0.05$

Appendix 2.6. Univariate linear mixed model analysis of the effects of age and gender on expression levels in the 30 BMI-related genes.

Gene	Chr	Age		Gender (Female=1, Male=0)	
		Beta	P-value	Beta	P-value
ADCY3	2	-0.0077	**4.89E-11**	-0.0150	0.500
ATP2A1	16	0.0005	0.266	-0.0119	0.189
BCDIN3D	12	-0.0016	**0.006**	0.0127	0.277
BDNF	11	0.0010	0.105	-0.0174	0.164
CADM2	3	-0.0010	0.285	-0.0125	0.495
CHST8	19	-0.0010	0.120	-0.0035	0.787
DGKG	3	0.0002	0.816	-0.0240	0.199
DNAJC27	2	0.0007	0.342	0.0315	**0.042**
ETV5	3	0.0078	**0.001**	-0.0341	0.435
FAIM2	12	-0.0001	0.888	0.0006	0.968
FTO	16	-0.0009	0.056	0.0134	0.160
GPRC5B	16	-0.0018	**0.040**	-0.0165	0.331
KCTD15	19	0.0007	0.336	-0.0129	0.360
LINGO2	9	-0.0010	0.179	-0.0103	0.499
LRP1B	2	-0.0008	0.169	-0.0122	0.302
MAP2K5	15	-0.0017	0.064	-0.0215	0.222
MTCH2	11	0.0057	**2.96E-09**	0.0059	0.739
MTIF3	13	-0.0012	0.070	0.0083	0.534
NEGR1	1	-0.0009	0.220	0.0005	0.975
NRXN3	14	-0.0010	0.660	-0.0178	0.705
NUDT3	6	0.0028	**1.69E-04**	0.0081	0.554
POC5	5	0.0043	**4.79E-06**	0.0282	0.119
QPCTL	19	0.0003	0.690	-0.0270	**0.037**
SEC16B	1	0.0004	0.627	-0.0126	0.403
SH2B1	16	-0.0004	0.444	-0.0110	0.309
SLC39A8	4	0.0008	0.427	-0.0547	**0.004**
STK33	11	0.0041	**0.005**	0.0313	0.245
TFAP2B	6	0.0084	**1.69E-05**	0.0021	0.956
TMEM18	2	0.0016	0.021	-0.0117	0.368
TNNI3K	1	0.0001	0.925	-0.0038	0.768

Bold:$p<0.05$

Appendix 2.7. Q-Q plot of the association between age and gene expression level for the 27 BP-related genes within each bacth.



*Total number of batches =14.

Chapter 3

Genetic Variants, Gene Expression and their association with Blood Pressure among non-Hispanic Whites in Genetic Epidemiology Network of Arteriopathy (GENOA)

Introduction

Hypertension affects approximately one-third of U.S adults [4] and is a major risk factor for a wide range of common chronic diseases associated with target-organ damage of the heart, brain, kidneys, and peripheral arteries.[3] In addition to the epidemiological risk factors, previous family studies and our study in chapter 2 have shown substantial evidence of heritability (30%-40%) for blood pressure.[5,13]

Recently, three large genome wide association study (GWAS) consortia have identified several new genomic regions where single nucleotide polymorphisms (SNPs) associated with blood pressure are highly significant ($p<5x10^{-8}$) and replicate across multiple studies in non-Hispanic whites.[14,16,17] The International Consortium for Blood Pressure Genome-Wide Association Studies (ICBP), which is the largest of the three blood pressure GWAS consortium including more than 200,000 non-Hispanic whites identified SNPs associated with blood pressure at 16 novel genetic loci. ICBP also replicated 12 of the 13 genetic loci reported from the other two consortia: Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) and Global Blood Pressure Genetics (Global BPgen) (Appendix 3.1).[17] Most of these identified SNPs fall in non-coding regions and their functional effects are not clear.

Gene expression levels in transformed beta-lymphocytes are influenced both by *cis* and *trans* genetic variants and are highly heritable.[39] Thus, investigating the genetic influence on the gene expression levels in these genomic regions may allow us to develop the knowledge of the potential functional mechanism at the molecular genetic level. In addition, examining whether the gene expression levels in these identified genomic regions are associated with blood pressure could help us to better understand about the combined effects of multiple DNA variations within an individual's regional haplotype that cannot be detected with a single SNP.

The goal of the present study is to investigate the functional genetic factors of blood pressure variation using a three stage approach (a) evaluating the associations between genetic variations in the significant and replicated genomic regions and blood pressure traits, (b) investigating the association between these genetic variations and proximal gene expression, and (c) examining the relationship between proximal gene expression and blood pressure traits in the non-Hispanic whites of the Genetic Epidemiology Network of Arteriopathy (GENOA) (Figure 3.1).

Methods

Study Population

The study population for this dissertation consists of non-Hispanic, white individuals from Rochester, Minnesota (MN) that were initially enrolled in the GENOA study. GENOA is a multicenter, community-based study of hypertensive sibships collected to identify genes influencing blood pressure levels and development of target organ damage due to hypertension.[41] In Rochester, MN, the Mayo Clinic diagnostic index and medical record linkage system of the Rochester Epidemiology Project were used to identify non-Hispanic white residents of Olmsted County.[43] In the first phase (Phase I:1995-2000) of GENOA (N=1583 non-Hispanic whites),

sibships containing at least two individuals with clinically diagnosed essential hypertension before age 60 were invited to participate, including both hypertensive and normotensive siblings. Participants were considered to have hypertension if they had either 1) a previous clinical diagnosis of hypertension by a physician with current anti-hypertensive treatment, or 2) an average systolic blood pressure (SBP) ≥140 mm Hg or diastolic blood pressure (DBP) ≥90 mm Hg on the second and third blood pressure readings. Exclusion criteria were secondary hypertension, pregnancy, and insulin-dependent diabetes mellitus. There were 1583 people with phenotype data, and we removed people who had secondary hypertension (N=9), had gender reassignment (N=1), or lacked SNP data (N=138). Participants who did not self report as non-Hispanic whites (N=5) or were outliers (> mean + 4 standard deviation or < mean - 4 standard deviation) for the outcome (SBP, DBP, and BMI) (N=7) were also excluded. 1423 people with phenotype and genotype data were used for genetic association analysis. We then merged this sample with gene expression data and 789 people were retained in the sample for gene expression analysis. The GENOA study was approved by the institutional review board of the Mayo Clinic, Rochester, MN. Written informed consent was obtained from all participants.

Phenotype Measurement

Blood was drawn by venipuncture after an overnight fast of at least eight hours. Resting SBP and DBP were measured with random zero sphygmomanometers (Hawksley and Sons, West Sussex, England) and cuffs appropriate for arm size. Three readings were taken in the right arm after the participant rested in the sitting position for at least five minutes according to standards set by The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High BP (JNC-7) guidelines.[46] Of the three readings, the last two readings were averaged for the analyses. Height was measured by wall stadiometer, weight by

electronic balance, and BMI was calculated by definition as weight in kilograms divided by height in meters squared.

Genotyping and Quality Control of SNPs

In GENOA, a majority of non-Hispanic whites (N=1386) have been successfully genotyped on the Affymetrix® Genome-Wide Human SNP Array 6.0. There were 123 subjects whose genotypes were not successfully measured on the Affymetrix 6.0 platform, so they were re-genotyped using the Illumina Human 1M-DUO BeadChip.

Genotyping was carried out at the Mayo Clinic in Rochester, Minnesota. Briefly, 500ng genomic DNA at 50ng/ul in low ethylene-diamine-tetraacetic acid (EDTA)-Tris buffer was digested in two separate reaction mixtures using the appropriate restriction enzyme (StyI and NspI, 250ng of DNA for each mixture). This step was followed by ligation of an adaptor sequence containing a universal primer sequence. Samples were then subjected to polymerase chain reaction (PCR) (four PCR reactions per sample for the NspI mixture and three for StyI) with conditions designed to amplify 200-2,000 base pairs. The seven PCR products were then combined with Agencourt Ampure beads, passed over an E & K Scientific filter plate, and eluted with elution buffer. Agarose gel analysis of the PCR products and quantification of the amount of PCR product was performed. PCR product concentration were confirmed to be at least 5ug DNA in 1ul EB buffer. Product was then fragmented with DNase I and an agarose gel analysis of the fragmented DNA was used to confirm this step. Following fragmentation, DNA was labeled with Terminal Deoxynucleotidyl Transferase (TdT), hybridized to the appropriate GeneChip, and incubated overnight. The chip was stained and washed on the Affymetrix 450 Fluidics station and then scanned on the Affymetrix 3000 GeneChip scanner.

Preliminary SNP genotype calls were generated using the Dynamic Model (DM)

algorithm.[58] The final SNP genotype calls were generated by Birdseed, an algorithm designed especially for the Affymetrix® Genome-Wide Human SNP Array 6.0, and based on the robust linear model with Mahalanobis distance classifier algorithm (RLMM).[59] In order to call genotypes while best accounting for experimental variability and population-specific allele frequencies, Birdseed utilizes information about variation across samples to modify pre-computed genotype calling models from Affymetrix for each SNP probe set. Birdseed has been shown to reduce the bias against heterozygous calls and boost call rates to over 99% while simultaneously increasing concordance rates.[60]

To obtain the final dataset, the following quality control (QC) thresholds were applied: sample call rates >95% and SNP call rates >95%. These thresholds are comparable to those used in the published GWAS.[8,9]

Imputation and Quality Control

Because only a portion of SNPs of interest have been genotyped, imputation methods were used to infer missing or untyped SNP genotypes based on known haplotype information from HapMap.[61] HapMap CEU (release 22 build 36) was used as a reference panel containing 30 trios that were collected from Utah residents with northern and western European descent by the Centre d'Etude du Polymorphisme Humain (CEPH).

Imputation was performed by using MACH v 1.0.16 which implements a Markov Chain based algorithm[62] to infer possible pairs of haplotypes for each individual's genotypes up to ~2.5 million SNPs. SNP imputation combined genotype data from the non-Hispanic white sample in Rochester, MN with the HapMap CEU samples and then inferred genotypes probabilistically according to shared haplotype stretches between study samples and HapMap. Imputation results are summarized as an “allele dosage” defined as the expected number of copies of the

alphabetically higher allele at the SNP (a fractional value between 0.0 and 2.0) for each genotype. The imputation accuracy rates of MACH have been shown to be similar to IMPUTE[63] and higher than several other imputation methods (e.g. fast PHASE, Beagle, and PLINK).[64] Quality control thresholds were applied as follows: SNPs with imputation quality score Rsq>0.3, and Hardy-Weiberg equilibrium p-value $>10^{-3}$ (tested in the GENOA unrelated sample).

Principal component analysis was conducted as another quality control step to identify and remove samples with outlying genotype profiles. First, we removed SNPs that had moderate to poor imputation quality as measured by the estimated R^2 between imputed and true genotypes ($R^2<0.8$) from MACH output. Because GENOA includes sibships, we obtained an unrelated sample of individuals (N=570) by randomly selecting one individual from each sibship. The first ten principal components (PCs) were then calculated on the set of SNPs that were common in both Affymetrix and Illumina platforms and were also available in HapMap (N=226,619 SNPs). The imputed genotypes were used to ensure no missing values for SNPs. An additive model was assumed for the SNPs with standardization of mean equals 0 and variance equals 1. We then used the loading matrix for these PCs to calculate the PC values in the full sample. Individuals with outlier values more than 6 standard deviations on any of the 10 PCs (N=45) were removed from the analysis sample.

Gene Expression Assessment and Quality Control

Blood samples for beta-lymphocyte extraction were collected during GENOA Phase I (1995-2000) study and GENOA Phase II (2000-2005) study. Beta-lymphocytes were isolated and transformed using Epstein-Barr virus at the time of collection. For the present study, individuals who participated in an ancillary GENOA study: Genetics of Microangiopathic Brain

Injury Study (GMBI) had their lymphocytes re-initiated for gene expression assessment. Specifically, cell lines of 237 subjects and 552 subjects were collected and transformed during 1995-2000 and 2001-2005, respectively. Immortalized lymphocytes were stored at -180°C in a freezing medium containing 10% dimethyl sulfoxide (DMSO) as a cryoprotectant. To re-initiate the lymphocytes, a vial of cells is retrieved from cryostorage, thawed at 37°C, rinsed twice in Roswell Park Memorial Institute (RPMI) 1640 culture medium (Irvine Scientific, Santa Ana CA), and transferred to a culture flask along with fresh culture medium supplemented with 15% fetal bovine serum (Irvine Scientific, Santa Ana CA). All cultured lymphocytes were maintained in the standardized culture medium, supplemented with 1% penicillin/streptomycin/L-glutamine (Sigma-Aldrich, St. Louis MO), at 37°C and 90% humidity in 5% CO2 for two to three weeks with regular changes of culture medium until a cell density of 5×10^5 cells/mL was achieved.

RNA samples were extracted using standard protocols. RNA quality was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Foster City, CA) and quantified by spectrophotometry using the Nanodrop ND-1000 (Nanodrop Inc., Wilmington, DE). All RNA samples used in the present study yielded both an A260/A280 absorbance ratio greater than 2.0 and a RNA Integrity Number (RIN) ≥ 8. One μg of RNA was labeled using the WT Expression labeling assay (Applied Biosystems/ Ambion, Foster City, CA) including the labeling controls from the GeneChip Eukaryotic Poly-A RNA Control Kit (Affymetrix, Santa Clara, CA). Each step of the labeling protocol was monitored using the Agilent 2100 Bioanalyzer or the Nanodrop spectrophotometer, as recommended by the manufacturer. Hybridization buffer, Eukaryotic Hybridization Controls, and OligoB2 controls were added to the cDNA fragments just prior to hybridization to the Affymetrix Human Exon 1.0 ST Array. Hybridization was performed at

45°C for 17 hours. Following hybridization, the chips were washed and stained with a phycoerythrin-strepavidin conjugate and were scanned at an excitation wavelength of 488 nm.

Array quality control was performed using Affymetrix Expression Console™ (v 1.1) at the transcript level using core-level probe sets. All array images passed visual inspection. Hybridization controls were all present with signal increases following concentration. Labeling control signal strengths followed the order Lys < Phe < Thr < Dap. Signal intensity plots were examined for raw and processed data to identify outliers.

Raw intensity data were processed using the Affymetrix Power Tool software. Probe summarization and probe set normalization were performed using Robust Multi-array Analysis (RMA), which included background correction, quantile normalization, log2-transformation, and probe set summarization.[47] Only core probesets were used to assess exon-level expression. Gene-level expression was assessed by averaging all the core probe sets for that gene.

Quality control of exon array data were done by examining the QC matrix to identify possible outliers or non-performing samples using the Partek Genomic Suite software (V.6.6). Analyses were restricted to core probe sets. Probe sets which are known to cross-hybridize and those with undetectable expression were also excluded.

Criteria to select SNPs, Genes, and Gene Expression Levels of interest

Both CHARGE and Global BPgen studies were included in ICBP. Therefore, top SNPs that were identified to be associated with blood pressure from ICBP were used as the index SNPs to select for proximal SNPs for study (Appendix 3.1).

Because only some of the index SNPs identified from previous GWAS fall within gene boundaries, we used four strategies to choose the set of SNPs and gene expression data to investigate the specific aims based on the location of index SNPs. First, if the index SNP is

inside of a gene, then all of the SNPs that are in a gene and ± 20 kb on either side of that specific gene were selected. Gene expression data for that specific gene was also selected. Second, if the index SNP is outside of a gene and the distance between the index SNP and the closest gene is within 100 kb, then SNPs that are within ± 20 kb of the index SNPs were selected. As for the gene expression data, the closest gene upstream and the closest gene downstream of the index SNP were selected. Third, in gene sparse regions, SNPs that are within ± 20 kb of the index SNPs were selected. As for the gene expression data, there is typically only one gene either upstream or downstream within ± 100kb of the index SNP. In this case, the single gene was chosen for gene expression data. Finally, in gene deserts where there is no gene within ± 100kb of the index SNP, we excluded the index SNP from our analysis. Detailed information of selected SNPs and genes for blood pressure are shown in Appendix 3.2.

Statistical Methods

All data management and statistical analyses were conducted using R version 2.14. Allele frequencies, genotype frequencies, and linkage disequilibrium (LD), as measured by r^2, were estimated for all pairs of SNPs using an expectation maximization (EM) algorithm. Before performing the statistical analysis, outcome variable distributions were examined for normality and outliers by using histograms and QQ-plots. Outliers greater than four standard deviations from the mean were removed.

Quantitative variables of interest include SNP allele dosage (a fractional value between 0.0 and 2.0), age (years), BMI (kg/m^2), SBP (mm Hg), DBP (mm Hg), and gene expression data (gene level). Categorical variables include gender (0=male 1=female). Before analysis, 15 mm Hg was added to observed SBP values and 10 mm Hg was added to DBP values for participants who were taking antihypertensive medication to adjust for the drug effect of blood pressure

lowering medications.[65] This adjustment was consistent with the antihypertensive drug adjustment in ICBP.

Association Testing

The goal of the present study is to investigate the functional genetic factors of blood pressure variation using a three stage approach (a) evaluating the associations between blood pressure traits and genetic variations (model a in Figure 3.1), (b) investigating the association between these genetic variations and proximal gene expression (model b) and (c) examining the relationship between proximal gene expression and blood pressure traits (model c).

For model a (Figure 3.1), the outcome variables (SBP and DBP) were first adjusted for age, age^2, gender and BMI and the residual variables were treated as the outcome variables in the linear mixed model to test for the association with the predictor (SNP). A Family ID was treated as a random intercept to control for family correlation in this dataset. This two step approach was also performed in models b and c. In model b, gene expression data was first adjusted by age, gender, and batch effect and the residuals were treated as the outcome predicted by the proximal SNPs around that specific gene. For model c, we investigated the association between gene expression (adjusted for age, gender, and batch effect) and blood pressure (adjustment for age, age^2, gender and BMI) by using the same statistical procedure.

The statistical models for the three tests are listed as follows and statistical significance of the main effect of each respective SNP was determined based on model p-value (two-sided).

Model (a):

$$\text{Blood Pressure residual}_{ij} = \beta_{oi} + \beta_1(\text{SNP}_{ij}) + \varepsilon_{ij}$$

$$\mathbf{H_o}: \beta_1 = 0, \qquad \mathbf{H_a}: \beta_1 \neq 0 \qquad \sim\chi^2_{(df=1)}$$

Model (b):

$$\text{Gene Expression residual}_{ij} = \beta_{oi} + \beta_1(\text{SNP}_{ij}) + \varepsilon_{ij}$$

$$\mathbf{H_o}: \beta_1 = 0, \qquad \mathbf{H_a}: \beta_1 \neq 0 \qquad \sim\chi^2_{(df=1)}$$

Model (c):

$$\text{Blood Pressure residual}_{ij} = \beta_{oi} + \beta_1(\text{Gene Expression residual}_{ij}) + \varepsilon_{ij}$$

$$\mathbf{H_o}: \beta_1 = 0, \qquad \mathbf{H_a}: \beta_1 \neq 0 \qquad \sim\chi^2_{(df=1)}$$

In the linear mixed effect model, i represents family and j represents individual. β_{oi} is a random intercept for the i^{th} family, and it is normally distributed with mean β_o and variance σ_b^2. ε_{ij} is the residual variation within the j^{th} individual from the i^{th} family, and it is normally distributed with mean 0 and variance σ^2. ε_{ij} is assumed to be independent from $\beta_{oi.}$

Results

Descriptive statistics for covariates and outcome variables in GENOA non-Hispanic whites are listed in Table 3.1. In the sample of 1423 individuals with both phenotype and genotype information, the average age was 55.3 years old, 55.31% were female, nearly half of them (45.26%) were obese, and around three-fourths of them (72.59%) had hypertension. The average SBP and DBP was 132 mm Hg and 78 mm Hg, respectively. For those subjects with both SNP and gene expression data (N=789), the demographic statistics were approximately the same except there were fewer people with obesity (41.44% vs. 45.26%). Compared to the sample who were removed (N=634) due to the lack of gene expression data, subjects with both SNP and gene expression data (N=789) had lower BMI (29.73 vs. 31.10, *p*=0.0001), had more females (58.05% vs. 51.89% , *p*=0.023) and were less obese (41.44 % vs. 50.00% , *p*=0.023).

Before we performed our full investigation, we evaluated the association of the 24 index SNPs reported from ICBP for their association with blood pressure and their influence on their proximal gene's expression (*cis*-effects) in GENOA. Out of 24 index SNPs, 3 SNPs were identified to be associated with both SBP and DBP in 3 genetic regions: *MTHFR* (rs17367504, SBP: p= 0.003; DBP: p= 0.012), *APT2B1* (rs17249754, SBP: p= 0.024; DBP: p= 0.045), and *PLCE1* (rs932764, SBP: p= 0.027; DBP: p= 0.02). Furthermore,10 index SNPs were found to have *cis*-effects around the following 10 genetic regions: *ULK4* (rs3774372, p=1.44 x10^{-55}), *MTHFR* (rs17367504, p=2.84 x10^{-11}), *FES* (rs2521501, p=3.55 x10^{-9}), *SLC4A7* (rs13082711, p=1.30 x10^{-7}), *CAPZA1* (rs2932538, p=5.81 x10^{-6}), *PLEKHA7* (rs381815, p=6.21 x10^{-6}), *CSK* (rs1378942, p=1.72 x10^{-5}), *HFE* (rs1799945, p=5.00 x10^{-4}), *ADM* (rs7129220, p=0.004), *CNNM2* (rs11191548, p=0.04). Specifically, rs17367504 in *MTHFR* was associated with both blood pressure and *MTHFR* expression (Table 3.2). Results for association of 24 index SNPs and blood pressure between ICBP and GENOA were compared and listed in Appendix 3.3. Overall, the allele frequencies of the coded allele between GENOA and ICBP were very close. The beta values of the three significant SNPs (*PLCE1*: rs932764, *MTHFR*: rs17367504, and *ATP2B1*: rs17249754) associated with blood pressure in GENOA were larger than those in ICBP (SBP: beta= -1.49 vs. -0.48; 2.94 vs. 0.90; -2.17 vs. -0.90, DBP: beta= -0.95 vs. -0.19; 1.50 vs. 0.55; -1.17 vs. -0.52 respectively). Because GENOA is a much smaller sample compared to ICBP, we expected that only larger beta estimates would be identified to be significant. In summary, 70% (=17/24 β coefficients) of beta coefficients in GENOA had consistent direction with those in ICBP.

Model a: Association between SNPs and Blood Pressure

Using 3537 SNPs proximal to 27 genes, the association of SNPs and blood pressure were

evaluated in 1423 GENOA non-Hispanic whites. Overall, 202 SNPs and 363 SNPs were identified to be associated with SBP and DBP, respectively (p<0.05) (Table 3.3). *ULK4* had the greatest number of significant SNPs associated with SBP (N=57) and DBP (N=38). Among all of the SNPs with significant signals (p<0.05), 64 SNPs were found to be associated with both SBP and DBP around the gene regions as follows: *MTHFR, PLCD3, PLEKHA7, ARHGAP42, PLCE1, MECOM, FGF5,* and *ATP2B1*. After applying Bonferroni corrected alpha-level (0.05/number of proximal SNPs for each specific gene), only 1 SNP (rs11105358, listed in Table 3.4) located around *ATP2B1* was found to be associated with both SBP and DBP (Table 3.3). For each gene, the most significant SNPs associated with SBP and DBP are listed in Table 3.4.

Model b: Association between Proximal SNPs and Gene Expression

To determine whether SNP variation in previously found blood pressure related gene regions may have *cis*-effects on gene expression level in GENOA non-Hispanic whites, 3537 SNPs were investigated for their association with their own gene's expression. *ULK4* was found to have the most significant signals. Out of 512 association tests, 278 SNPs were associated with *ULK4* expression (p<0.05), and 158 SNPs passed Bonferroni corrected alpha-level at 9.77×10^{-5} (Table 3.5). The second most significant *cis*-effects were found in the examination of *MTHFR* expression. Among 66 association tests, 46 SNPs had significant association with *MTHFR* expression, and 38 SNPs passed Bonferroni corrected alpha-level (p<7.58×10^{-4}). In summary, out of 3537 association tests, 829 SNPs proximal to 24 genes showed significant *cis*-effects on gene expression (p<0.05). Among these significant results, 320 SNPs located around 14 genes passed Bonferroni corrected alpha-level. The SNPs with strongest *cis*-effects for these 14 genes were listed as follows: *ULK4* (rs3934103, p=8.72×10^{-58}), *MTHFR* (rs2066462, p=7.74×10^{-18}), *FES* (rs8032315, p=6.37×10^{-12}), *FGF5* (rs7688596, p=4.05×10^{-9}), *GOSR2* (rs17676978, p=4.71

x10^{-8}), *SLC4A7* (rs13082711, *p*=1.29 x10^{-7}), *PLEKHA7* (rs10832696, *p*=3.89 x10^{-7}), *CAPZA1* (rs2932534, *p*=9.58 x10^{-7}), *CSK* (rs4886606, *p*=6.88 x10^{-6}), *ARHGAP42* (rs12283290, *p*=1.65 x10^{-5}), *GUCY1A3* (rs2625276, *p*=2.37 x10^{-4}), *MOV10* (rs3748655, *p*=4.24 x10^{-4}), *HFE* (rs198851, *p*=4.30 x10^{-4}) and *BAT3* (rs2736176, *p*=7.27 x10^{-4}) (Table 3.5).

LocusZoom was then used to plot the regional association of proximal SNPs and gene expression for these 14 genes (Appendix 3.4 - Appendix 3.17). The LocusZoom plot for *ULK4* shows that the majority of significant SNPs were in high LD with the most significant SNP (r^2>0.8) (Appendix 3.4). For *MTHFR*, there were two clusters of significant SNPs and they were in different LD blocks with each other(r^2<0.4) which suggested that different alleles in this locus could cause *MTHFR* expression and this phenomenon is called allelic heterogeneity (Appendix 3.5). LocusZoom plots for the remaining genes are shown in Appendix 3.6- Appendix 3.17. Allelic heterogeneity is shown in expression of several genetic regions such as *FES* expression, *FGF5* expression, *PLEKHA7* expression, *CSK* expression, *ARHGAP42* expression, and *HFE* expression.

Model c: Association between Gene Expression and Blood Pressure

Gene expression levels for 27 genes were then investigated for their association with blood pressure. Expression levels of four genes were found to be associated with SBP: *GOSR2* (*p*=0.0062), *PLCD3* (*p*=0.0396), *SLC39A8* (*p*=0.0120), and *ULK4* (*p*=0.0149). Eight genes had expression level significantly associated with DBP: *GOSR2* (*p*=0.0204), *ZNF652* (*p*=0.0155), *SLC39A8* (*p*=0.0256), *MECOM* (*p*=0.0409), *ULK4* (*p*=0.0005), *CAPZA1* (*p*=0.0206), *CNNM2* (*p*=0.0300) and *NPR3* (*p*=0.0219) (Table 3.6). After applying Bonferroni correction, only one association passed the alpha level of 0.00185 (=0.05/27 genes): *ULK4* associated with DBP (*p*=0.0005).

Last, the results from all three models (i.e. model a, b and c) were summarized in table 3.6 based on alpha level of 0.05. Among 27 genes, only *ULK4* showed significant associations for all three models. Specifically, 35 SNPs in *ULK4* were found to have effects on both DBP and *ULK4* expression while *ULK4* expression was also found to be associated with DBP. Greater *ULK4* expression level was significantly associated with increased DBP level (β=4.03, p=0.0004) and this association passed the Bonferroni corrected alpha level of 0.00185. All of these 35 DBP associated SNPs were also highly significantly associated with *ULK4* expression ($p<10^{-46}$) and they all passed the Bonferroni corrected alpha level of 9.77×10^{-5}(=0.05/512 proximal SNPs, Appendix 3.18). These associations were consistent with the observed positive association between *ULK4* expression and DBP. For example, the T allele of rs17218264 was associated with significantly higher *ULK4* expression and increased DBP. As for those 28 SBP-associated SNPs that were also associated with *ULK4* expression, only 1 SNP (rs9311289, p=0.000011) passed the Bonferroni corrected alpha level ($p=1.1 \times 10^{-5}$, Appendix 3.19). Similarly, the effects of these 28 significantly SNP on *ULK4* expression and SBP were consistent with the positive association between *ULK4* expression and SBP. LocusZoom was then used to plot these 35 DBP associated SNPs and 28 SBP associated SNPs and their association with *ULK4* expression (Figure 3.2). Figure 3.2 shows that all of these 35 DBP-associated SNPs were in high LD with the most significant SNP ($r^2>0.8$) and they were in different LD block from those 28 SBP-associated SNPs. In addition, 27 SBP-associated SNPs were in high LD with the second most significant SNP and they were not in LD with the most significant SNP, rs9311289, that was highly associated with *ULK4* expression ($p= 1.1 \times 10^{-5}$).

There were 7 and 23 SNPs found to be associated with both blood pressure and expression level of the *MTHFR* gene and the *PLEKHA7* gene, respectively (Appendix 3.20 and Appendix

3.21). However, no significant associations were found between their gene expression level and blood pressure. All of the 7 SNPs associated with *MTHFR* expression were highly significant and passed the Bonferroni corrected alpha level of 7.58×10^{-4} (Appendix 3.20). The LocusZoom plot shows that among these 7 SNPs, 6 SNPs were in high LD with the most significant SNP and the index SNP , rs17367504, reported from ICBP was one of them (Figure 3.3). For *PLEKHA7* gene, only 3 SNPs passed the Bonferroni corrected alpha level of 1.67×10^{-4} and two of them were in high LD ($r^2 > 0.8$) (Figure 3.4).

Sensitivity Analysis

Because the sample sizes for genetic association (N=1423) and gene expression analysis (n=789) were different, we did a sensitivity analysis by using 789 subjects to test for the SNP-blood pressure associations to check if the results we found earlier were also present in this smaller sample. A total of 3537 SNPs were examined for their association with blood pressure and gene expression levels in this smaller sample (N=789). For *ULK4* gene, 22 SNPs were found to be associated with both SBP and *ULK4* expression and 21 SNPs were found to be associated with both DBP and *ULK4* expression. In the previous analysis with the larger sample (N=1423), 28 SNPs in *ULK4* were associated with both SBP and *ULK4* expression, and 22 of them were identified to be associated with both SBP and *ULK4* expression in the smaller sample (N=789) (Appendix 3.22). Similarly, we found 35 SNPs associated with both DBP and *ULK4* expression in the larger sample, and 21 of them were identified to be associated with both SBP and *ULK4* expression in the smaller sample (Appendix 3.23). In addition, we did not identify any new findings in this smaller sample.

Assessment of the Mediating Effects of Gene Expression Levels between SNPs and Blood Pressure

We found 22 SNPs associated with both SBP and *ULK4* expression and 21 SNPs associated with both DBP and *ULK4* expression (Appendix 3.22 and Appendix 3.23). In addition, *ULK4* expression was found to be associated with both SBP and DBP (Table 3.6). We then investigated whether *ULK4* expression mediated the associations between these SNPs and blood pressure. The associations between these identified SNPs and blood pressure were investigated before and after adjusting for *ULK4* expression. After adjusting for *ULK4* expression, the associations of these SNPs with SBP and DBP were not statistically significant at the alpha level of 0.05 (Appendix 3.24 and Appendix 3.25, respectively). However, the associations of *ULK4* expression and blood pressure remained significant after adjusting for these SNPs ($p<0.05$). These results suggested that *ULK4* expression might mediate the association between these SNPs and blood pressure.

Discussion

In this study of hypertensive, non-Hispanic white sibships, only one gene, *ULK4,* was found to have proximal SNPs associated with both blood pressure and *ULK4* expression and it's expression was also identified to be associated with blood pressure ($p<0.05$). After Bonferroni correction was applied, proximal SNPs still had significant *cis*-effects on gene expression levels but their association with blood pressure did not passed Bonferroni correction alpha level. This phenomenon could be explained by three reasons. First, we estimated the effective sample sizes[66] based on the sibship structure (Appendix 3.26 and Appendix 3.27) in GENOA and performed power calculations (Appendix 3.28 and Appendix 3.29) by using effective sample sizes and standardized phenotypes (mean= 0, standard deviation=1). With effective sample size of 989 individuals at alpha level of 0.05, the power to detect the associations between SNPs and blood pressure only ranged from 0.43 (MAF=0.2, β=0.1) to 0.61 (MAF=0.5, β=0.1) (Appendix 3.28).

On the other hand, the power to detect the *cis*-effects of SNPs on gene expression levels was supposed to be smaller due to the decreased effective sample size of 590 (Appendix 3.29). However, since the standardized effect sizes for the associations of SNPs and gene expression levels were much larger (0.26-0.83) than the the associations of SNPs and blood pressure (SBP:0.05-0.17, DBP:0.07-0.26), the estimated power ranged from 0.93 (MAF=0.2, β=0.25) to 0.99 (MAF=0.5, β=0.25) at alpha level of 0.05 (Appendix 3.29). We also compared the results of SNP-blood pressure association between ICBP and GENOA (Appendix 3.3). Out of 24 associations tests, only three significant SNP-blood pressure associations (p<0.05) were identified in GENOA non-Hispanic whites. These three beta coefficients have the same direction compared to those in ICBP but the effect sizes were much larger and they were probably overestimated in GENOA. Due to the much smaller sample size and limited power to detect the modest effect size of SNPs associated with blood pressure in GENOA, it could partially explain why we only identified a few significant SNP-blood pressure associations (p<0.05) at the beginning and none after Bonferroni correction. Second, around 70% of individuals in this current study were hypertensive while subjects in ICBP were from the general population. The lack of replication of most of the SNP-blood pressure associations could be due to the differences among study populations in distributions of genetic and environmental factors. For example, BMI is an important risk factor for blood pressure and the distribution of mean BMI ranged from 22.8 to 28.6 in the numerous studies in ICBP while the average BMI in GENOA is larger (BMI=31.1). Last, it is possible that SNPs that influence blood pressure in hypertensive sibships are not completely the same as those from the general population. Nevertheless, these blood pressure-associated SNPs reported from ICBP were highly significant and replicate across many studies (total sample size>200,000) and it is important to investigate whether these SNPs

have similar effects on blood pressure in the hypertensive sibships.

Previous literature has demonstrated that some proximal SNPs are associated with both *ULK4* expression and blood pressure. A genome-wide association study of global gene expression identified three missense SNPs (rs1716975, rs2272007 and rs1052501, $p<10^{-7}$) associated with altered expression of *ULK4* among HapMap CEU in lymphoblastoid cell lines.[39] The CHARGE consortium then examined these three SNPs in their sample and found all of them were significantly associated with DBP ($p<5x10^{-5}$) but not with SBP (p=0.84-0.94).[16] Although our sample only identified these three SNPs to be associated with *ULK4* expression ($p<10^{-55}$), but not with blood pressure (data not shown). All of these three SNPs are in high LD (r^2 >0.8) with those 35 SNPs associated with both DBP and *ULK4* expression in GENOA non-Hispanic whites. In addition, HapMap CEU also identified these 35 SNPs associated with *ULK4* expression in lymphoblastoid cell lines ($p<10^{-9}$).[67] However, for those 28 SNPs associated with both SBP and gene expression in GENOA, HapMap CEU sample did not identify these SNPs and so far no studies in the literature have reported evidence of these signals. The most significant SNP, rs9311289, is the only SNP that passed the Bonferroni corrected alpha level and it is in different LD block from the remaining 27 SNPs. A replication sample will be needed to confirm and validate this finding.

Among 35 SNPs associated with both DBP and *ULK4* expression, rs17215589 is the only non-synonymous SNP. The G-to-A transition at rs17215589 results in an alanine to threonine alternation at amino acid 715. The *in silico* SIFT and Polyphen algorithms predicted it to be tolerated or benign, respectively. *ULK4* (chromosome 3p22.1) encodes a serine/threonine-protein kinase and this is an enzyme that transfers phosphates to serine or threonine in proteins. Little is known about *ULK4* (serine/threonine-protein kinase *ULK4*) and how genetic variation and gene

expression in these genes might influence blood pressure.

The second gene that showed a significant result is *MTHFR*. The index SNP, rs17367504, in *MTHFR* was found to be associated with both blood pressure and *MTHFR* expression in GENOA. ICBP also identified the *cis*-effects of rs17367504 on *MTHFR* expression in blood (p=1x10^{-11}). *MTHFR* expression was not found to be associated with blood pressure in GENOA, and it is likely that there are many factors that influence serum and plasma protein levels which are likely to be more directly implicated in the disease process than gene expression level.[68] Other mechanisms could include epigenetic factors, environmental factors influencing regulation of expression, and post-translation modification. Studying these factors will help us to better understand the etiology of disease process linking from gene expression to protein and then to diseases. rs17367504 is located in an intron of *MTHFR*, methylenetetrahydrofolate reductase (chromosome 1p36.3). This SNP is 6.4 kb away from rs1801133 (C677T, LD r^2=0.06), a coding variant that has been shown to be associated with high plasma homocysteine concentration, pre-eclampsia, and hypertension.[69] *MTHFR* is an enzyme that catalyzes the conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate. This reaction is required for the remethylation process that converts homocysteine to methionine. Individuals with the TT genotype at codon 677 tend to have lower plasma folate levels, have less active *MTHFR* available to produce 5-methyltetrahydrofolate, and thus are predisposed to have mild high blood homocysteine levels. A high level of blood homocysteine may injure vascular endothelium and thus predispose to atherosclerosis and result in hypertension.[70]

Last, 23 SNPs located in *PLEKHA7* (11p15.1) showed significant association with both blood pressure and *PLEKHA7* expression and three of these SNPs passed the Bonferroni corrected alpha level. *PLEKHA7* encodes pleckstrin homology domain- containing protein,

family A member 7. This protein was reported to be involved in regulating integrity of zonula adherens, an epithelial cadherin-based cell-cell junction.[71] However, little is known about how *PLEKHA7* could influence blood pressure. In addition to non-Hispanic whites, rs381815 in *PLEKHA7* was shown to be associated with blood pressure in both Chinese[72] and Korean[73] population, and our results are consistent with theirs. Specifically, the T allele of rs381815 is associated with increased blood pressure across these three populations which suggest that the genetic architecture involved with *PLEKHA7* may have similar effects on blood pressure across these three populations. The potential functions for the 27 BP-related genes are listed in Appendix 3.30.1- Appendix 3.30.3.

There are some limitations of the current study that need to be considered. 1) Immortalized celllines were used and it is possible that transformation of beta-lymphocytes could alter the biology of the cell and have an effect on expression profile. However, previous studies have shown that a large proportion of genetically determined differences in gene expression can be measured in immortalized lymphocytes in culture.[74] Also, standardization of the environment in cell culture removes individual-level environmental variation. As a result, the measured gene expression levels could optimally reflect individual DNA variations. 2) Because the sample in this current study is non-Hispanic whites, the inferences may not be generalizable to other ethnic groups. 3) We adjusted antihypertensive drug treatment by adding 15 mm Hg and 10 mm Hg to SBP and DBP, respectively. This approach has been adopted by many GWAS consortia rather than excluding individuals on therapy or ignoring antihypertensive treatment. However, other factors such as medication categories, drug dosage, and variation in prescription patterns could influence blood pressure regulation and this adjustment scheme might be an oversimplification. Nevertheless, this drug adjustment should generally lead the bias toward the null. An alternative

method for treatment adjustment was also suggested by Cui et al. (addition of 10/5 mm Hg to SBP and DBP, respectively)[75] which gave very similar results in an interim analysis in this study. 4) Other environment factors such as dietary sodium and potassium intake, physical activity and stress also contribute to interindividual differences in blood pressure. Because most of the variables were not measured or measured adequately, we could not adjust for them in this study. (5) Because of the relatively small sample size and limited power in our GENOA non-Hispanic whites, for those index SNPs that were not within a gene, we only focused on proximal SNPs that are within ±20kb of the index SNPs to maintain the appropriate number of SNPs to obtain adequate power. We chose 20kb because the average length of LD blocks in the European population is about 22 kb.[76] The disadvantage of this approach is we were not able to capture and study the whole spectrum of SNPs within those specific genes close to the index SNPs. This could explain why we found almost no SNPs associated with both blood pressure and gene expression in those genetic regions (Table 3.6, gene dense regions and gene sparse region).

In conclusion, we found very strong evidence connecting SNP level variation with *ULK4* expression variation ($p<10^{-45}$) and identified a significant association between *ULK4* expression variation and DBP ($p=0.0004$) in GENOA non-Hispanic whites. This result underscores the power of combing gene expression data with SNP variation data to identify genes with potential functional significance for blood pressure. *MTHFR,* and *PLEKHA7* also showed suggestive evidence of associations. Although the full implications and biological significance of *ULK4*, *MTHFR,* and *PLEKHA7* are not yet completely understood, our findings in these genes warrants future biological studies to improve our understanding of the molecular mechanisms connecting DNA variants and gene expression to blood pressure.

Figure 3.1. Association analysis among SNPs, gene expression and blood pressure.

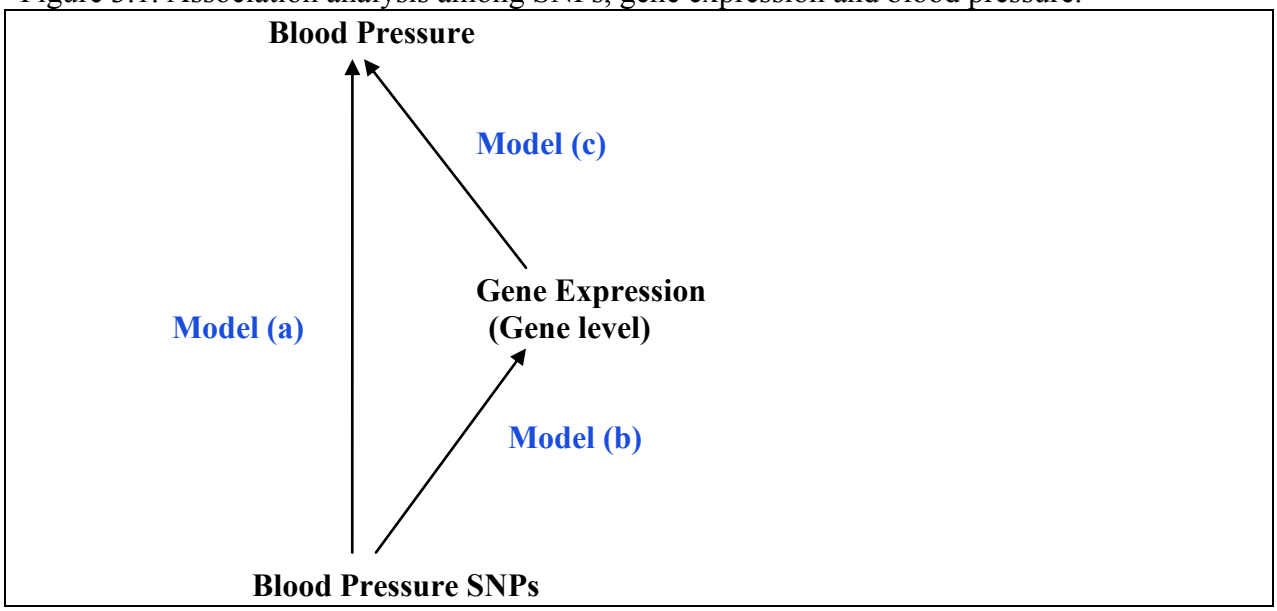


Table 3.1. Descriptive characteristics of non-Hispanic whites in GENOA.

	Genetic Association (N=1423)		With Gene Expression Data (N=789)		Without Gene Expression Data (N=634)		
Continuous Variables	Mean	SD	Mean	SD	Mean	SD	T-test P-value†
Age, years	55.30	10.76	55.65	10.36	54.87	11.23	0.173
Height, cm	168.75	9.24	168.41	9.09	169.17	9.41	0.125
Weight, kg	86.59	19.46	84.53	18.64	89.14	20.15	**0.0001**
BMI, kg/m^2	30.34	6.16	29.73	5.82	31.10	6.48	**0.0001**
SBP, mm Hg	132.89	16.91	132.8	16.66	133.00	17.22	0.827
Adjusted SBP, mm Hg*	142.53	19.73	142.73	19.27	142.30	20.30	0.683
DBP, mm Hg	78.48	9.62	78.52	9.43	78.44	9.86	0.877
Adjusted DBP, mm Hg*	84.91	10.92	85.14	10.69	84.64	11.19	0.391
Categorical Variables	N	%	N	%	N	%	χ^2 –test P-value†
Female	787	55.31	458	58.05	329	51.89	**0.023**
Obesity	644	45.26	327	41.44	317	50.00	**0.002**
Hypertension	1033	72.59	585	74.14	448	70.66	0.160
Antihypertensive Medication	915	64.30	522	66.16	393	61.99	0.115

† Statistical tests compared the subjects with both SNP and gene expression data (n=789) to subjects whithout gene expression data (n=634).

*Adjusted SBP=SBP+15mm Hg for taking hypertension drug.
Adjusted DBP=DBP+10mm Hg for taking hypertension drug.

Table 3.2. Association of index SNP, blood pressure and gene expression in GENOA.

Index SNP	Chr	SNP Position	Nearest Gene	SBP		DBP		Gene Expression	
				Beta	P-value	Beta	P-value	Beta	P-value
rs3774372	3	41,877,414	*ULK4*	0.24	0.784	0.41	0.443	**0.317**	**1.44E-55**
rs17367504	1	11,862,778	*MTHFR*	**2.94**	**0.003**	**1.50**	**0.012**	**-0.095**	**2.84E-11**
rs2521501	15	91,437,388	*FES*	-1.05	0.223	-0.28	0.590	**0.072**	**3.55E-09**
rs13082711	3	27,537,909	*SLC4A7*	0.10	0.902	0.14	0.771	**-0.083**	**1.30E-07**
rs2932538	1	113,216,543	*CAPZA1*	-1.08	0.189	-0.05	0.928	**0.043**	**5.81E-06**
rs381815	11	16,902,268	*PLEKHA7*	-1.05	0.186	-0.77	0.112	**0.093**	**6.21E-06**
rs1378942	15	75,077,367	*CSK*	-0.66	0.363	0.14	0.756	**-0.035**	**1.72E-05**
rs1799945	6	26,091,179	*HFE*	0.87	0.379	0.70	0.248	**0.046**	**5.00E-04**
rs7129220	11	10,350,538	*ADM*	-1.27	0.238	-1.15	0.078	**0.083**	**0.004**
rs11191548	10	104,846,178	*CNNM2*	0.03	0.978	-0.19	0.788	**-0.030**	**0.040**
rs805303	6	31,616,366	*BAT3*	-0.11	0.869	-0.20	0.631	-0.010	0.102
rs13107325	4	103,188,709	*SLC39A8*	0.14	0.918	-0.42	0.615	0.030	0.230
rs3184504	12	111,884,608	*SH2B3*	0.66	0.364	0.03	0.943	-0.011	0.241
rs17608766	17	45,013,271	*GOSR2*	-1.23	0.216	-0.64	0.289	-0.014	0.249
rs419076	3	169,100,886	*MECOM*	-0.61	0.371	-0.25	0.555	-0.009	0.270
rs11191548	10	104,846,178	*NT5C2*	0.03	0.978	-0.19	0.788	0.021	0.280
rs1813353	10	18,707,448	*CACNB2*	-0.37	0.619	-0.53	0.243	0.049	0.286
rs1173771	5	32,815,028	*NPR3*	-0.32	0.652	-0.20	0.637	-0.015	0.324
rs633185	11	100,593,538	*ARHGAP42*	-0.10	0.891	0.27	0.557	-0.027	0.332
rs13139571	4	156,645,513	*GUCY1A3*	0.45	0.566	0.02	0.969	-0.050	0.366
rs1458038	4	81,164,723	*PRDM8*	-0.97	0.225	-0.79	0.104	0.011	0.394
rs1458038	4	81,164,723	*FGF5*	-0.97	0.225	-0.79	0.104	0.016	0.474
rs2932538	1	113,216,543	*MOV10*	-1.08	0.189	-0.05	0.928	0.008	0.564
rs12940887	17	47,402,807	*ZNF652*	-0.25	0.716	-0.70	0.097	-0.004	0.774
rs17249754	12	90,060,586	*ATP2B1*	**-2.17**	**0.024**	**-1.17**	**0.045**	-0.005	0.801
rs12946454	17	43,208,121	*PLCD3*	0.07	0.923	0.26	0.581	-0.002	0.869
rs932764	10	95,895,940	*PLCE1*	**-1.49**	**0.027**	**-0.95**	**0.020**	0.001	0.876

Bold:$p<0.05$.

Table 3.3. Significant results from association of SNPs and blood pressure.†

Index SNP	Chr	Nearest Gene	SNP Selection Range		No. of Proximal SNPs	Bonferroni Corrected Alpha-Level	No. of SNPs associated with (p<0.05)			No. of SNPs associated with (p<Bonferroni Correction)		
			Star	End			SBP	DBP	Both SBP and DBP	SBP	DBP	Both SBP and DBP
Gene Region with Index SNP inside of it												
rs1799945	6	*HFE*	26,067,509	26,115,469	53	9.43E-04	0	0	0	0	0	0
rs2521501	15	*FES*	91,407,665	91,459,006	25	2.00E-03	0	0	0	0	0	0
rs805303	6	*BAT3*	31,586,805	31,640,170	65	7.69E-04	2	0	0	0	0	0
rs17608766	17	*GOSR2*	44,980,486	45,038,732	48	1.04E-03	3	0	0	0	0	0
rs17367504	1	*MTHFR*	11,825,787	11,886,115	66	7.58E-04	7	22	7	0	0	0
rs12946454	17	*PLCD3*	43,169,009	43,229,891	51	9.80E-04	2	10	2	0	0	0
rs1378942	15	*CSK*	75,054,425	75,115,539	28	1.79E-03	0	0	0	0	0	0
rs3184504	12	*SH2B3*	111,823,752	111,909,426	20	2.50E-03	0	0	0	0	0	0
rs13139571	4	*GUCY1A3*	156,567,862	156,678,212	103	4.85E-04	2	0	0	0	0	0
rs12940887	17	*ZNF652*	47,346,569	47,459,835	66	7.58E-04	0	15	0	0	0	0
rs13107325	4	*SLC39A8*	103,162,821	103,286,655	133	3.76E-04	9	0	0	0	0	0
rs381815	11	*PLEKHA7*	16,789,212	17,055,963	299	1.67E-04	10	21	1	0	0	0
rs633185	11	*ARHGAP42*	100,538,407	100,881,656	497	1.01E-04	9	49	5	0	0	0
rs932764	10	*PLCE1*	95,733,746	96,108,146	255	1.96E-04	19	49	6	0	0	0
rs1813353	10	*CACNB2*	18,409,606	18,850,688	542	9.23E-05	10	1	0	0	0	0
rs419076	3	*MECOM*	168,781,287	169,401,563	563	8.88E-05	49	38	29	0	0	0
rs3774372	3	*ULK4*	41,268,091	42,023,660	512	9.77E-05	57	38	0	0	0	0
Gene Dense Region (Multiple Genes within ±100 kb of Index SNP)												
rs1458038	4	*FGF5*	81,144,723	81,207,742	40	1.25E-03	6	4	3	0	0	0
	4	*PRDM8*	81,144,723	81,138,657								
rs2932538	1	*MOV10*	113,196,543	113,237,048	24	2.08E-03	0	0	0	0	0	0
	1	*CAPZA1*	113,196,543	113,182,075								
rs11191548	10	*NT5C2*	104,826,178	104,867,774	49	1.02E-03	0	0	0	0	0	0
	10	*CNNM2*	104,826,178	104,698,075								
rs7129220	11	*ADM*	10,330,538	10,346,642	35	1.43E-03	2	0	0	0	0	0
Gene Sparse Region (Only One Gene within ±100 kb of Index SNP)												
rs17249754	12	*ATP2B1*	90,040,586	90,001,828	14	3.57E-03	9	11	8	1	2	1
rs1173771	5	*NPR3*	32,795,028	32,731,665	25	2.00E-03	0	0	0	0	0	0
rs13082711	3	*SLC4A7*	27,517,909	27,434,215	24	2.08E-03	0	1	0	0	0	0
Total	---	----------	----------	----------	3537	----------	202	363	64	1	2	1

†Genes were sorted by gene size. See Appendix 3.2.

Table 3.4. Summary of most significant SNPs for SBP and DBP.

Gene	Chr	No. of Proximal SNPs	Bonferroni Corrected Alpha-Level	Most significant SNP associated with SBP			Most significant SNP associated with DBP		
				SNP	SNP Beta	SNP P-value	SNP	SNP Beta	SNP P-value
Gene Region with Index SNP inside of it									
HFE	6	53	9.43E-04	rs1800562	-2.49	0.0592	rs198851	0.77	0.2346
FES	15	25	2.00E-03	rs1029420	1.82	0.0599	rs6496737	0.69	0.1194
BAT3	6	65	7.69E-04	rs2178899	-2.38	**0.0249**	rs2242657	-0.98	0.1037
GOSR2	17	48	1.04E-03	rs6504622	1.36	**0.0488**	rs9303532	-0.71	0.0966
MTHFR	1	66	7.58E-04	rs12567136	3.05	**0.0019**	rs4846041	1.35	**0.0038**
PLCD3	17	51	9.80E-04	rs16939858	-1.67	**0.0446**	rs16939858	-1.46	**0.0037**
CSK	15	28	1.79E-03	rs1378938	-1.48	0.0518	rs12438333	1.78	0.0520
SH2B3	12	20	2.50E-03	rs11065898	-1.10	0.1725	rs12580300	0.79	0.1192
GUCY1A3	4	103	4.85E-04	rs3796591	-2.18	**0.0321**	rs3796591	-0.86	0.1665
ZNF652	17	66	7.58E-04	rs8082083	3.20	0.0786	rs12937634	0.97	**0.0195**
SLC39A8	4	133	3.76E-04	rs6829701	-2.15	**0.0059**	rs170871	-0.80	0.0530
PLEKHA7	11	299	1.67E-04	rs2215085	-1.54	**0.0277**	rs7109209	-1.17	**0.0067**
ARHGAP42	11	497	1.01E-04	rs662453	-1.56	**0.0256**	rs12294970	2.64	**0.0034**
PLCE1	10	255	1.96E-04	rs1998709	-2.30	**0.0009**	rs10786152	-1.12	**0.0067**
CACNB2	10	542	9.23E-05	rs10741083	-1.77	**0.0111**	rs1409204	-1.17	**0.0483**
MECOM	3	563	8.88E-05	rs16853637	-2.66	**0.0059**	rs16853637	-1.79	**0.0023**
ULK4	3	512	9.77E-05	rs1874354	-6.18	**0.0022**	rs17284313	-1.49	**0.0309**
Gene Dense Region (Multiple Genes within ±100 kb of Index SNP)									
FGF5	4	40	1.25E-03	rs7686601	-1.78	**0.0115**	rs7686601	-1.06	**0.0139**
PRDM8	4			rs7686601	-1.78	**0.0115**	rs7686601	-1.06	**0.0139**
MOV10	1	24	2.08E-03	rs3795819	-1.44	0.0793	rs6537748	-0.48	0.2457
CAPZA1	1			rs3795819	-1.44	0.0793	rs6537748	-0.48	0.2457
NT5C2	10	49	1.02E-03	rs2274341	-0.60	0.4712	rs7074395	-0.60	0.1758
CNNM2	10			rs2274341	-0.60	0.4712	rs7074395	-0.60	0.1758
ADM	11	35	1.43E-03	rs11822548	3.28	**0.0388**	rs16907654	1.20	0.0620
Gene Sparse Region (Only One Gene within ±100 kb of Index SNP)									
ATP2B1	12	14	3.57E-03	rs11105358	-2.40	**0.0006***	rs11105358	-1.41	**0.0010***
NPR3	5	25	2.00E-03	rs1173726	1.51	0.1442	rs1173766	0.63	0.1390
SLC4A7	3	24	2.08E-03	rs820428	-1.78	0.1461	rs6551207	1.59	**0.0361**

Bold:$p<0.05$.*P-value that passed Bonferroni Corrected Alpha-Level (0.05/number of proximal SNPs for each specific gene).

Table 3.5. Significant results from association of proximal SNPs and gene expression.

Gene	Chr	SNP Selection Range		Proximal SNP Number	Bonferroni Corrected Alpha-Level	No. SNPs associated with Gene Expression (p<0.05)	No. of SNPs based on Bonferroni Correction	Top SNP with lowest P-value		
		Star	End					SNP	Beta	P-value
ULK4	3	41,268,091	42,023,660	512	9.77E-05	278	158	rs3934103	0.322	8.72E-58
MTHFR	1	11,825,787	11,886,115	66	7.58E-04	46	38	rs2066462	0.150	7.74E-18
FES	15	91,407,665	91,459,006	25	2.00E-03	10	10	rs8032315	-0.065	6.37E-12
FGF5	4	81,144,723	81,207,742	40	1.25E-03	19	11	rs7688596	-0.117	4.05E-09
GOSR2	17	44,980,486	45,038,732	48	1.04E-03	12	4	rs17676978	0.073	4.71E-08
SLC4A7	3	27,517,909	27,434,215	24	2.08E-03	15	8	rs13082711	-0.083	1.29E-07
PLEKHA7	11	16,789,212	17,055,963	299	1.67E-04	48	15	rs10832696	-0.135	3.89E-07
CAPZA1	1	113,196,543	113,182,075	24	2.08E-03	15	10	rs2932534	-0.044	9.58E-07
CSK	15	75,054,425	75,115,539	28	1.79E-03	18	18	rs4886606	0.037	6.88E-06
ARHGAP42	11	100,538,407	100,881,656	497	1.01E-04	174	36	rs12283290	0.155	1.65E-05
GUCY1A3	4	156,567,862	156,678,212	103	4.85E-04	9	3	rs2625276	-0.382	2.37E-04
MOV10	1	113,196,543	113,237,048	24	2.08E-03	7	2	rs3748655	-0.051	4.24E-04
HFE	6	26,067,509	26,115,469	53	9.43E-04	10	6	rs198851	0.050	4.30E-04
BAT3	6	31,586,805	31,640,170	65	7.69E-04	19	1	rs2736176	0.024	7.27E-04
SLC39A8	4	103,162,821	103,286,655	133	3.76E-04	17	0	rs17032436	0.087	0.0025
ADM	11	10,330,538	10,346,642	35	1.43E-03	20	0	rs16907654	-0.086	0.0028
CACNB2	10	18,409,606	18,850,688	542	9.23E-05	39	0	rs7901640	0.128	0.0030
MECOM	3	168,781,287	169,401,563	563	8.88E-05	7	0	rs17507954	0.036	0.0112
NT5C2	10	104,826,178	104,867,774	49	1.02E-03	22	0	rs11191553	0.032	0.0119
NPR3	5	32,795,028	32,731,665	25	2.00E-03	9	0	rs7723507	-0.039	0.0276
ZNF652	17	47,346,569	47,459,835	66	7.58E-04	18	0	rs11657365	0.035	0.0316
CNNM2	10	104,826,178	104,698,075	49	1.02E-03	11	0	rs11191558	-0.030	0.0384
SH2B3	12	111,823,752	111,909,426	20	2.50E-03	5	0	rs4766573	-0.024	0.0388
PLCD3	17	43,169,009	43,229,891	51	9.80E-04	1	0	rs2306828	0.022	0.0499
PLCE1	10	95,733,746	96,108,146	255	1.96E-04	0	0	rs17109869	0.015	0.0569
PRDM8	4	81,144,723	81,138,657	40	1.25E-03	0	0	rs16998073	0.024	0.0969
ATP2B1	12	90,040,586	90,001,828	14	3.57E-03	0	0	rs17782841	-0.026	0.0970
Total	--	----------	----------	3537	----------	829	320	----------	----------	----------

Table 3.6. Summary results from association of SNPs, gene expression and blood pressure.†

Gene	Chr	Proximal SNP Number	No. of Proximal SNPs associated with		Association of SBP & Gene Expression		Association of DBP & Gene Expression	
			SBP & Gene Expression (p<0.05)	DBP & Gene Expression (p<0.05)	Beta	P-value	Beta	P-value
Gene Region with Index SNP inside of it								
HFE	6	53	0	0	-3.69	0.2597	-1.60	0.4184
FES	15	25	0	0	-4.46	0.1691	-2.67	0.1723
BAT3	6	65	0	0	9.53	0.0576	3.80	0.2087
GOSR2	17	48	0	0	10.11	**0.0062**	5.16	**0.0204**
MTHFR	1	66	7	7	-1.03	0.7441	-1.57	0.4096
PLCD3	17	51	0	1	-7.73	**0.0396**	-4.07	0.0724
CSK	15	28	0	0	-1.20	0.7569	-3.26	0.1609
SH2B3	12	20	0	0	2.24	0.5005	-0.07	0.9733
GUCY1A3	4	103	0	0	-1.04	0.1177	-0.57	0.1561
ZNF652	17	66	0	0	2.94	0.2155	3.46	**0.0155**
SLC39A8	4	133	0	0	5.84	**0.0120**	3.12	**0.0256**
PLEKHA7	11	299	5	19	-2.72	0.1267	-0.97	0.3672
ARHGAP42	11	497	5	2	-1.11	0.3753	0.83	0.2685
PLCE1	10	255	0	0	-6.40	0.1291	-3.34	0.1892
CACNB2	10	542	0	0	-1.28	0.0906	-0.85	0.0615
MECOM	3	563	0	0	-3.38	0.3565	-4.51	**0.0409**
ULK4‡	3	512	28	35	4.64	**0.0149**	4.03	**0.0005***
Gene Dense Region (Multiple Genes within ±100 kb of Index SNP)								
FGF5	4	40	1	1	2.09	0.2019	0.80	0.4148
PRDM8	4		0	0	1.46	0.6054	-7.73	**0.0396**
MOV10	1	24	0	0	1.58	0.5815	1.01	0.5579
CAPZA1	1		0	0	2.85	0.4727	5.53	**0.0206**
NT5C2	10	49	0	0	3.59	0.1768	2.53	0.1138
CNNM2	10		0	0	-5.82	0.1020	-4.65	**0.0300**
ADM	11	35	0	0	0.43	0.8050	0.75	0.4715
Gene Sparse Region (Only One Gene within ±100 kb of Index SNP)								
ATP2B1	12	14	0	0	3.18	0.1462	1.37	0.3001
NPR3	5	25	0	0	-2.86	0.1715	-2.88	**0.0219**
SLC4A7	3	24	0	0	2.78	0.2378	1.09	0.4395

†Genes were sorted by gene size. See Appendix 3.2. ‡*ULK4* is the only gene that has significant associations in three models: a,b, c. Bold: p-value <0.05. *Bonferroni Corrected Alpha-Level: 0.00185 (0.05/27 genes).

Figure 3.2. The 35 DBP-associated SNPs, 28 SBP-associated SNPs, and their *cis*-effects on *ULK4* expression.

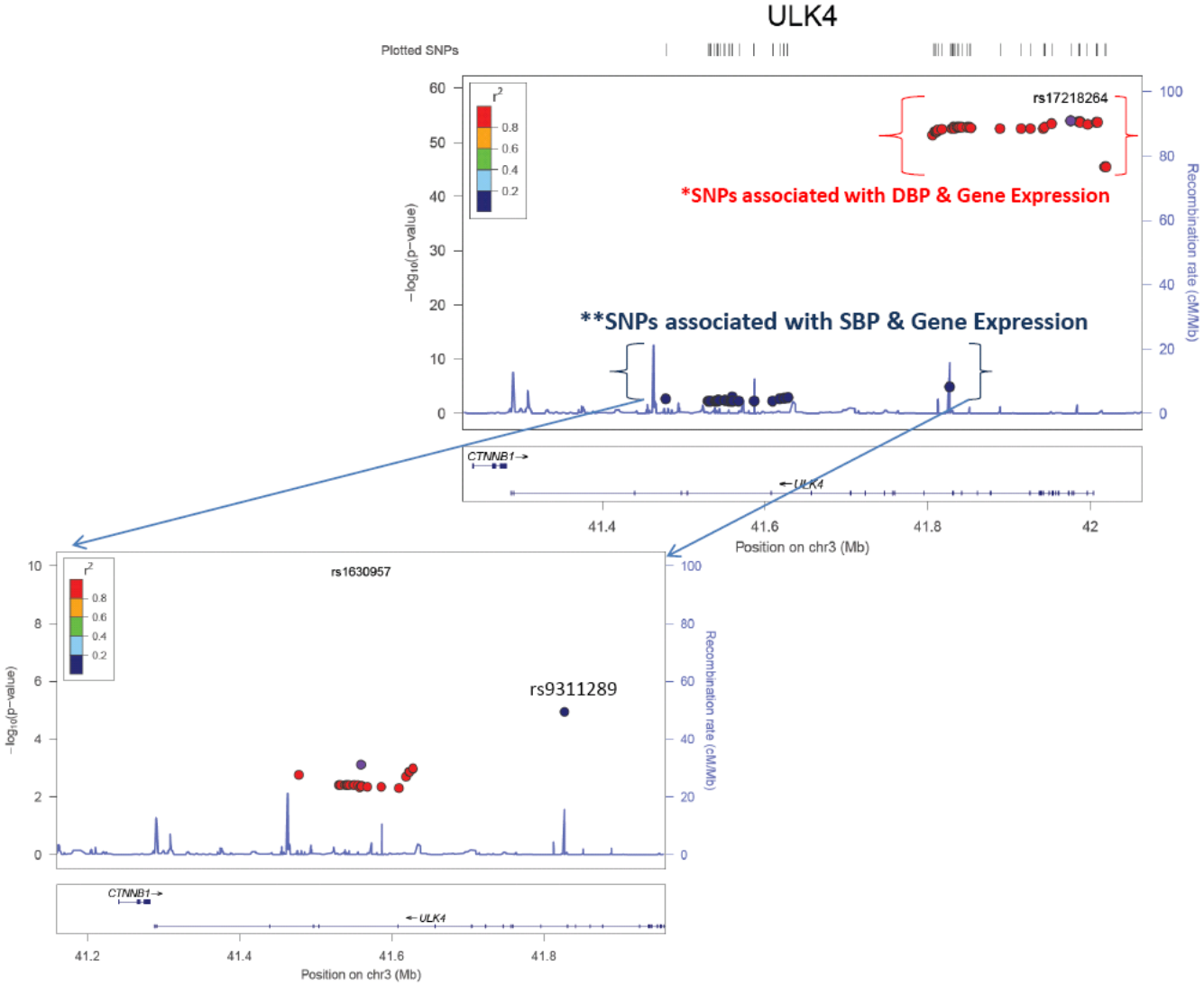
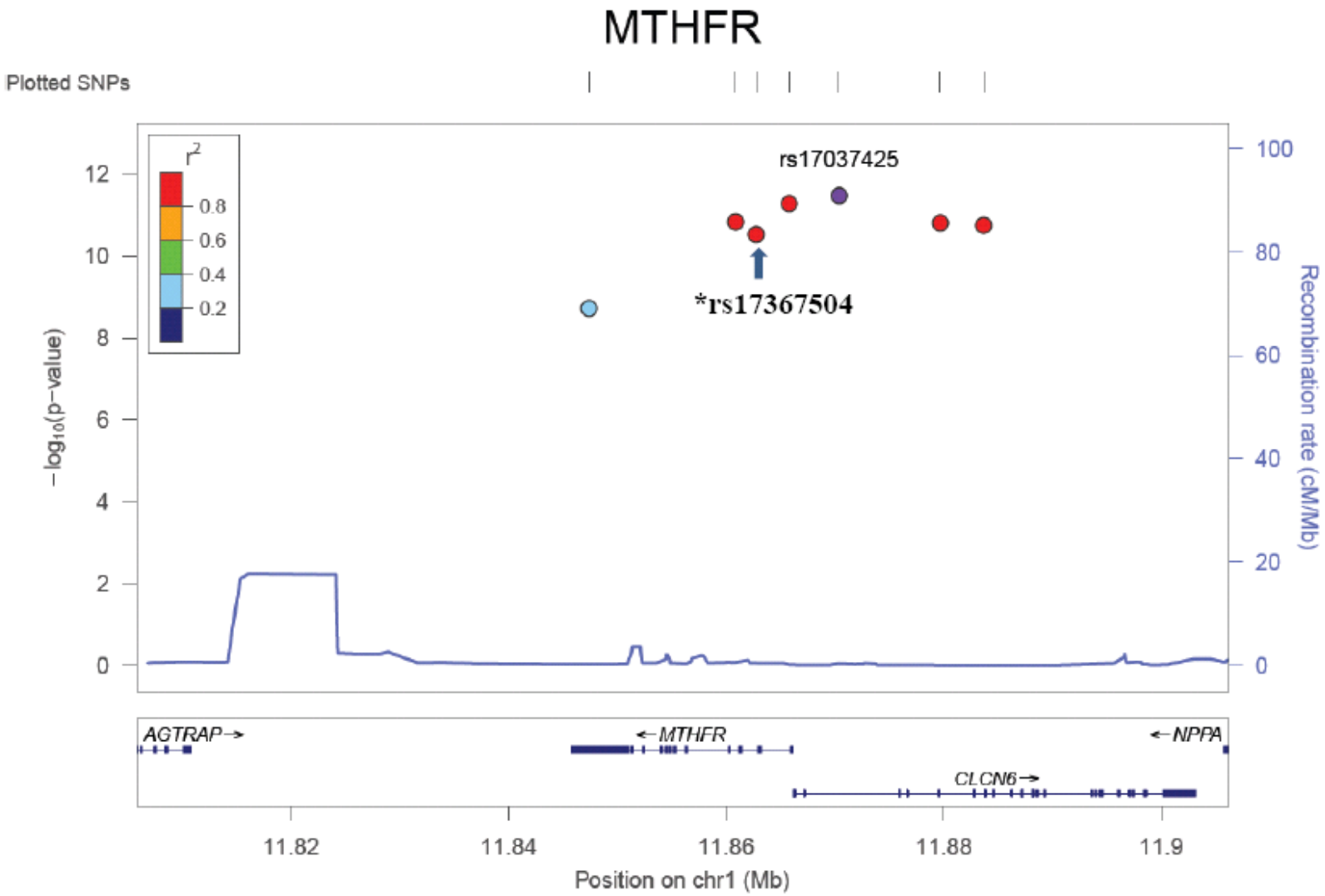
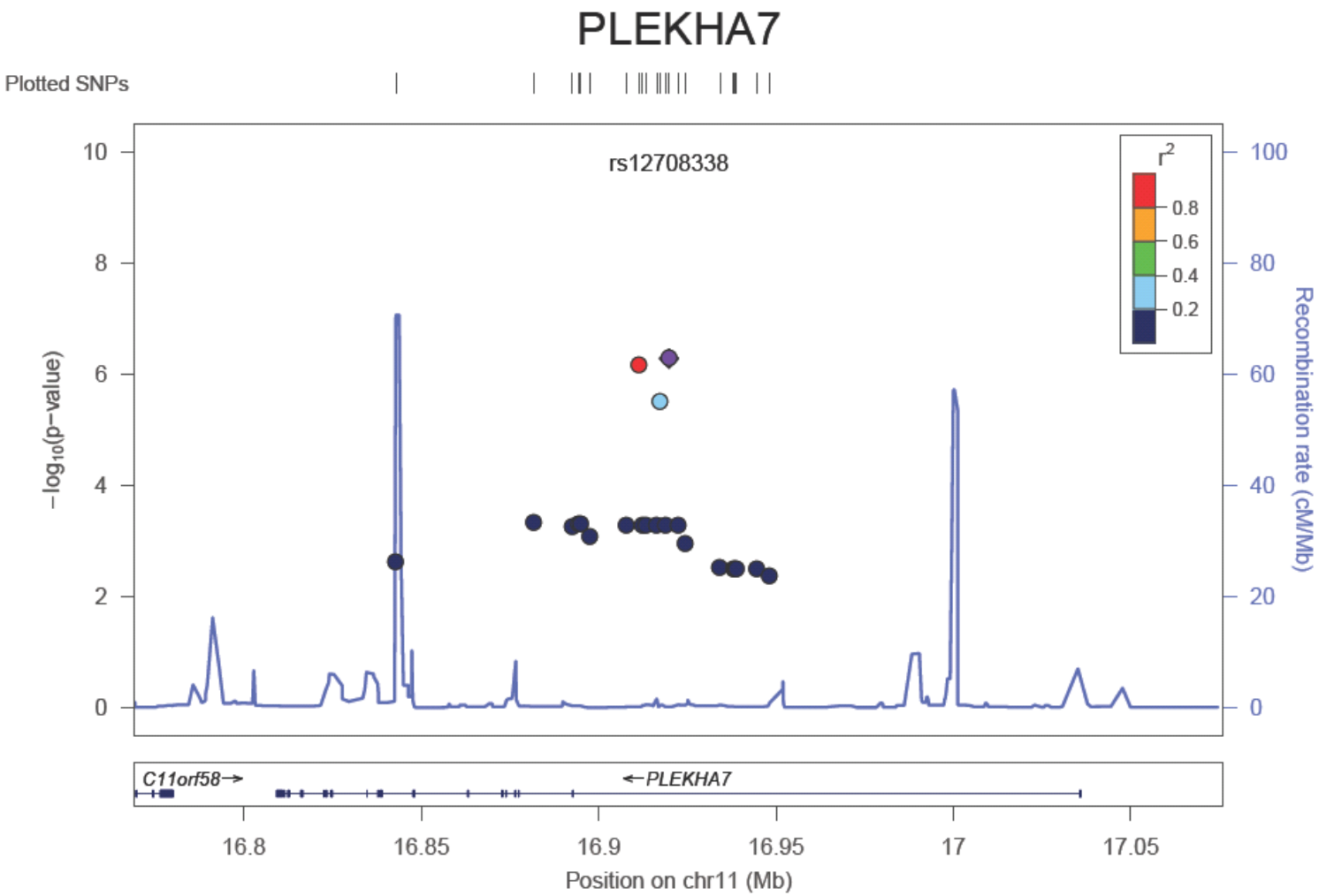


Figure 3.3. The 7 SNPs and their *cis*-effects on *MTHFR* expression.



*rs17367504 is the index SNP reported from ICBP. **Purple circle is the SNP with lowest p-value.

Figure 3.4. The 23 SNPs and their *cis*-effects on *PLEKHA7* expression.



*Purple circle is the SNP with lowest p-value.

Appendix 3.1. Summary association results for 29 blood pressure associated SNPs from the International Consortium for Blood Pressure (ICBP).

Index SNP	Nearest Gene	Function	Chr	SNP Position	Coded Allele/Non-Coded allele	Coded Allele Freq.	SBP		DBP	
							Beta	P-value	Beta	P-value
16 Novel loci identified by ICBP										
rs2932538	*MOV10*	Intergenic	1	113,216,543	A/G	0.25	-0.39	1.20E-09	-0.24	9.90E-10
rs13082711	*SLC4A7*	Intron	3	27,537,909	C/T	0.22	0.32	1.50E-06	0.24	3.80E-09
rs419076	*MECOM*	Intron	3	169,100,886	C/T	0.53	-0.41	1.80E-13	-0.24	2.10E-12
rs13107325	*SLC39A8*	Missense	4	103,188,709	C/T	0.95	0.98	3.30E-14	0.68	2.30E-17
rs13139571	*GUCY1A3*	Intron	4	156,645,513	A/C	0.24	-0.32	1.20E-06	-0.26	2.20E-10
rs1173771	*NPR3*	Intergenic	5	32,815,028	A/G	0.4	-0.5	1.80E-16	-0.26	9.10E-12
rs11953630	*EBF1*	Intergenic	5	157,845,402	T/C	0.37	-0.41	3.00E-11	-0.28	3.80E-13
rs1799945	*HFE*	Missense	6	26,091,179	C/G	0.86	-0.63	7.70E-12	-0.46	1.50E-15
rs805303	*BAT3*	Intron	6	31,616,366	A/G	0.39	-0.38	1.50E-11	-0.23	3.00E-11
rs932764	*PLCE1*	Intron	10	95,895,940	A/G	0.56	-0.48	7.10E-16	-0.19	8.10E-07
rs7129220	*ADM*	Intergenic	11	10,350,538	A/G	0.11	0.62	3.00E-12	0.3	6.40E-08
rs633185	*ARHGAP42*	Intron	11	100,593,538	C/G	0.72	0.57	1.20E-17	0.33	2.00E-15
rs2521501	*FES*	Intron	15	91,437,388	A/T	0.69	-0.65	5.20E-19	-0.36	1.90E-15
rs17608766	*GOSR2*	Intron	17	45,013,271	C/T	0.14	0.56	1.10E-10	0.13	0.017
rs1327235	*JAG1*	Intergenic	20	10,969,030	G/A	0.46	0.34	1.90E-08	0.3	1.40E-15
rs6015450	*EDN3*	Intergenic	20	57,751,117	G/A	0.12	0.9	3.90E-23	0.56	5.60E-23
12 Loci that were identified in ICBP and also have been reported in CHARGE and Global BPgen										
rs17367504	*MTHFR*	Intron	1	11,862,778	A/G	0.85	0.9	8.70E-22	0.55	3.50E-19
rs3774372	*ULK4*	Missense	3	41,877,414	C/T	0.17	0.07	0.39	0.37	9.00E-14
rs1458038	*FGF5*	Intergenic	4	81,164,723	C/T	0.71	-0.7	1.50E-23	-0.46	8.50E-25
rs1813353	*CACNB2*	Intron	10	18,707,448	C/T	0.32	-0.57	2.60E-12	-0.42	2.30E-15
rs4590817	*C10orf107*	Intron	10	63,467,553	G/C	0.84	0.65	4.00E-12	0.42	1.30E-12
rs11191548	*NT5C2*	Intron	10	104,846,178	A/G	0.85	0.9	8.70E-22	0.55	3.50E-19
rs381815	*PLEKHA7*	Intron	11	16,902,268	C/T	0.17	0.07	0.39	0.37	9.00E-14
rs17249754	*ATP2B1*	Intron	12	90,060,586	C/T	0.71	-0.7	1.50E-23	-0.46	8.50E-25
rs3184504	*SH2B3*	Missense	12	111,884,608	C/T	0.32	-0.57	2.60E-12	-0.42	2.30E-15
rs10850411	*TBX3*	Intergenic	12	115,387,796	T/C	0.7	0.35	5.40E-08	0.25	5.40E-10
rs1378942	*CSK*	Intron	15	75,077,367	A/C	0.65	-0.61	5.70E-23	-0.42	2.70E-26
rs12940887	*ZNF652*	Intron	17	47,402,807	C/T	0.62	-0.36	1.80E-10	-0.27	2.30E-14
Locus that was identified in Global BPgen but not in CHARGE or ICBP										
rs12946454	*PLCD3**	Intron	17	43,208,121	A/T	0.73	0.57	1.00E-08	NA	NA

*Because PLCD3 is not identified in ICBP, the statistical results for this gene was extracted from Global BPgen.

Appendix 3.2. Summary of twenty-nine blood pressure associated index SNPs and their closest genes. †

Index SNP	Nearest Gene	Chr	SNP Position	Gene Start	Gene End	Gene Size (bp)	SNP Selection Region		SNP Number (Total=3537)
							Star	End	
Gene Region with Index SNP inside of it									
rs1799945	*HFE*	6	26,091,179	26,087,509	26,095,469	7,960	26,067,509	26,115,469	53
rs2521501	*FES*	15	91,437,388	91,427,665	91,439,006	11,341	91,407,665	91,459,006	25
rs805303	*BAT3*	6	31,616,366	31,606,805	31,620,170	13,365	31,586,805	31,640,170	65
rs17608766	*GOSR2*	17	45,013,271	45,000,486	45,018,732	18,246	44,980,486	45,038,732	48
rs17367504	*MTHFR*	1	11,862,778	11,845,787	11,866,115	20,328	11,825,787	11,886,115	66
rs12946454	*PLCD3*	17	43,208,121	43,189,009	43,209,891	20,882	43,169,009	43,229,891	51
rs1378942	*CSK*	15	75,077,367	75,074,425	75,095,539	21,114	75,054,425	75,115,539	28
rs3184504	*SH2B3*	12	111,884,608	111,843,752	111,889,426	45,674	111,823,752	111,909,426	20
rs13139571	*GUCY1A3*	4	156,645,513	156,587,862	156,658,212	70,350	156,567,862	156,678,212	103
rs12940887	*ZNF652*	17	47,402,807	47,366,569	47,439,835	73,266	47,346,569	47,459,835	66
rs13107325	*SLC39A8*	4	103,188,709	103,182,821	103,266,655	83,834	103,162,821	103,286,655	133
rs381815	*PLEKHA7*	11	16,902,268	16,809,212	17,035,963	226,751	16,789,212	17,055,963	299
rs633185	*ARHGAP42*	11	100,593,538	100,558,407	100,861,656	303,249	100,538,407	100,881,656	497
rs932764	*PLCE1*	10	95,895,940	95,753,746	96,088,146	334,400	95,733,746	96,108,146	255
rs1813353	*CACNB2*	10	18,707,448	18,429,606	18,830,688	401,082	18,409,606	18,850,688	542
rs419076	*MECOM*	3	169,100,886	168,801,287	169,381,563	580,276	168,781,287	169,401,563	563
rs3774372	*ULK4*	3	41,877,414	41,288,091	42,003,660	715,569	41,268,091	42,023,660	512
Gene Dense Region (Multiple Genes within ±100 kb of Index SNP)									
rs1458038	*FGF5*	4	81,164,723	81,187,742	81,212,170	24,428	81,144,723	81,207,742	40
	PRDM8	4	81,164,723	81,118,657	81,125,482	6,825	81,144,723	81,138,657	
rs2932538	*MOV10*	1	113,216,543	113,217,048	113,243,367	26,319	113,196,543	113,237,048	24
	CAPZA1	1	113,216,543	113,162,075	113,214,241	52,166	113,196,543	113,182,075	
rs11191548	*NT5C2*	10	104,846,178	104,847,774	104,953,063	105,289	104,826,178	104,867,774	49
	CNNM2	10	104,846,178	104,678,075	104,838,344	160,269	104,826,178	104,698,075	
rs7129220	*ADM**	11	10,350,538	10,326,642	10,328,923	2,281	10,330,538	10,346,642	35
Gene Sparse Region (Only One Gene within ±100 kb of Index SNP)									
rs17249754	*ATP2B1*	12	90,060,586	89,981,828	90,049,844	68,016	90,040,586	90,001,828	14
rs1173771	*NPR3*	5	32,815,028	32,711,665	32,787,252	75,587	32,795,028	32,731,665	25
rs13082711	*SLC4A7*	3	27,537,909	27,414,215	27,498,245	84,030	27,517,909	27,434,215	24
Index SNPs Excluded (No gene within ±100 kb of Index SNP)									
rs10850411	*TBX3*	12	115,387,796	115,108,059	115,121,969	13,910	NA	NA	NA
rs6015450	*EDN3*	20	57,751,117	57,875,499	57,901,047	25,548	NA	NA	NA
rs1327235	*JAG1*	20	10,969,030	10,618,332	10,654,694	36,362	NA	NA	NA
rs11953630	*EBF1*	5	157,845,402	158,122,924	158,526,788	403,864	NA	NA	NA
SNPs Excluded (No gene expression data available)									
rs4590817	*C10orf107*	10	63,467,553	63,422,719	63,526,089	103,370	NA	NA	NA

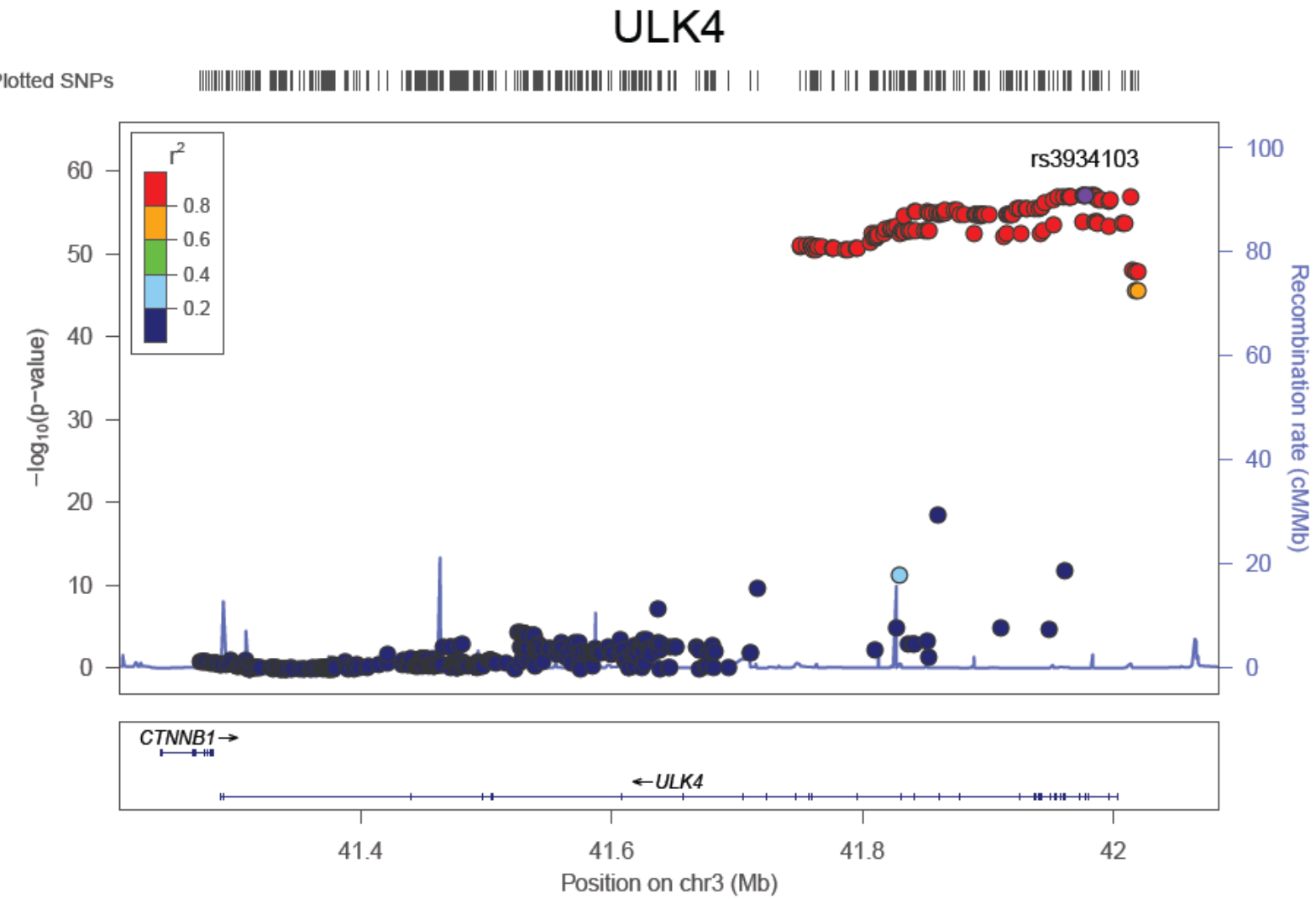
*There is one gene ,"SBF2", upstream of the index SNP but no genes downstream. †Genes were sorted by gene size.

Appendix 3.3. Association of index SNP, and blood pressure in ICBP and GENOA.

Index SNP	Nearest Gene	Chr	SNP Position	ICBP						GENOA				
				Coded /Non-Coded Allele	Coded Allele Freq.	SBP		DBP		Coded Allele Freq.	SBP		DBP	
						Beta	P-value	Beta	P-value		Beta	P-value	Beta	P-value
Novel loci identified by ICBP														
rs2932538	*MOV10*	1	113,216,543	A/G	0.25	-0.39	1.20E-09	-0.24	9.90E-10	0.25	-1.08	0.189	-0.05	0.928
rs13082711	*SLC4A7*	3	27,537,909	C/T	0.22	0.32	1.50E-06	0.24	3.80E-09	0.26	0.10	0.902	0.14	0.771
rs419076	*MECOM*	3	169,100,886	C/T	0.53	-0.41	1.80E-13	-0.24	2.10E-12	0.53	-0.61	0.371	-0.25	0.555
rs13107325	*SLC39A8*	4	103,188,709	C/T	0.95	0.98	3.30E-14	0.68	2.30E-17	0.93	0.14	0.918	-0.42	0.615
rs13139571	*GUCY1A3*	4	156,645,513	A/C	0.24	-0.32	1.20E-06	-0.26	2.20E-10	0.26	0.45	0.566	0.02	0.969
rs1173771	*NPR3*	5	32,815,028	A/G	0.40	-0.50	1.80E-16	-0.26	9.10E-12	0.41	-0.32	0.652	-0.20	0.637
rs1799945	*HFE*	6	26,091,179	C/G	0.86	-0.63	7.70E-12	-0.46	1.50E-15	0.86	0.87	0.379	0.70	0.248
rs805303	*BAT3*	6	31,616,366	A/G	0.39	-0.38	1.50E-11	-0.23	3.00E-11	0.39	-0.11	0.869	-0.20	0.631
rs932764	*PLCE1*	10	95,895,940	A/G	0.56	-0.48	7.10E-16	-0.19	8.10E-07	0.52	**-1.49**	**0.027**	**-0.95**	**0.020**
rs7129220	*ADM*	11	10,350,538	A/G	0.11	0.62	3.00E-12	0.30	6.40E-08	0.11	-1.27	0.238	-1.15	0.078
rs633185	*ARHGAP42*	11	100,593,538	C/G	0.72	0.57	1.20E-17	0.33	2.00E-15	0.72	-0.10	0.891	0.27	0.557
rs2521501	*FES*	15	91,437,388	A/T	0.69	-0.65	5.20E-19	-0.36	1.90E-15	0.64	-1.05	0.223	-0.28	0.590
rs17608766	*GOSR2*	17	45,013,271	C/T	0.14	0.56	1.10E-10	0.13	0.017	0.15	-1.23	0.216	-0.64	0.289
Loci that were identified in ICBP and also have been reported in CHARGE and Global BPgen														
rs17367504	*MTHFR*	1	11,862,778	A/G	0.85	0.90	8.70E-22	0.55	3.50E-19	0.86	**2.94**	**0.003**	**1.50**	**0.012**
rs3774372	*ULK4*	3	41,877,414	C/T	0.17	0.07	0.39	0.37	9.00E-14	0.18	0.24	0.784	0.41	0.443
rs1458038	*FGF5*	4	81,164,723	C/T	0.71	-0.70	1.50E-23	-0.46	8.50E-25	0.70	-0.97	0.225	-0.79	0.104
rs1813353	*CACNB2*	10	18,707,448	C/T	0.32	-0.57	2.60E-12	-0.42	2.30E-15	0.29	-0.37	0.619	-0.53	0.243
rs11191548	*NT5C2*	10	104,846,178	C/T	0.09	-1.10	6.90E-26	-0.46	9.40E-13	0.10	0.03	0.978	-0.19	0.788
rs381815	*PLEKHA7*	11	16,902,268	C/T	0.74	-0.58	5.30E-11	-0.35	5.30E-10	0.75	-1.05	0.186	-0.77	0.112
rs17249754	*ATP2B1*	12	90,060,586	A/G	0.16	-0.90	1.80E-18	-0.52	1.20E-14	0.15	**-2.17**	**0.024**	**-1.17**	**0.045**
rs3184504	*SH2B3*	12	111,884,608	C/T	0.53	-0.60	3.80E-18	-0.45	3.60E-25	0.50	0.66	0.364	0.03	0.943
rs1378942	*CSK*	15	75,077,367	A/C	0.65	-0.61	5.70E-23	-0.42	2.70E-26	0.66	-0.66	0.363	0.14	0.756
rs12940887	*ZNF652*	17	47,402,807	C/T	0.62	-0.36	1.80E-10	-0.27	2.30E-14	0.60	-0.25	0.716	-0.70	0.097
Locus that was identified in Global BPgen but not in CHARGE or ICBP														
rs12946454	*PLCD3*	17	43,208,121	A/T	0.73	0.57	1.00E-08	NA	NA	0.72	0.07	0.923	0.26	0.581

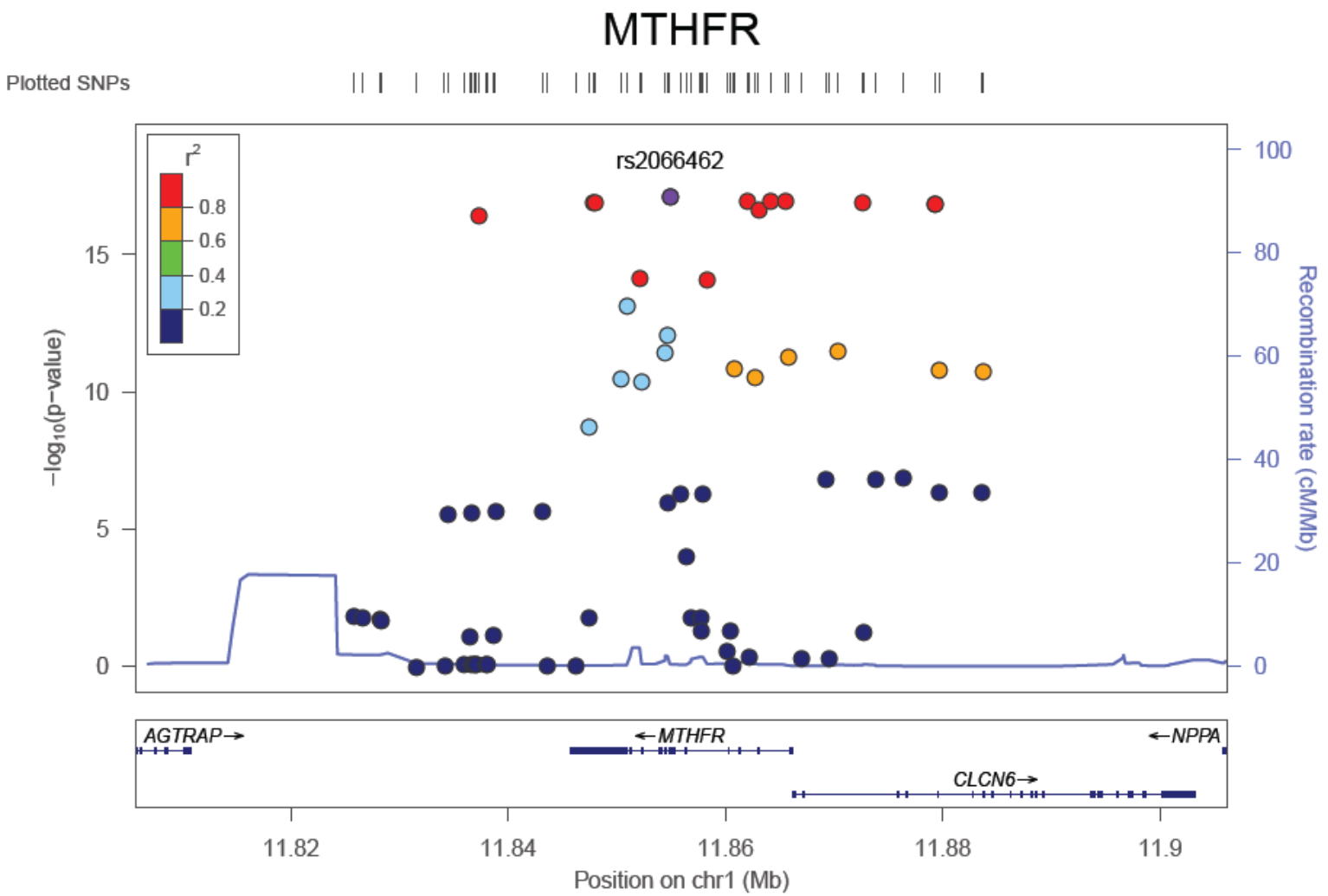
Bold:*p*-value<0.05 in GENOA.

Appendix 3.4. SNPs and their *cis*-effects on *ULK4* expression.



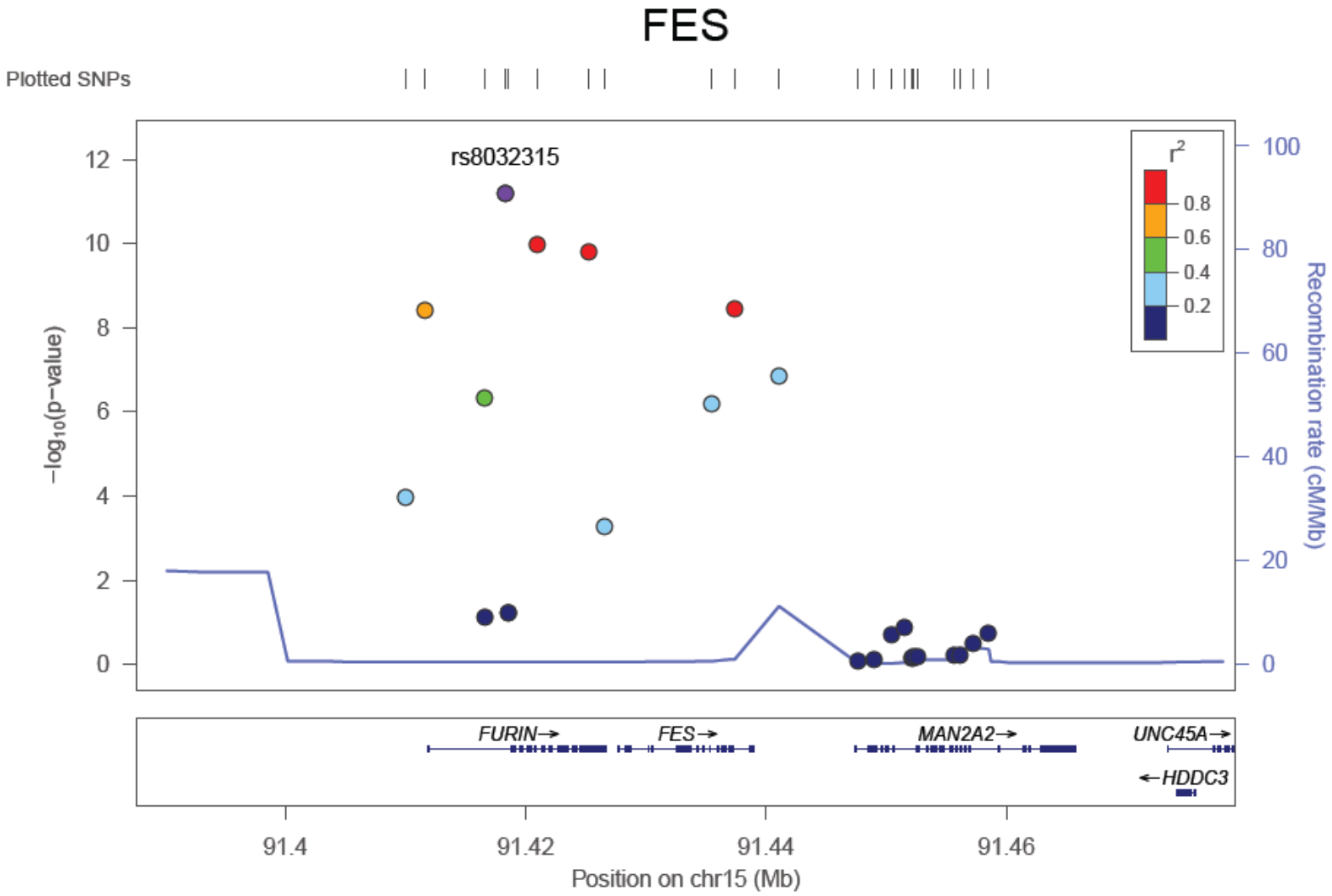
*Purple circle is the SNP with lowest p-value.

Appendix 3.5. SNPs and their *cis*-effects on *MTHFR* expression.



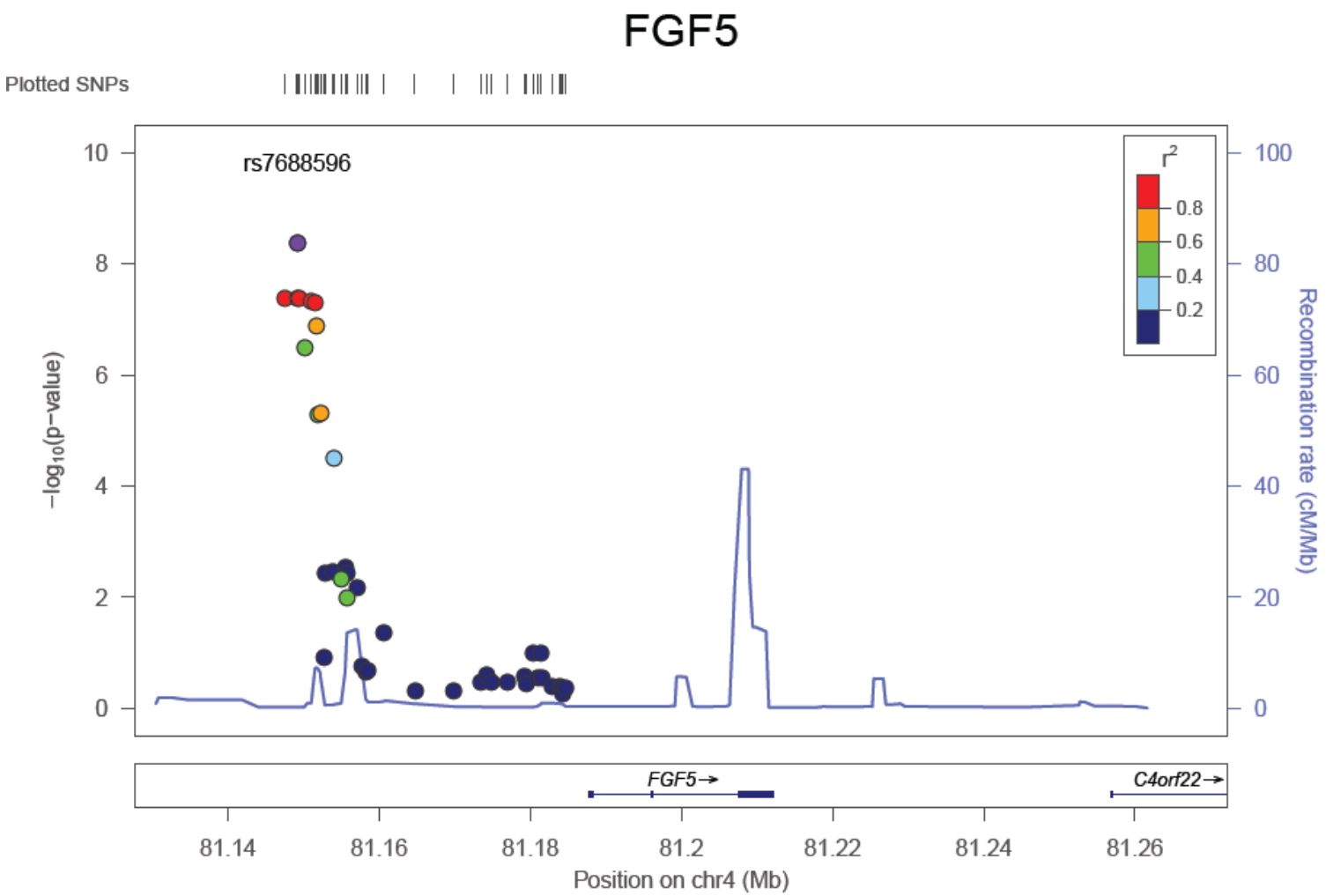
*Purple circle is the SNP with lowest p-value.

Appendix 3.6. SNPs and their *cis*-effects on *FES* expression.



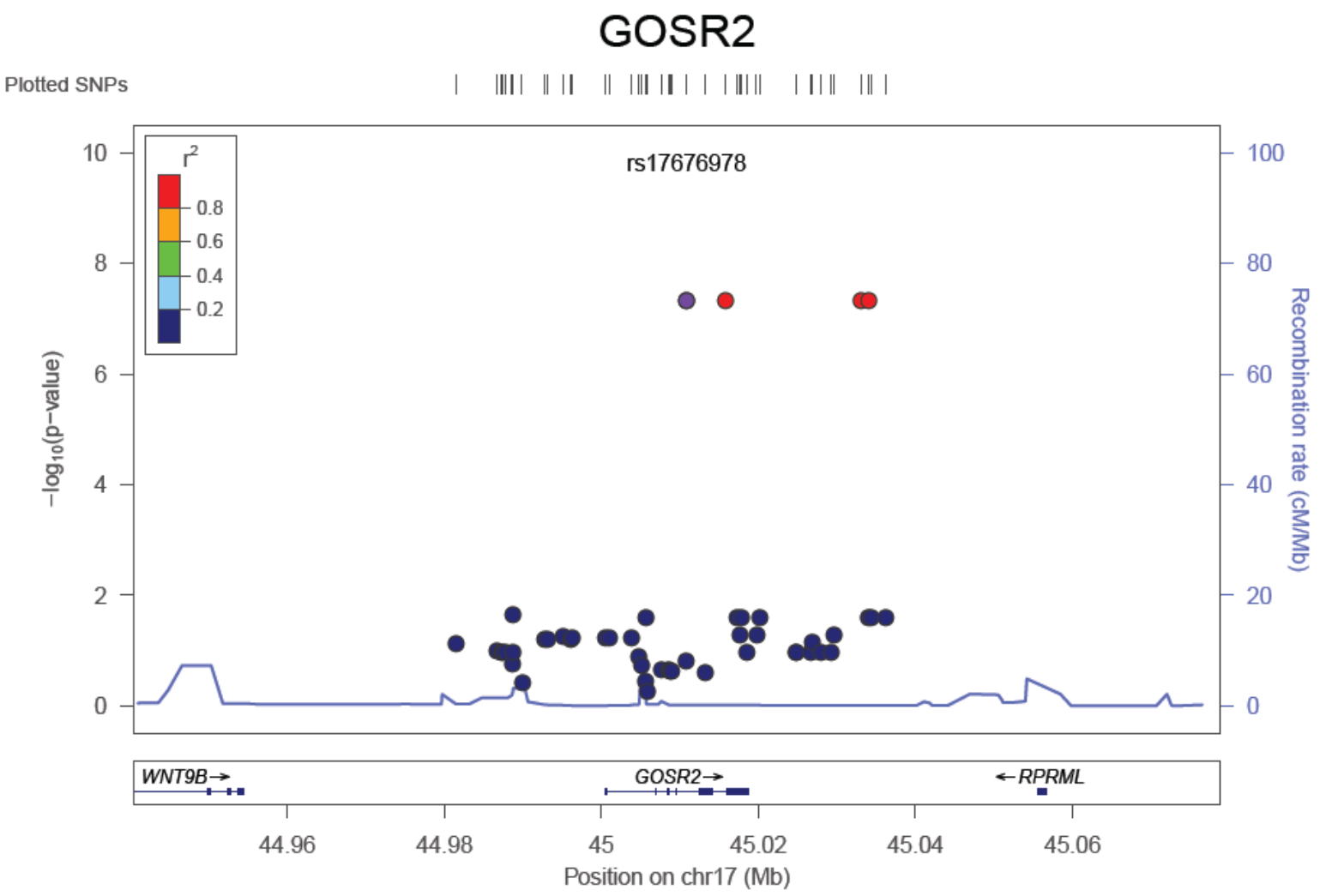
*Purple circle is the SNP with lowest p-value.

Appendix 3.7. SNPs and their *cis*-effects on *FGF5* expression.



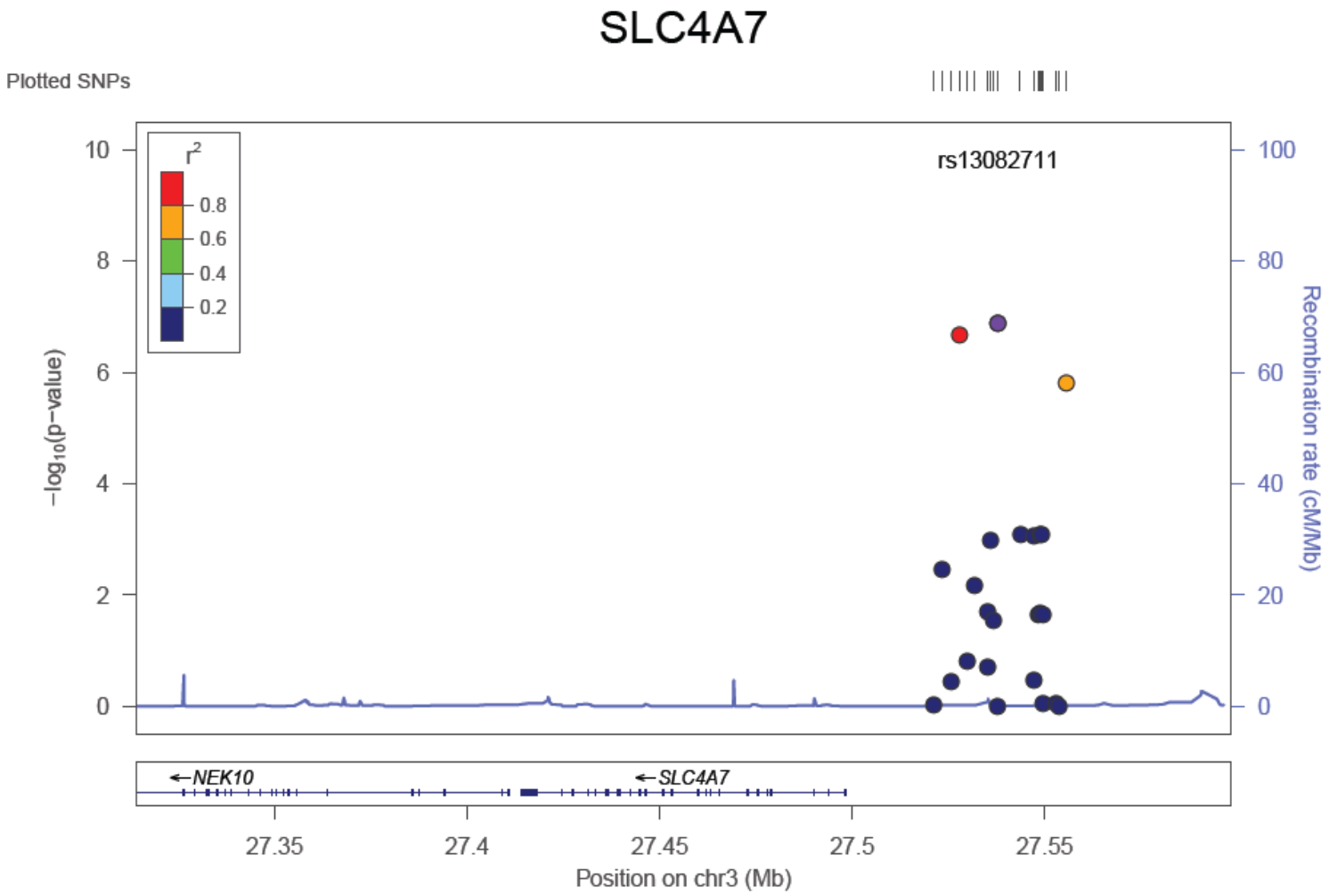
*Purple circle is the SNP with lowest p-value.

Appendix 3.8. SNPs and their *cis*-effects on *GOSR2* expression.



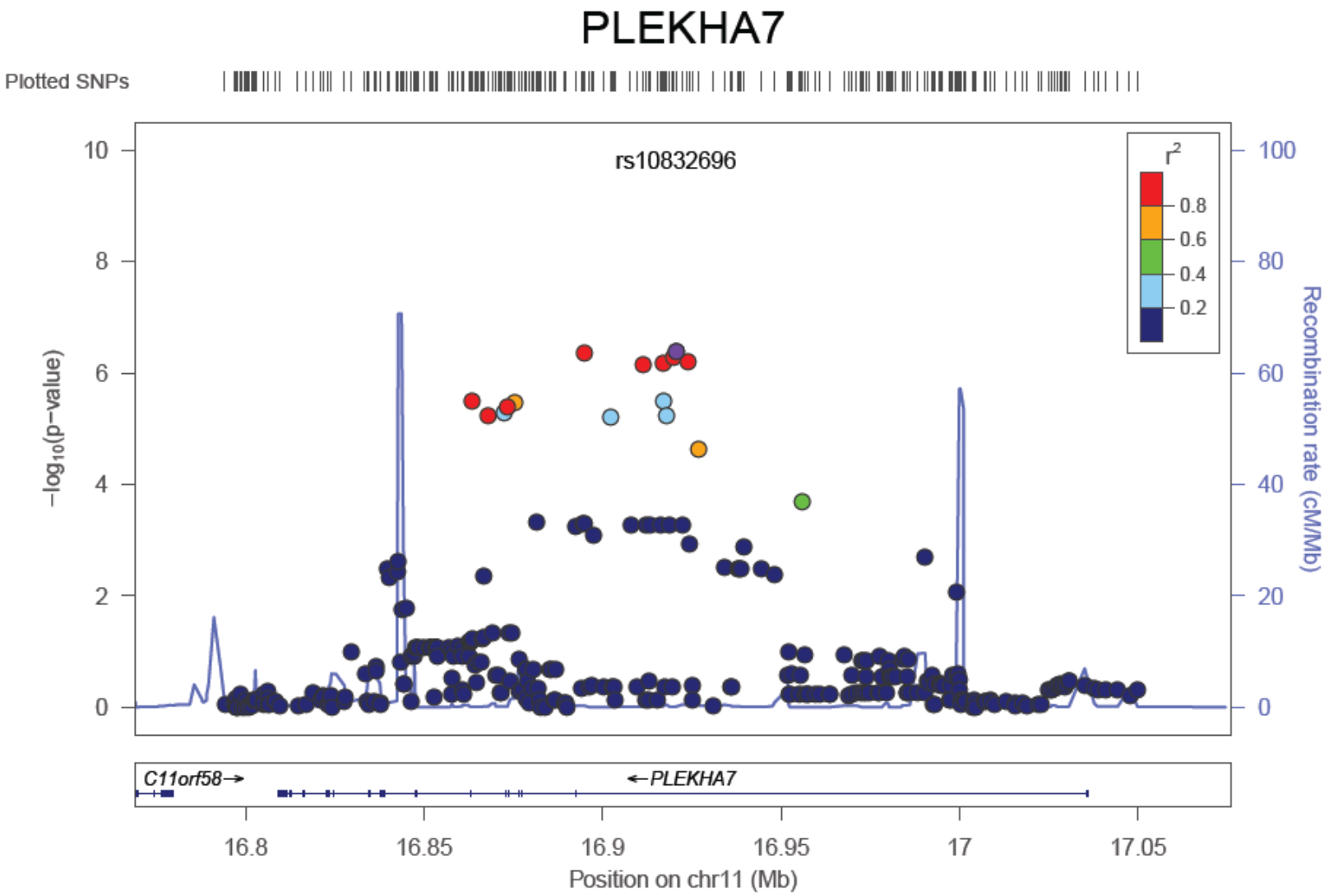
*Purple circle is the SNP with lowest p-value.

Appendix 3.9. SNPs and their *cis*-effects on *SLC4A7* expression.



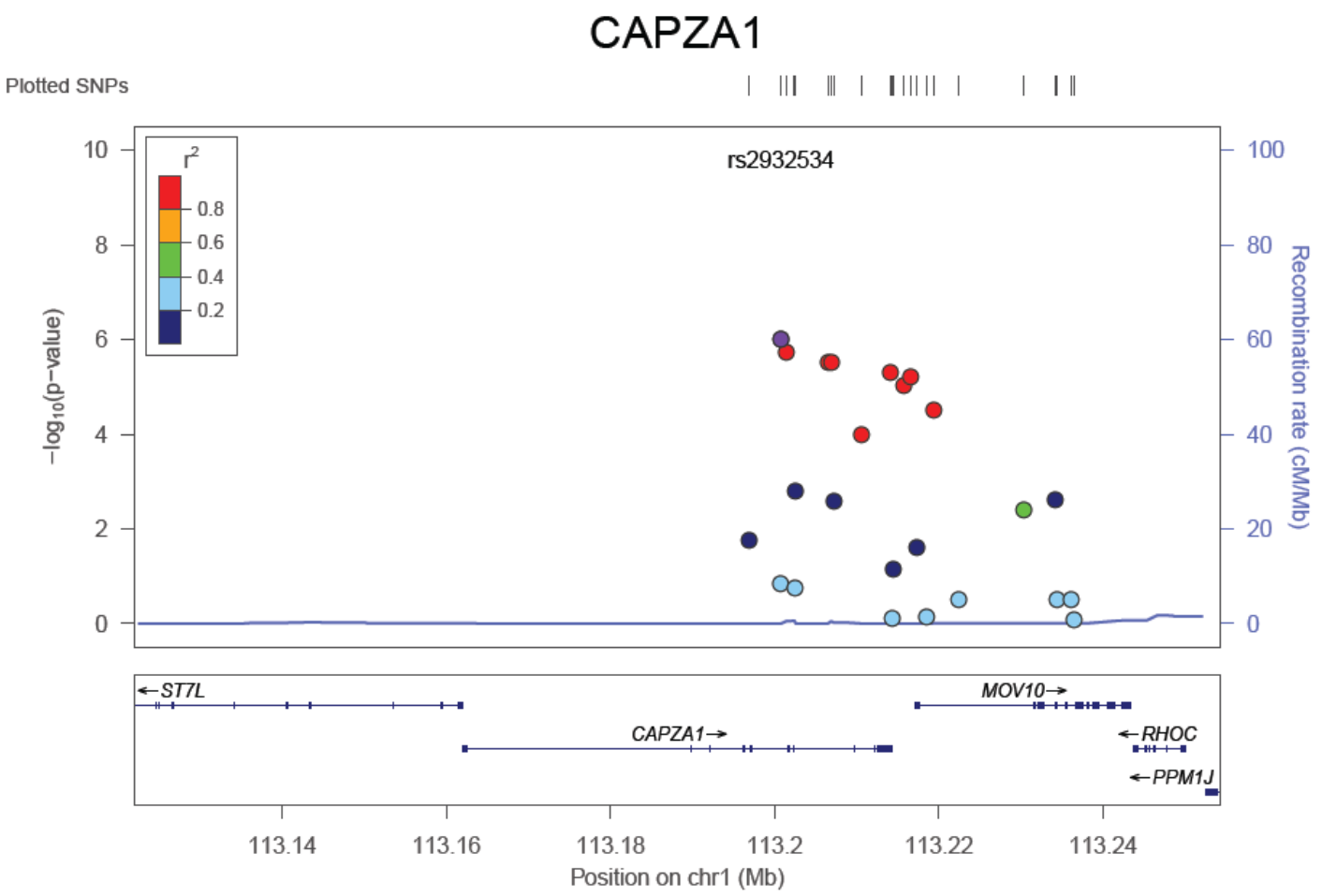
*Purple circle is the SNP with lowest p-value.

Appendix 3.10. SNPs and their *cis*-effects on *PLEKHA7* expression.



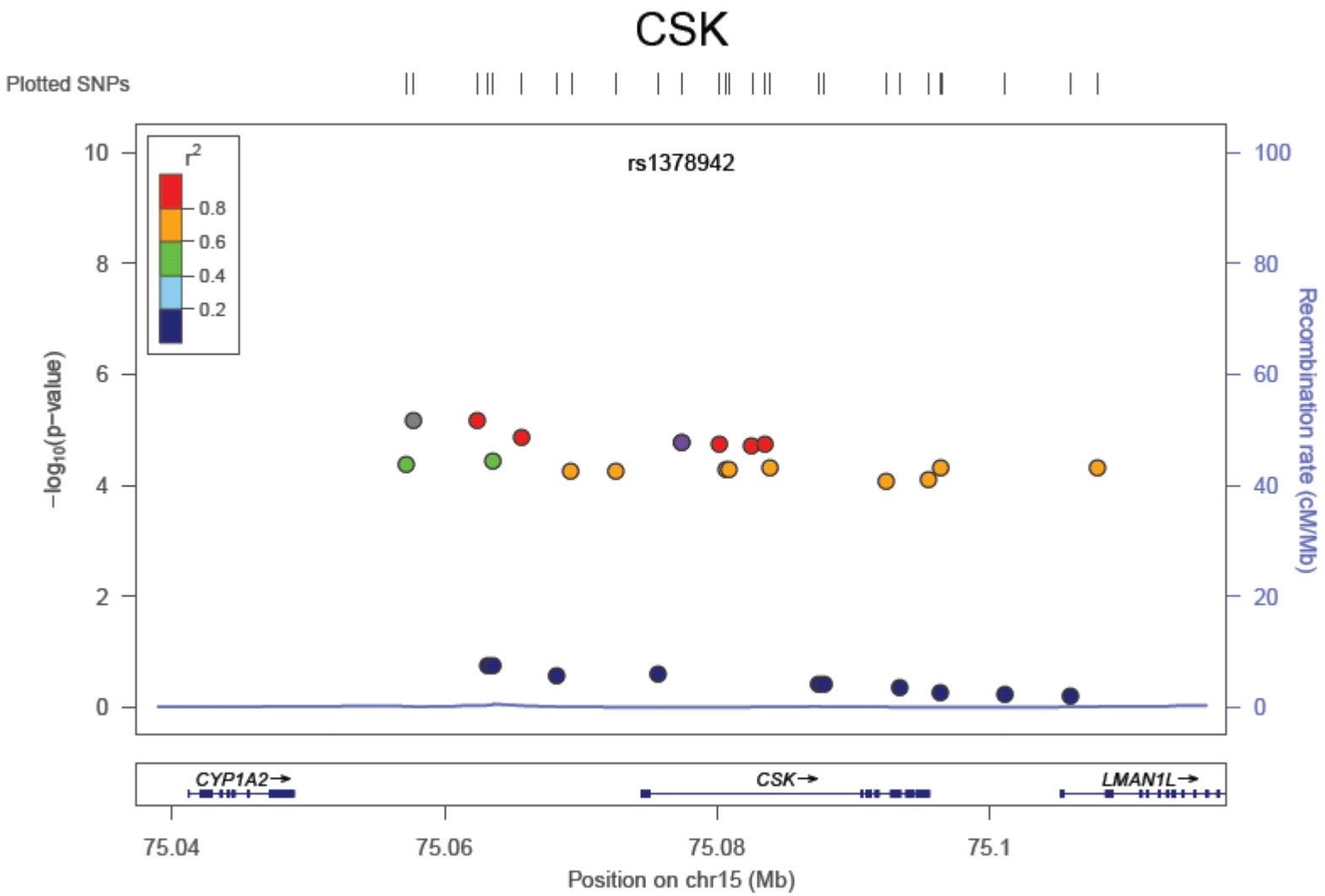
*Purple circle is the SNP with lowest p-value.

Appendix 3.11. SNPs and their *cis*-effects on *CAPZA1* expression.



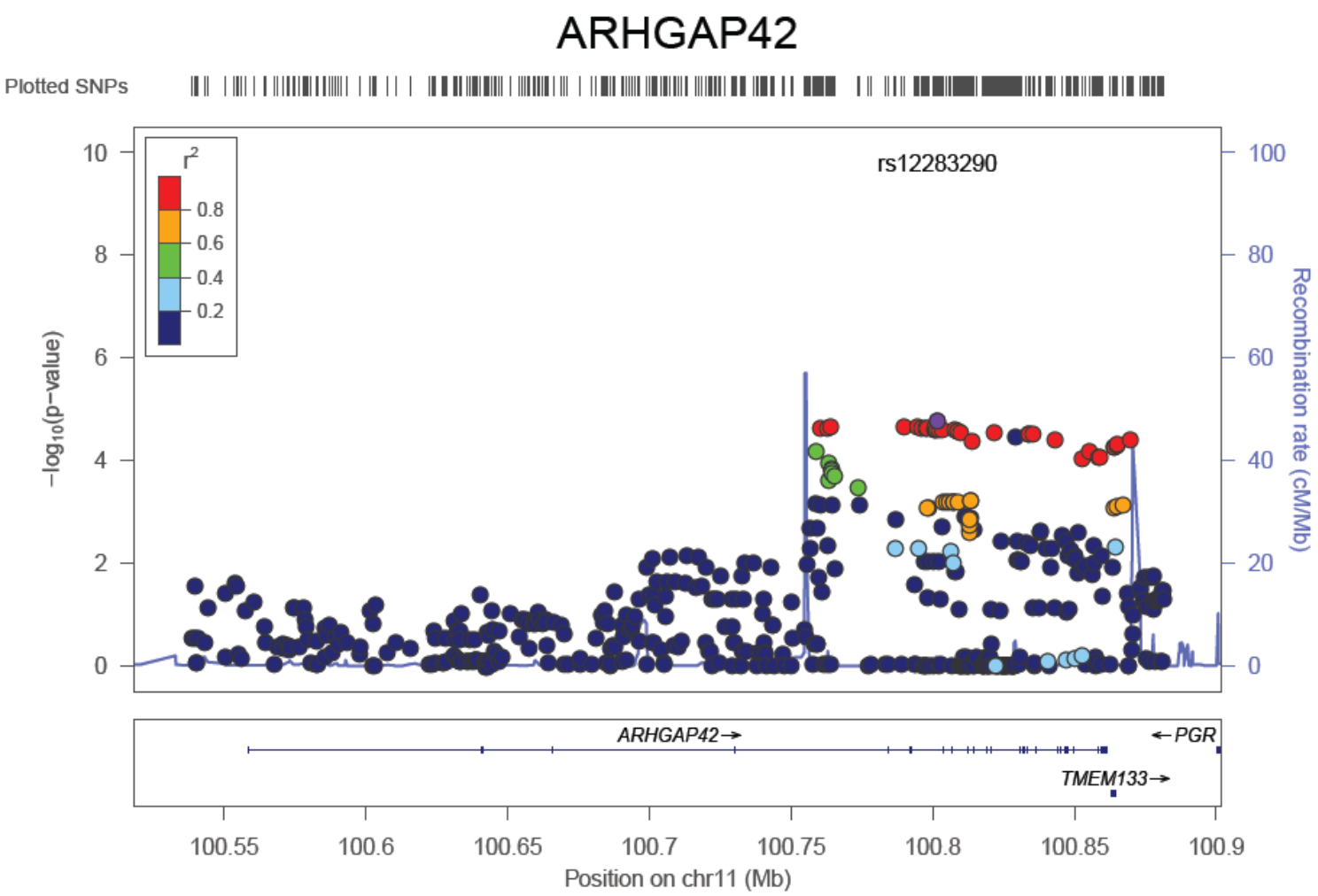
*Purple circle is the SNP with lowest p-value.

Appendix 3.12. SNPs and their *cis*-effects on *CSK* expression.



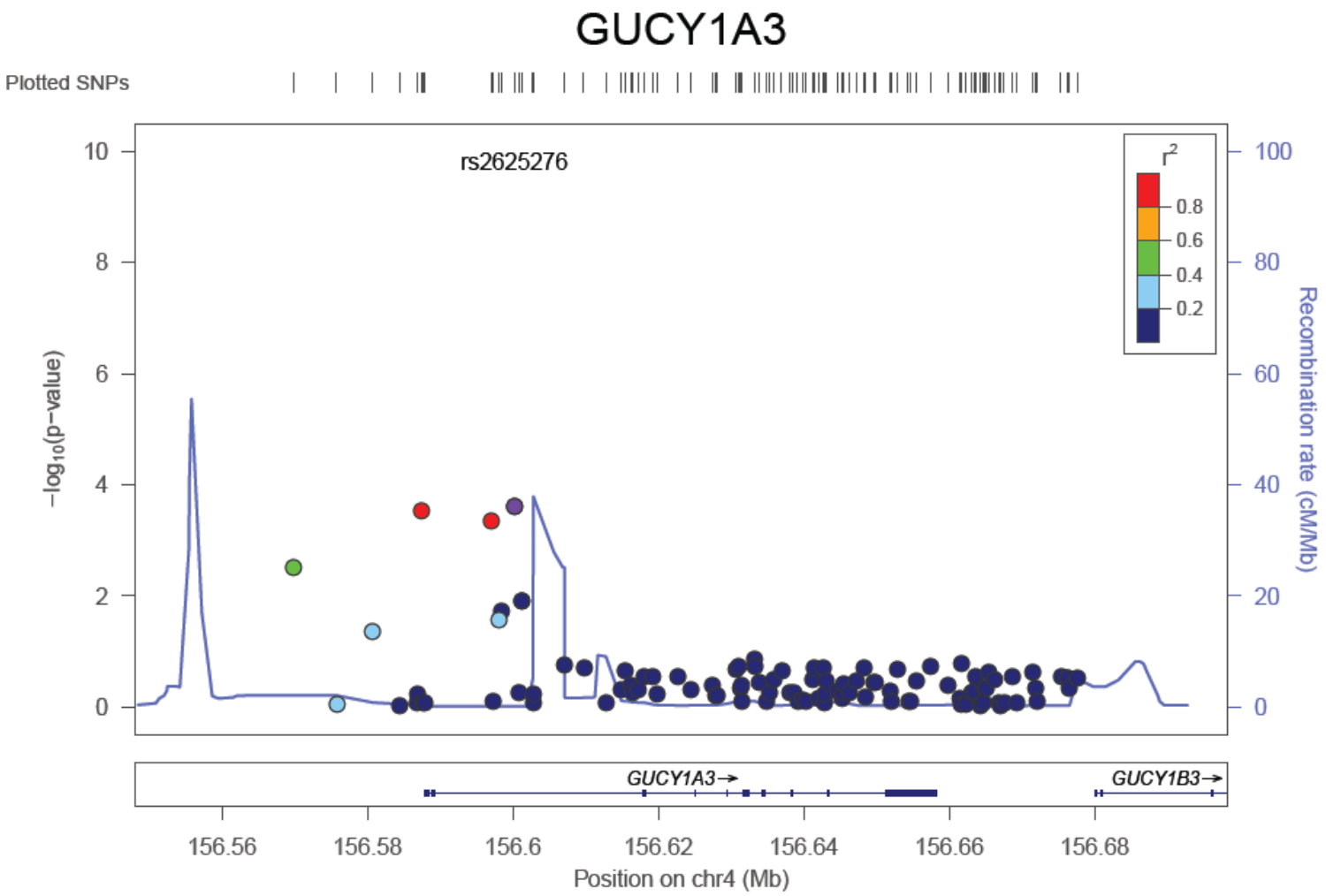
*Purple circle is the SNP with lowest p-value.

Appendix 3.13. SNPs and their *cis*-effects on *ARHGAP42* expression.



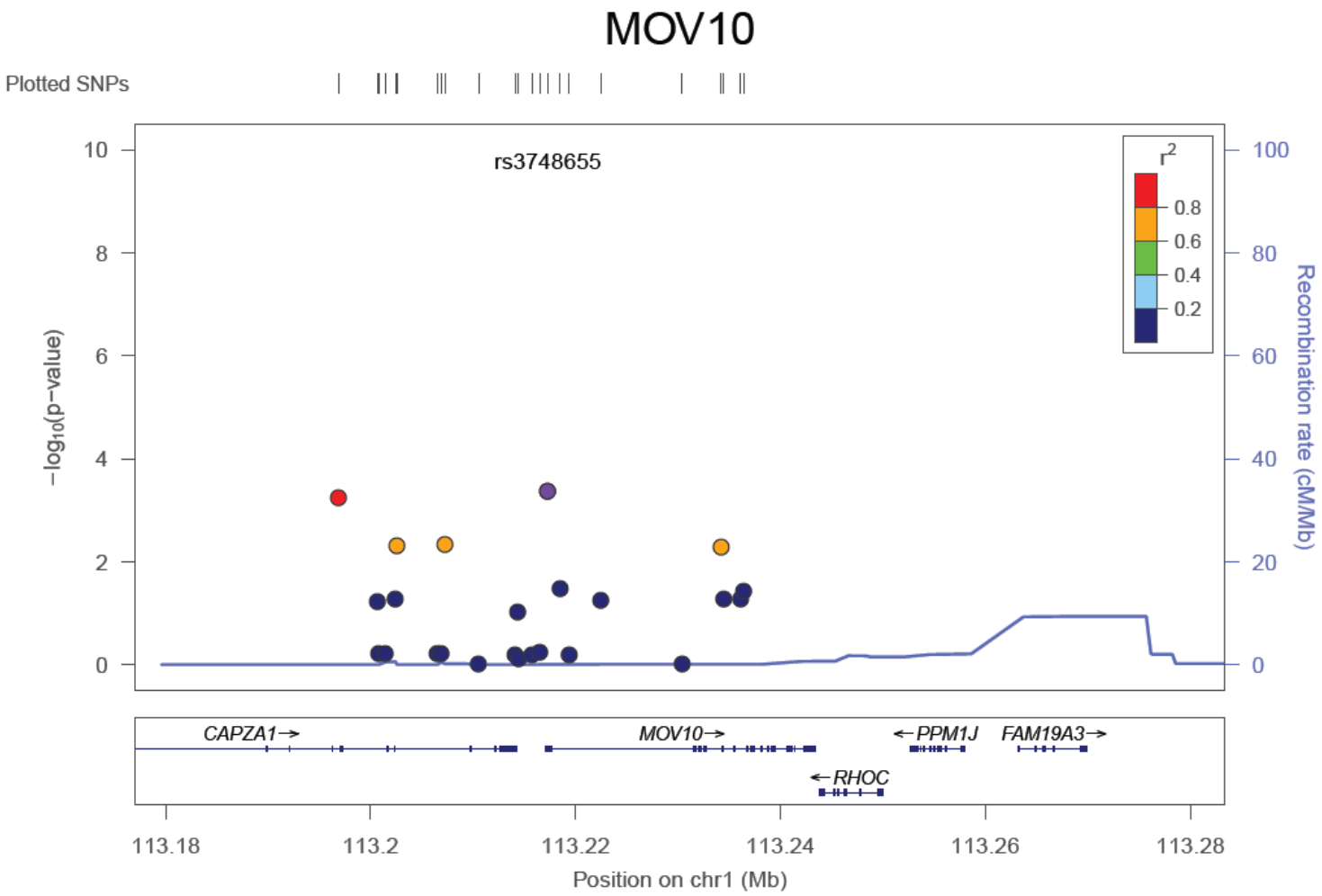
*Purple circle is the SNP with lowest p-value.

Appendix 3.14. SNPs and their *cis*-effects on *GUCY1A3* expression.



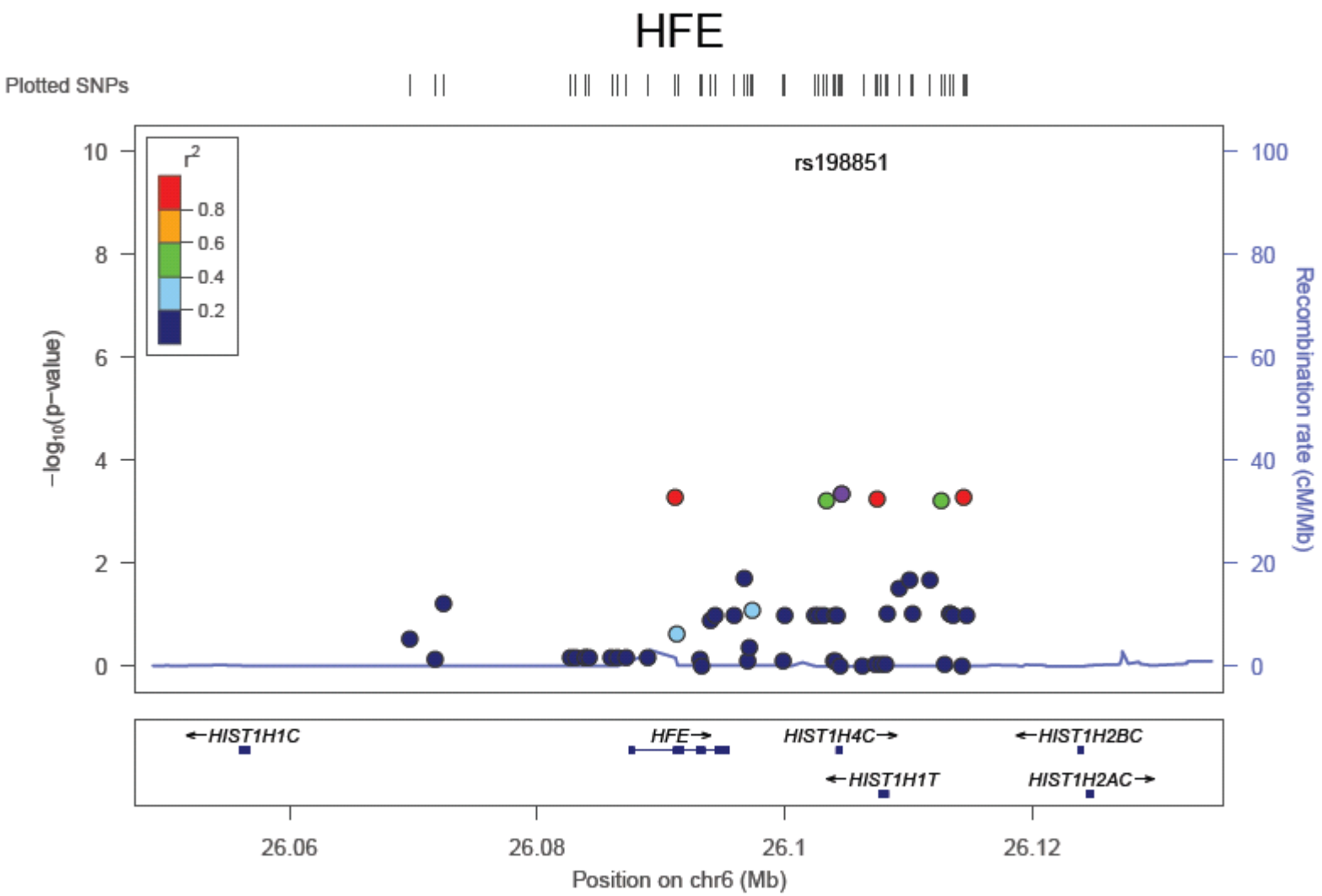
*Purple circle is the SNP with lowest p-value.

Appendix 3.15. SNPs and their *cis*-effects on *MOV10* expression.



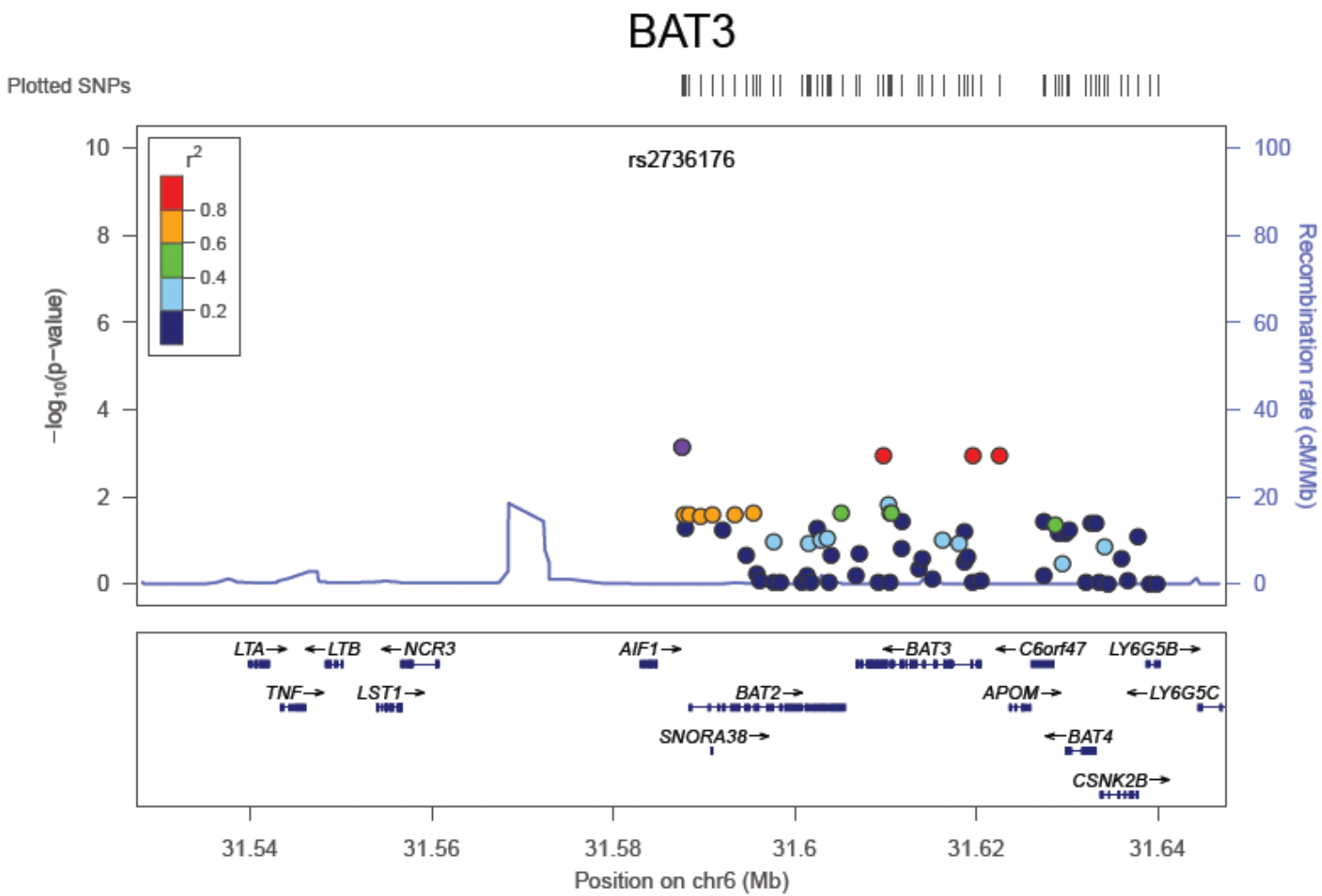
*Purple circle is the SNP with lowest p-value.

Appendix 3.16. SNPs and their *cis*-effects on *HFE* expression at gene level.



*Purple circle is the SNP with lowest p-value.

Appendix 3.17. SNPs and their *cis*-effects on *BAT3* expression.



*Purple circle is the SNP with lowest p-value.

Appendix 3.18. The 35 SNPs associated with diastolic blood pressure and *ULK4* expression.

SNP	Position	Function	Coded / Non-Coded Allele	Coded Allele Freq.	Systolic BP		Diastolic BP		Gene Level	
					Beta	P-value	Beta	P-value	Beta	P-value
rs17218264	41975847	Intron	C/T	0.17	1.18	0.302	1.11	0.027	0.32	**1.05E-54**
rs11129935	41985203	Intron	A/C	0.83	-1.18	0.301	-1.11	0.027	-0.32	**1.10E-54**
rs17284313	41986787	Intron	A/C	0.83	-1.20	0.296	-1.11	0.026	-0.32	**1.30E-54**
rs17218441	41987044	Intron	A/G	0.83	-1.18	0.304	-1.11	0.026	-0.32	**1.44E-54**
rs10510733	42009192	NA	C/T	0.17	1.07	0.349	1.09	0.030	0.32	**1.54E-54**
rs17218504	41987278	Intron	A/G	0.83	-1.15	0.314	-1.11	0.027	-0.32	**1.71E-54**
rs10510732	42007339	NA	A/G	0.17	1.04	0.366	1.08	0.031	0.32	**1.95E-54**
rs4299460	41952445	Intron	A/C	0.83	-1.14	0.320	-1.09	0.029	-0.32	**3.02E-54**
rs17284472	41996304	Intron	A/G	0.17	1.08	0.349	1.10	0.028	0.32	**4.11E-54**
rs12185934	41836558	Intron	A/C	0.83	-0.98	0.393	-1.04	0.038	-0.32	**1.37E-53**
rs12186050	41836879	Intron	C/G	0.17	0.98	0.393	1.03	0.038	0.32	**1.37E-53**
rs766286	41838329	Intron	A/G	0.17	0.98	0.393	1.03	0.039	0.32	**1.39E-53**
rs11129932	41838767	Intron	A/C	0.17	0.98	0.393	1.03	0.039	0.32	**1.41E-53**
rs10510729	41831998	Intron	A/G	0.17	0.98	0.394	1.04	0.037	0.32	**1.42E-53**
rs17215883	41841873	Intron	C/T	0.83	-0.98	0.392	-1.03	0.039	-0.32	**1.46E-53**
rs17215589	41831203	Missense	C/T	0.83	-0.98	0.395	-1.05	0.036	-0.32	**1.47E-53**
rs10510730	41849013	Intron	C/G	0.83	-0.98	0.391	-1.03	0.039	-0.32	**1.48E-53**
rs17215988	41851320	Intron	A/C	0.83	-0.98	0.390	-1.03	0.040	-0.32	**1.57E-53**
rs17218103	41943724	Intron	A/G	0.83	-1.10	0.336	-1.07	0.032	-0.32	**1.60E-53**
rs10510731	41853310	Intron	C/T	0.83	-0.99	0.388	-1.02	0.040	-0.32	**1.71E-53**
rs17281609	41833432	Intron	A/G	0.18	0.92	0.419	1.01	0.041	0.32	**2.48E-53**
rs17283677	41926148	Intron	A/G	0.84	-1.09	0.341	-1.07	0.033	-0.32	**2.74E-53**
rs17216919	41888938	Intron	C/G	0.17	1.09	0.341	1.07	0.033	0.32	**2.74E-53**
rs17283929	41941898	Intron	C/G	0.17	1.09	0.341	1.07	0.033	0.32	**2.74E-53**
rs17283243	41914792	Intron	C/T	0.84	-1.09	0.341	-1.07	0.033	-0.32	**2.75E-53**
rs17215498	41829679	Intron	C/G	0.18	0.92	0.420	1.02	0.039	0.32	**2.88E-53**
rs17281182	41817263	Intron	C/T	0.83	-1.00	0.383	-1.07	0.032	-0.32	**3.45E-53**
rs17215183	41812334	Intron	C/T	0.17	1.01	0.379	1.08	0.031	0.32	**4.28E-53**
rs17280980	41810949	Intron	A/T	0.84	-1.03	0.370	-1.09	0.029	-0.32	**7.67E-53**
rs17214945	41808660	Intron	G/T	0.16	1.01	0.377	1.09	0.030	0.32	**8.74E-53**
rs17280952	41810834	Intron	C/T	0.16	1.03	0.368	1.09	0.029	0.32	**8.84E-53**
rs17214987	41810481	Intron	A/G	0.84	-1.04	0.366	-1.10	0.029	-0.32	**9.86E-53**
rs17280777	41806167	Intron	C/T	0.84	-0.92	0.430	-1.09	0.033	-0.32	**4.31E-52**
rs17218866	42017634	Intron	C/T	0.15	1.22	0.304	1.13	0.029	0.31	**2.80E-46**
rs17218936	42019392	NA	C/T	0.85	-1.21	0.307	-1.12	0.030	-0.31	**3.06E-46**
SBP	NA	NA	NA	NA	NA	NA	NA	NA	4.64	**0.0148**
DBP	NA	NA	NA	NA	NA	NA	NA	NA	4.03	**0.0004**

Appendix 3.19. The 28 SNPs associated with systolic blood pressure and *ULK4* expression.

SNP	Position	Function	Coded / Non-Coded Allele	Coded Allele Freq.	Systolic BP		Diastolic BP		Gene Level	
					Beta	P-value	Beta	P-value	Beta	P-value
rs9311289	41827063	Intron	A/G	0.33	-2.08	0.032	-1.26	0.077	-0.10	**0.0000**
rs1630957	41559534	Intron	A/G	0.07	-3.67	0.026	-1.06	0.377	-0.13	**0.0007**
rs1256354	41628177	Intron	G/T	0.94	4.23	0.010	1.09	0.360	0.12	**0.0010**
rs1256355	41627589	Intron	A/G	0.06	-4.23	0.010	-1.09	0.356	-0.12	**0.0011**
rs2576194	41624012	Intron	C/T	0.94	4.29	0.009	1.13	0.337	0.12	**0.0013**
rs2700465	41622938	Intron	A/C	0.94	4.30	0.008	1.14	0.332	0.12	**0.0013**
rs17057270	41477890	Intron	C/T	0.06	-4.33	0.008	-1.48	0.208	-0.12	**0.0017**
rs1691986	41618420	Intron	A/C	0.07	-4.19	0.009	-1.02	0.379	-0.11	**0.0020**
rs1256396	41551051	Intron	A/T	0.06	-4.85	0.002	-1.40	0.218	-0.10	**0.0037**
rs1691983	41542342	Intron	A/G	0.94	4.85	0.002	1.40	0.218	0.10	**0.0037**
rs2643976	41541841	Intron	A/G	0.06	4.85	0.002	1.40	0.217	0.10	**0.0037**
rs2643977	41541987	Intron	C/T	0.06	-4.85	0.002	-1.40	0.217	-0.10	**0.0037**
rs4258937	41544423	Intron	A/G	0.94	-4.85	0.002	-1.40	0.217	-0.10	**0.0037**
rs9831248	41549933	Intron	C/G	0.06	-4.85	0.002	-1.40	0.217	-0.10	**0.0037**
rs11919885	41541456	Intron	C/T	0.94	-4.85	0.002	-1.40	0.217	-0.10	**0.0038**
rs12639280	41540052	Intron	C/T	0.06	4.86	0.002	1.40	0.216	0.10	**0.0038**
rs1495700	41532869	Intron	A/C	0.94	-4.86	0.002	-1.40	0.215	-0.10	**0.0038**
rs2017454	41555563	Intron	A/G	0.06	-4.86	0.002	-1.41	0.215	-0.10	**0.0038**
rs2643968	41538775	Intron	C/G	0.06	4.87	0.002	1.41	0.212	0.10	**0.0038**
rs1795317	41531540	Intron	C/T	0.06	-4.88	0.002	-1.42	0.211	-0.10	**0.0039**
rs7613493	41530347	Intron	C/T	0.06	-4.89	0.002	-1.43	0.209	-0.10	**0.0039**
rs1495693	41560092	Intron	C/T	0.06	-4.87	0.002	-1.45	0.208	-0.10	**0.0041**
rs1631168	41559614	Intron	A/C	0.06	-4.87	0.002	-1.45	0.208	-0.10	**0.0041**
rs1795312	41560382	Intron	A/T	0.06	-4.87	0.002	-1.45	0.207	-0.10	**0.0041**
rs1565960	41567692	Intron	A/G	0.06	-4.56	0.006	-1.35	0.256	-0.11	**0.0044**
rs1629284	41586308	Intron	C/G	0.06	-4.86	0.003	-1.48	0.203	-0.10	**0.0044**
rs1874354	41557727	Intron	A/T	0.07	-5.25	0.002	-1.32	0.276	-0.11	**0.0046**
rs1795346	41609466	Intron	C/G	0.06	-4.65	0.004	-1.42	0.219	-0.10	**0.0048**
SBP	NA	NA	NA	NA	NA	NA	NA	NA	4.64	**0.0148**
DBP	NA	NA	NA	NA	NA	NA	NA	NA	4.03	**0.0004**

Appendix 3.20. The 7 SNPs associated with blood pressure and *MTHFR* expression.

SNP	Position	Function	Coded / Non-Coded Allele	Coded Allele Freq.	Systolic BP		Diastolic BP		Gene Level	
					Beta	P-value	Beta	P-value	Beta	P-value
rs17037425	11870383	Intron	A/G	0.13	-2.631	**0.0108**	-1.282	**0.0413**	0.104	**3.31E-12**
rs13306561	11865804	Intron	A/G	0.84	2.860	**0.0041**	1.471	**0.0150**	-0.099	**5.17 E-12**
rs17037390	11860843	Intron	A/G	0.14	-3.049	**0.0021**	-1.537	**0.0106**	0.097	**1.40 E-11**
rs2050265	11879699	Intron	A/G	0.86	3.029	**0.0021**	1.629	**0.0065**	-0.096	**1.54 E-11**
rs12567136	11883731	Intron	C/T	0.86	3.053	**0.0019**	1.657	**0.0057**	-0.096	**1.69 E-11**
rs17367504*	11862778	Intron	A/G	0.86	2.938	**0.0028**	1.499	**0.0122**	-0.095	**2.84 E-11**
rs3737967	11847449	Missense C1orf167	A/G	0.05	-3.284	**0.0395**	-1.976	**0.0416**	0.142	**1.79 E-09**
Systolic BP	NA	Intron	NA	NA	NA	NA	NA	NA	-1.030	0.74
Diastolic BP	NA	Intron	NA	NA	NA	NA	NA	NA	-1.570	0.41

*rs17367504 is the index SNP reported from ICBP.

Appendix 3.21. The 23 SNPs associated with blood pressure and *PLEKHA7* expression.

SNP	Position	Function	Coded / Non-Coded Allele	Coded Allele Freq.	Systolic BP		Diastolic BP		Gene Level	
					Beta	P-value	Beta	P-value	Beta	P-value
rs12708338	16919790	Intron	G/T	0.14	1.83	0.0618	1.38	**0.0213**	-0.126	**5.21E-07**
rs7123174	16911431	Intron	A/T	0.87	-1.79	0.0757	-1.43	**0.0193**	0.128	**6.79E-07**
rs11024074	16917219	Intron	C/T	0.27	1.39	0.0725	1.04	**0.0261**	-0.095	**3.04E-06**
rs10832694	16881656	Intron	C/T	0.59	0.98	0.1625	1.05	**0.0139**	-0.065	**0.0005**
rs370161	16894877	Intron	A/C	0.48	-1.15	0.1011	-0.86	**0.0437**	0.064	**0.0005**
rs442389	16894413	Intron	A/G	0.52	1.16	0.0992	0.86	**0.0434**	-0.064	**0.0005**
rs2041236	16922349	Intron	A/T	0.49	-1.10	0.1115	-0.83	**0.0477**	0.063	**0.0005**
rs10766356	16918796	Intron	A/G	0.51	1.10	0.1119	0.83	**0.0475**	-0.063	**0.0005**
rs10832695	16916347	Intron	A/G	0.49	-1.10	0.1123	-0.84	**0.0472**	0.063	**0.0005**
rs10766354	16913292	Intron	C/T	0.51	1.10	0.1128	0.84	**0.0472**	-0.063	**0.0005**
rs378755	16912251	Intron	A/C	0.49	-1.10	0.1131	-0.84	**0.0472**	0.063	**0.0005**
rs432373	16907844	Intron	C/T	0.51	1.10	0.1137	0.84	**0.0471**	-0.063	**0.0005**
rs402973	16892477	Intron	C/G	0.47	-1.17	0.0927	-0.86	**0.0436**	0.063	**0.0006**
rs11024066	16897496	Intron	C/T	0.62	1.21	0.0926	1.18	**0.0070**	-0.063	**0.0008**
rs7109209	16924324	Intron	A/G	0.39	-1.22	0.0867	-1.17	**0.0067**	0.060	**0.0012**
rs1027617	16842787	Intron	A/G	0.54	1.79	0.0744	1.37	**0.0248**	0.082	**0.0024**
rs20818	16925684	Intron	A/G	0.47	-1.38	**0.0457**	-0.89	**0.0332**	0.054	**0.0029**
rs10766358	16934149	Intron	C/G	0.37	-1.14	0.1148	-0.93	**0.0354**	0.056	**0.0030**
rs7117745	16938792	Intron	C/T	0.55	1.48	**0.0350**	0.72	0.0933	-0.054	**0.0031**
rs11024076	16938450	Intron	A/G	0.38	-1.41	0.0502	-1.09	**0.0130**	0.056	**0.0032**
rs2215085	16944512	Intron	C/T	0.45	-1.54	**0.0277**	-0.71	0.0956	0.054	**0.0032**
rs4757444	16937968	Intron	A/G	0.45	-1.43	**0.0418**	-0.71	0.0964	0.054	**0.0032**
rs4757446	16948199	Intron	C/T	0.55	1.43	**0.0403**	0.65	0.1233	-0.052	**0.0042**
Systolic BP	NA	Intron	NA	NA	NA	NA	NA	NA	-2.72	0.1267
Diastolic BP	NA	Intron	NA	NA	NA	NA	NA	NA	-0.97	0.3672

Appendix 3.22. The twenty-One SNPs that were associated with both systolic blood pressure and *ULK4* expression in both sample.

SNP	Position	Function	Coded / Non-Coded Allele	Sample with SNP Data (N=1423)			Sample with both SNP and Gene Expreession Data (N=789)				
				Systolic BP			Systolic BP			Gene Expression	
				Coded Allele Freq.	Beta	P-value	Coded Allele Freq.	Beta	P-value	Beta	P-value
rs17057270	41477890	Intron	C/T	0.06	-4.33	0.008	0.07	-5.10	0.009	-0.12	0.0017
rs1256396	41551051	Intron	A/T	0.06	-4.85	0.002	0.07	-5.71	0.002	-0.10	0.0037
rs1691983	41542342	Intron	A/G	0.94	4.85	0.002	0.93	5.71	0.002	0.10	0.0037
rs2643976	41541841	Intron	A/G	0.06	4.85	0.002	0.07	5.71	0.002	0.10	0.0037
rs2643977	41541987	Intron	C/T	0.06	-4.85	0.002	0.07	-5.71	0.002	-0.10	0.0037
rs4258937	41544423	Intron	A/G	0.94	-4.85	0.002	0.93	-5.71	0.002	-0.10	0.0037
rs9831248	41549933	Intron	C/G	0.06	-4.85	0.002	0.07	-5.71	0.002	-0.10	0.0037
rs11919885	41541456	Intron	C/T	0.94	-4.85	0.002	0.93	-5.71	0.002	-0.10	0.0038
rs12639280	41540052	Intron	C/T	0.06	4.86	0.002	0.07	5.72	0.002	0.10	0.0038
rs1495700	41532869	Intron	A/C	0.94	-4.86	0.002	0.93	-5.72	0.002	-0.10	0.0038
rs2017454	41555563	Intron	A/G	0.06	-4.86	0.002	0.07	-5.72	0.002	-0.10	0.0038
rs2643968	41538775	Intron	C/G	0.06	4.87	0.002	0.07	5.73	0.002	0.10	0.0038
rs1795317	41531540	Intron	C/T	0.06	-4.88	0.002	0.07	-5.74	0.002	-0.10	0.0039
rs7613493	41530347	Intron	C/T	0.06	-4.89	0.002	0.07	-5.75	0.002	-0.10	0.0039
rs1495693	41560092	Intron	C/T	0.06	-4.87	0.002	0.07	-5.73	0.003	-0.10	0.0041
rs1631168	41559614	Intron	A/C	0.06	-4.87	0.002	0.07	-5.73	0.003	-0.10	0.0041
rs1795312	41560382	Intron	A/T	0.06	-4.87	0.002	0.07	-5.73	0.003	-0.10	0.0041
rs1565960	41567692	Intron	A/G	0.06	-4.56	0.006	0.07	-5.37	0.007	-0.11	0.0044
rs1629284	41586308	Intron	C/G	0.06	-4.86	0.003	0.06	-5.72	0.003	-0.10	0.0044
rs1874354	41557727	Intron	A/T	0.07	-5.25	0.002	0.08	-6.18	0.002	-0.11	0.0046
rs1795346	41609466	Intron	C/G	0.06	-4.65	0.004	0.06	-5.47	0.004	-0.10	0.0048

Appendix 3.23. The twenty-two SNPs that were associated with both diastolic blood pressure and *ULK4* expression in both sample.

SNP	Position	Function	Coded / Non-Coded Allele	Sample with SNP Data (N=1423)			Sample with both SNP and Gene Expreession Data (N=789)				
				Diastolic Blood Pressure			Diastolic Blood Pressure			Gene Expression	
				Coded Allele Freq.	Beta	P-value	Coded Allele Freq.	Beta	P-value	Beta	P-value
rs17218264	41975847	Intron	C/T	0.17	1.11	0.027	0.18	1.47	0.032	0.32	1.05E-54
rs11129935	41985203	Intron	A/C	0.83	-1.11	0.027	0.82	-1.48	0.032	-0.32	1.10E-54
rs17284313	41986787	Intron	A/C	0.83	-1.11	0.026	0.82	-1.49	0.031	-0.32	1.30E-54
rs17218441	41987044	Intron	A/G	0.83	-1.11	0.026	0.82	-1.48	0.031	-0.32	1.44E-54
rs10510733	42009192	Intron	C/T	0.17	1.09	0.030	0.18	1.45	0.036	0.32	1.54E-54
rs17218504	41987278	Intron	A/G	0.83	-1.11	0.027	0.82	-1.48	0.032	-0.32	1.71E-54
rs4299460	41952445	Intron	A/C	0.83	-1.09	0.029	0.82	-1.45	0.035	-0.32	3.02E-54
rs17284472	41996304	Intron	A/G	0.17	1.10	0.028	0.18	1.47	0.034	0.32	4.11E-54
rs17218103	41943724	Intron	A/G	0.83	-1.07	0.032	0.82	-1.43	0.038	-0.32	1.60E-53
rs17283677	41926148	Intron	A/G	0.84	-1.07	0.033	0.83	-1.42	0.039	-0.32	2.74E-53
rs17216919	41888938	Intron	C/G	0.17	1.07	0.033	0.18	1.42	0.039	0.32	2.74E-53
rs17283929	41941898	Intron	C/G	0.17	1.07	0.033	0.18	1.42	0.039	0.32	2.74E-53
rs17283243	41914792	Intron	C/T	0.84	-1.07	0.033	0.83	-1.42	0.039	-0.32	2.75E-53
rs17281182	41817263	Intron	C/T	0.83	-1.07	0.032	0.82	-1.43	0.038	-0.32	3.45E-53
rs17215183	41812334	Intron	C/T	0.17	1.08	0.031	0.18	1.44	0.037	0.32	4.28E-53
rs17280980	41810949	Intron	A/T	0.84	-1.09	0.029	0.83	-1.46	0.035	-0.32	7.67E-53
rs17214945	41808660	Intron	G/T	0.16	1.09	0.030	0.17	1.45	0.036	0.32	8.74E-53
rs17280952	41810834	Intron	C/T	0.16	1.09	0.029	0.17	1.46	0.034	0.32	8.84E-53
rs17214987	41810481	Intron	A/G	0.84	-1.10	0.029	0.83	-1.46	0.034	-0.32	9.86E-53
rs17280777	41806167	Intron	C/T	0.84	-1.09	0.033	0.83	-1.45	0.039	-0.32	4.31E-52
rs17218866	42017634	Intron	C/T	0.15	1.13	0.029	0.16	1.50	0.035	0.31	2.80E-46
rs17218936	42019392	Intron	C/T	0.85	-1.12	0.030	0.84	-1.50	0.036	-0.31	3.06E-46

Appendix 3.24. Multivariable linear mixed model analysis of SNPs and gene expression levels on systolic blood pressure.

SNP	Position	Function	Coded / Non-Coded Allele	Coded Allele Freq.	Univariable Analysis				Multivariable Analysis			
					Systolic BP		Gene Expression		Systolic BP		Gene Expression	
					Beta	P-value	Beta	P-value	Beta	P-value	Beta	P-value
rs17057270	41477890	Intron	C/T	0.07	-5.10	0.009	-0.12	0.0017	-3.30	0.4063	5.15	0.0071
rs1256396	41551051	Intron	A/T	0.07	-5.71	0.002	-0.10	0.0037	-3.64	0.1032	5.20	0.0065
rs1691983	41542342	Intron	A/G	0.93	5.71	0.002	0.10	0.0037	3.64	0.1023	5.20	0.0065
rs2643976	41541841	Intron	A/G	0.07	5.71	0.002	0.10	0.0037	-3.64	0.1018	5.20	0.0065
rs2643977	41541987	Intron	C/T	0.07	-5.71	0.002	-0.10	0.0037	-3.64	0.1022	5.20	0.0065
rs4258937	41544423	Intron	A/G	0.93	-5.71	0.002	-0.10	0.0037	3.64	0.1024	5.20	0.0065
rs9831248	41549933	Intron	C/G	0.07	-5.71	0.002	-0.10	0.0037	-3.64	0.1025	5.20	0.0065
rs11919885	41541456	Intron	C/T	0.93	-5.71	0.002	-0.10	0.0038	3.64	0.101	5.20	0.0065
rs12639280	41540052	Intron	C/T	0.07	5.72	0.002	0.10	0.0038	-3.64	0.1007	5.20	0.0065
rs1495700	41532869	Intron	A/C	0.93	-5.72	0.002	-0.10	0.0038	3.64	0.0982	5.20	0.0065
rs2017454	41555563	Intron	A/G	0.07	-5.72	0.002	-0.10	0.0038	-3.64	0.1038	5.20	0.0065
rs2643968	41538775	Intron	C/G	0.07	5.73	0.002	0.10	0.0038	-3.64	0.1003	5.20	0.0065
rs1795317	41531540	Intron	C/T	0.07	-5.74	0.002	-0.10	0.0039	-3.64	0.0972	5.20	0.0065
rs7613493	41530347	Intron	C/T	0.07	-5.75	0.002	-0.10	0.0039	-3.67	0.0962	5.20	0.0065
rs1495693	41560092	Intron	C/T	0.07	-5.73	0.003	-0.10	0.0041	-3.64	0.1172	5.19	0.0066
rs1631168	41559614	Intron	A/C	0.07	-5.73	0.003	-0.10	0.0041	-3.64	0.1156	5.19	0.0066
rs1795312	41560382	Intron	A/T	0.07	-5.73	0.003	-0.10	0.0041	-3.64	0.1181	5.19	0.0066
rs1565960	41567692	Intron	A/G	0.07	-5.37	0.007	-0.11	0.0044	-3.64	0.1299	5.18	0.0067
rs1629284	41586308	Intron	C/G	0.06	-5.72	0.003	-0.10	0.0044	-3.44	0.3098	5.13	0.0073
rs1874354	41557727	Intron	A/T	0.08	-6.18	0.002	-0.11	0.0046	-3.92	0.0943	5.18	0.0066
rs1795346	41609466	Intron	C/G	0.06	-5.47	0.004	-0.10	0.0048	-3.47	0.2042	5.15	0.0071

Appendix 3.25. Multivariable linear mixed model analysis of SNPs and gene expression levels on diastolic blood pressure.

SNP	Position	Function	Coded / Non-Coded Allele	Coded Allele Freq.	Univariable Analysis				Multivariable Analysis			
					Diastolic BP		Gene Expression		Diastolic BP		Gene Expression	
					Beta	P-value	Beta	P-value	Beta	P-value	Beta	P-value
rs17218264	41975847	Intron	C/T	0.18	1.47	0.032	0.32	1.05E-54	0.28	0.730	3.75	0.006
rs11129935	41985203	Intron	A/C	0.82	-1.48	0.032	-0.32	1.10E-54	-0.28	0.728	3.75	0.006
rs17284313	41986787	Intron	A/C	0.82	-1.49	0.031	-0.32	1.30E-54	-0.30	0.714	3.73	0.006
rs17218441	41987044	Intron	A/G	0.82	-1.48	0.031	-0.32	1.44E-54	-0.29	0.718	3.74	0.006
rs10510733	42009192	Intron	C/T	0.18	1.45	0.036	0.32	1.54E-54	0.24	0.768	3.79	0.005
rs17218504	41987278	Intron	A/G	0.82	-1.48	0.032	-0.32	1.71E-54	-0.29	0.722	3.74	0.006
rs4299460	41952445	Intron	A/C	0.82	-1.45	0.035	-0.32	3.02E-54	-0.25	0.756	3.78	0.005
rs17284472	41996304	Intron	A/G	0.18	1.47	0.034	0.32	4.11E-54	0.27	0.736	3.76	0.005
rs17218103	41943724	Intron	A/G	0.82	-1.43	0.038	-0.32	1.60E-53	-0.23	0.777	3.80	0.005
rs17283677	41926148	Intron	A/G	0.83	-1.42	0.039	-0.32	2.74E-53	-0.22	0.783	3.80	0.005
rs17216919	41888938	Intron	C/G	0.18	1.42	0.039	0.32	2.74E-53	0.22	0.783	3.80	0.005
rs17283929	41941898	Intron	C/G	0.18	1.42	0.039	0.32	2.74E-53	0.22	0.783	3.80	0.005
rs17283243	41914792	Intron	C/T	0.83	-1.42	0.039	-0.32	2.75E-53	-0.22	0.783	3.80	0.005
rs17281182	41817263	Intron	C/T	0.82	-1.43	0.038	-0.32	3.45E-53	-0.23	0.771	3.79	0.005
rs17215183	41812334	Intron	C/T	0.18	1.44	0.037	0.32	4.28E-53	0.25	0.759	3.78	0.005
rs17280980	41810949	Intron	A/T	0.83	-1.46	0.035	-0.32	7.67E-53	-0.28	0.731	3.76	0.005
rs17214945	41808660	Intron	G/T	0.17	1.45	0.036	0.32	8.74E-53	0.27	0.741	3.77	0.005
rs17280952	41810834	Intron	C/T	0.17	1.46	0.034	0.32	8.84E-53	0.28	0.725	3.75	0.005
rs17214987	41810481	Intron	A/G	0.83	-1.46	0.034	-0.32	9.86E-53	-0.29	0.720	3.75	0.005
rs17280777	41806167	Intron	C/T	0.83	-1.45	0.039	-0.32	4.31E-52	-0.24	0.767	3.79	0.005
rs17218866	42017634	Intron	C/T	0.16	1.50	0.035	0.31	2.80E-46	0.37	0.652	3.70	0.005
rs17218936	42019392	Intron	C/T	0.84	-1.50	0.036	-0.31	3.06E-46	-0.36	0.659	3.71	0.005

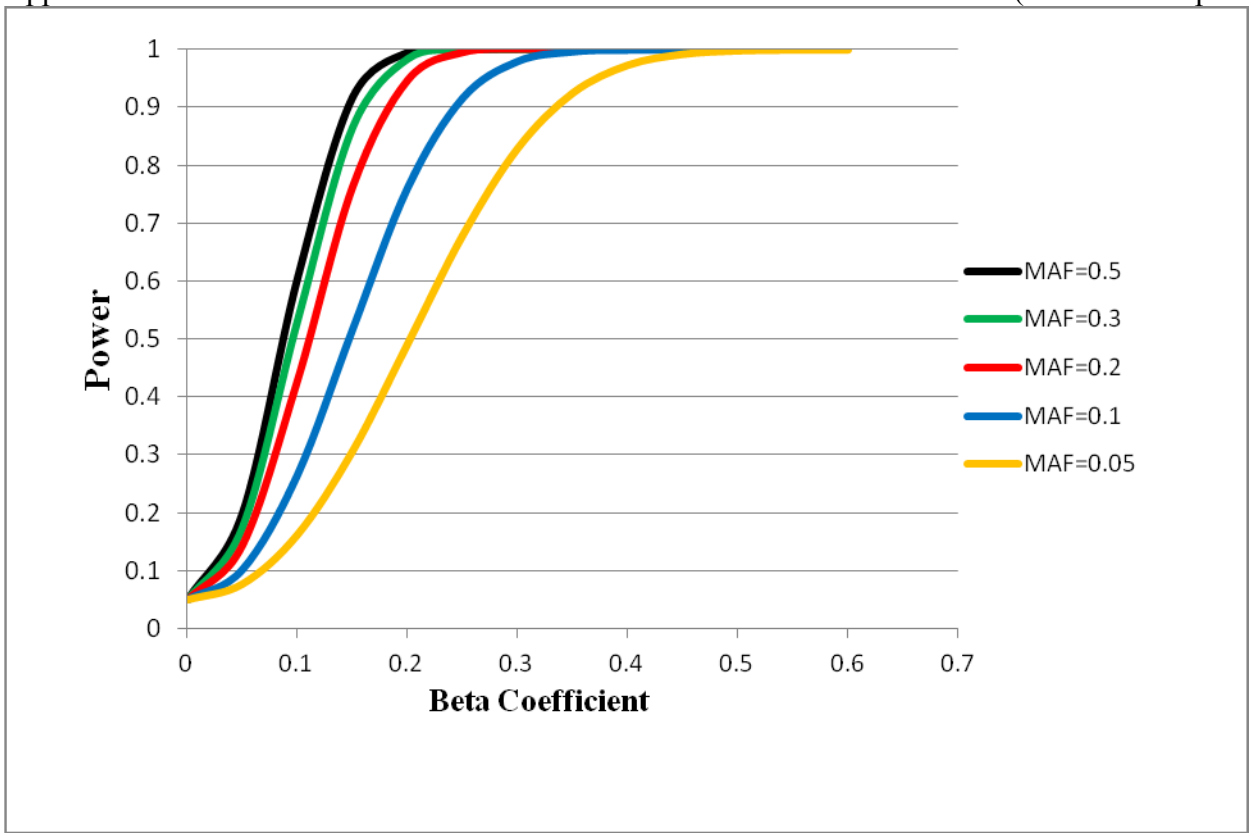
Appendix 3.26. Estimation of effective sample size for individuals with SNP data (Total sample size=1489, effective sample size=989).

Sibship Size	**Number of unrelated individuals corresponds to sibship size**	**Number of Sibships**	**Number of Unrelated individuals**
1	1	62	62
2	1.5	291	436.5
3	2	115	230
4	2.5	45	112.5
5	3	21	63
6	3.5	10	35
7	4	5	20
8	4.5	3	13.5
9	5	1	5
10	5.5	1	5.5
11	6	1	6
Total effective sample size in all sibships			989

Appendix 3.27. Estimation of effective sample size for individuals with both SNP data and gene expression data. (Total sample size=789, effective sample size=590).

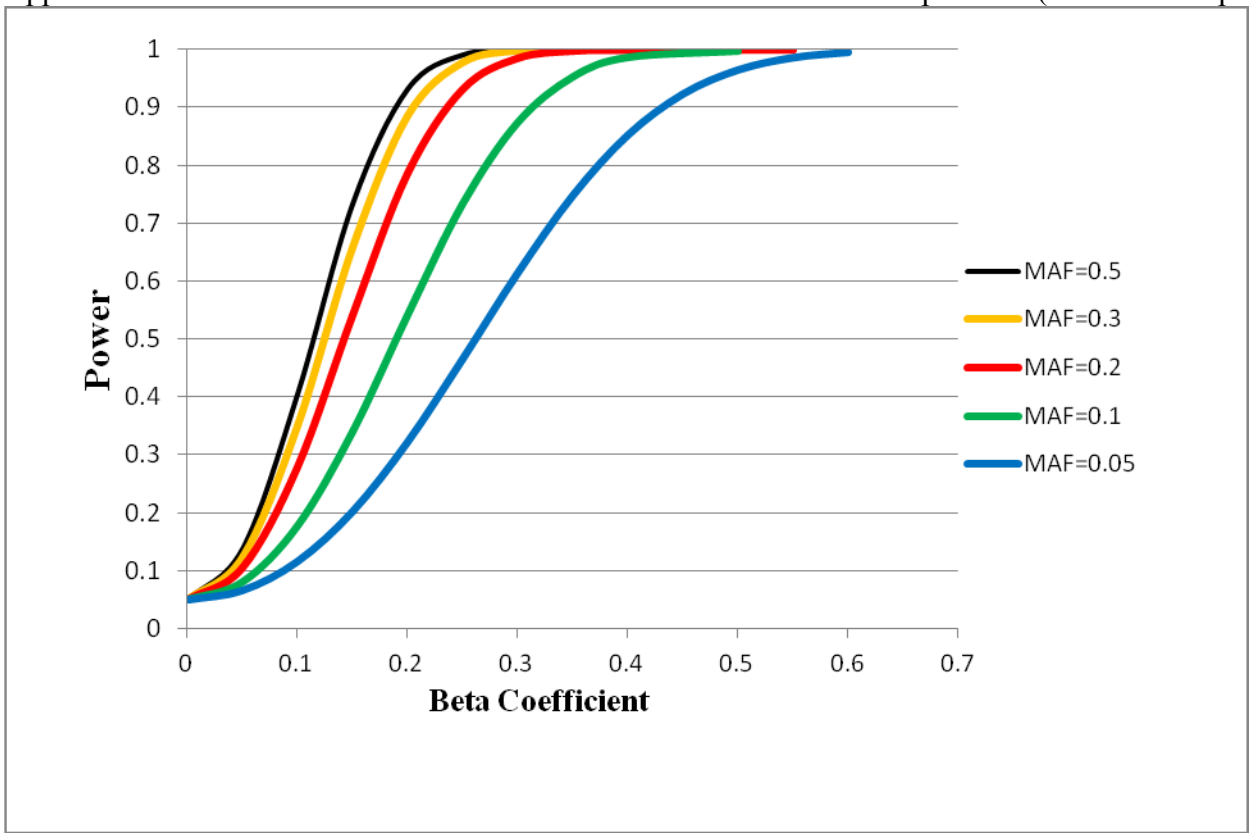
Sibship Size	Number of unrelated individuals corresponds to sibship size	Number of Sibships	Number of Unrelated individuals
1	1	131	131
2	1.5	181	271.5
3	2	47	94
4	2.5	16	40
5	3	10	30
6	3.5	4	14
7	4	0	0
8	4.5	1	4.5
9	5	1	5
Total effective sample size in all sibships			590

Appendix 3.28. Power calculation for association of SNPs and Blood Pressure (effective sample size=989).*



*Power calculation was estimated by using standardized phenotype with mean of zero and standard deviation of one.

Appendix 3.29. Power calculation for association of SNPs and Gene Expression (effective sample size=590).*



*Power calculation was estimated by using standardized phenotype with mean of zero and standard deviation of one.

Appendix 3.30.1. Potential functional for the 27 BP-related genes.

Gene	Chr	Gene Name	Function
CAPZA1	1	Capping protein (actin filament) muscle Z-line, alpha 1	This gene encodes the alpha subunit of the barbed-end actin binding protein. The protein regulates growth of the actin filament by capping the barbed end of growing actin filaments.[77]
MOV10	1	Moloney leukemia virus 10	This gene encodes putative helicase *MOV-10* enzyme.[77]
MTHFR	1	Methylenetetrahydrofolate reductase (NAD(P)H)	The protein encoded by this gene catalyzes the conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate. This reaction is required for the remethylation process that converts homocysteine to methionine. Genetic variation in this gene are associated with occlusive vascular disease, neural tube defects, colon cancer and acute leukemia, and mutations in this gene are associated with methylenetetrahydrofolate reductase deficiency.[77]
MECOM	3	MDS1 and EVI1 complex locus	The protein encoded by this gene is a transcriptional regulator that may be involved in hematopoiesis, apoptosis, development, and cell differentiation and proliferation.[77]
SLC4A7	3	Solute carrier family 4, sodium bicarbonate cotransporter, member 7	This gene encodes a sodium bicarbonate cotransporter. The encoded protein likely plays a critical role in regulation of intracellular pH involved in visual and auditory sensory transmission.[77]
ULK4	3	Unc-51-like kinase 4 (serine/threonine-protein kinase)	*ULK4* encodes a serine/threonine-protein kinase and this is an enzyme that transfers phosphates to serine or threonine in proteins.[77]
FGF5	4	Fibroblast growth factor 5	*FGF5* protein is a member of the fibroblast growth factor (*FGF*) family that stimulates cell growth and proliferation in multiple cell types, including cardiac myocytes, and has been associated with angiogenesis in the heart.[78]
GUCY1A3	4	Guanylate cyclase 1, soluble, alpha 3	The protein encoded by this gene is an alpha subunit of this complex and it interacts with a beta subunit to form the guanylate cyclase enzyme, which is activated by nitric oxide.[77]
PRDM8	4	PR domain containing 8	This gene encodes a member of the pantothenate kinase family. Pantothenate kinases are key regulatory enzymes in the biosynthesis of coenzyme A (CoA). The encoded protein catalyzes the first and rate-limiting enzymatic reaction in CoA biosynthesis and is regulated by CoA through feedback inhibition.[77]
SLC39A8	4	Solute carrier family 39 (zinc transporter), member 8	This gene encodes a member of the *SLC39* family of solute-carrier genes, which show structural characteristics of zinc transporters and it also transports cadmium and manganese.[79] The encoded protein is glycosylated and found in the plasma membrane and mitochondria, and functions in the cellular import of zinc at the onset of inflammation.

Appendix 3.30.2. Potential functional for the 27 BP-related genes.

Gene	Chr	Gene Name	Function
NPR3	5	Natriuretic peptide receptor C/guanylate cyclase C (atrionatriuretic peptide receptor C)	This gene encodes one of three natriuretic peptide receptors. Natriutetic peptides are small peptides which regulate blood volume and pressure, pulmonary hypertension, and cardiac function as well as some metabolic and growth processes.[77]
BAT3	6	BCL2-associated athanogene 6	This gene encodes a nuclear protein that is cleaved by caspase 3 and is implicated in the control of apoptosis. In addition, the protein forms a complex with E1A binding protein p300 and is required for the acetylation of p53 in response to DNA damage.[77]
HFE	6	Hemochromatosis	The protein encoded by this gene is a membrane protein that is similar to MHC class I-type proteins and associates with beta2-microglobulin (beta2M). It is thought that this protein functions to regulate iron absorption by regulating the interaction of the transferrin receptor with transferrin. The iron storage disorder, hereditary haemochromatosis, is a recessive genetic disorder that results from defects in this gene. [77]
CACNB2	10	Calcium channel, voltage-dependent, beta 2 subunit	This gene encodes a subunit of a voltage-dependent calcium channel protein which is a member of the voltage-gated calcium channel superfamily. The gene is expressed in the heart and a nonsynomymous variant in *CACNB2* was identified in affected individuals with Brugada syndrome. [80] *CACNB2* is one member of a family of voltage-gated calcium channel genes, several of which have effects on blood pressure regulation and serve as targets of calcium channel blockers. The beta-2 subunit interacts with alpha-1calcium channels (CaV1.2) and this is a mechanism by which variation in *CACNB2* may alter blood pressure.[81]
CNNM2	10	Cyclin M2	This gene encodes a member of the ancient conserved domain containing protein family. Members of this protein family contain a cyclin box motif and have structural similarity to the cyclins. The encoded protein may play an important role in magnesium homeostasis by mediating the epithelial transport and renal reabsorption of Mg2+. Mutations in this gene are associated with renal hypomagnesemia. Alternatively spliced transcript variants encoding multiple isoforms have been observed for this gene.[77]
NT5C2	10	5'-nucleotidase, cytosolic II	This gene encodes a hydrolase that serves as an important role in cellular purine metabolism by acting primarily on inosine 5'-monophosphate and other purine nucleotides.[77]
PLCE1	10	Phospholipase C, epsilon 1	*PLCE1* is important for normal podocyte development in the glomerulus; sequence variation in *PLCE1* has been implicated in familial nephrotic syndromes and end-stage kidney disease.[82]
ADM	11	Adrenomedullin	ADM encoding adrenomedullin—which has natriuretic, vasodilatory and blood-pressure-lowering properties.[19]

Appendix 3.30.3. Potential functional for the 27 BP-related genes.

Gene	Chr	Gene Name	Function
ARHGAP42	11	Rho GTPase activating protein 42	*ARHGAP* family members activate Rho GTPase proteins. Mutations of *ARHGAP* family genes lead to carcinogenesis through the dysregulation of Rho/Rac/Cdc42-like GTPases.[77]
PLEKHA7	11	Pleckstrin homology domain containing, family A member 7	*PLEKHA7* gene encodes pleckstrin homology domain- containing protein, family A member 7. This protein was reported to be involved in regulating integrity of zonula adherens, an epithelial cadherin-based cell-cell junction.[71]
ATP2B1	12	ATPase, Ca++ transporting, plasma membrane 1	This gene encodes plasma membrane calcium ATPase isoform 1 (PMCA1) that is expressed in vascular endothelium. It is involved in calcium pumping from the cytosol to the extracellular compartment.[83] An investigation of cultured rat aortic smooth muscle cells found elevated PMCA1 mRNA levels in spontaneously hypertensive rats compared to nonhypertensive controls.[84]
SH2B3	12	SH2B adaptor protein 3	This gene encodes a member of the *SH2B* adaptor family of proteins, which are involved in a range of signaling activities by growth factor and cytokine receptors. The encoded protein is a key negative regulator of cytokine signaling and plays a critical role in hematopoiesis. Mutations in this gene have been associated with susceptibility to celiac disease type 13 and susceptibility to insulin-dependent diabetes mellitus.[77]
CSK	15	C-src tyrosine kinase	*CSK* is involved in angiotensin II–dependent vascular smooth muscle cell proliferation.[85]
FES	15	Feline sarcoma oncogene	This gene encodes the human cellular counterpart of a feline sarcoma retrovirus protein with transforming capabilities. The gene product has tyrosine-specific protein kinase activity and that activity is required for maintenance of cellular transformation.[77]
GOSR2	17	Golgi SNAP receptor complex member 2	This gene encodes a trafficking membrane protein which transports proteins among the medial- and trans-Golgi compartments.[77]
PLCD3	17	Phospholipase C, delta 3	*PLCD3* is a member of the phospholipase C family of enzymes, important in vascular smooth muscle signaling and activated by the vasoactive peptides angiotensin II and endothelin.[86]
ZNF652	17	Zinc finger protein 652	*CBFA2T3-ZNF652* functions as a transcriptional corepressor complex.[77]

Chapter 4

Exploring the association among Genetic Variants, Gene Expression and BMI among non-Hispanic Whites in Genetic Epidemiology Network of Arteriopathy (GENOA)

Introduction

Obesity is a major public health problem that is associated with multiple chronic diseases, including hypertension, cardiovascular disease, and type II diabetes.[87]Approximately 32% and 34% of American adults are overweight and obese, respectively.[4] Although life style changes have shown to contribute to increased BMI, heritability studies provide evidence that genetic factors contribute to almost 25-60% of the population variation in BMI.[6,7,25]

Three large genome wide association study (GWAS) consortia of body mass index (BMI) have revealed 32 common genetic variants at 32 loci that contribute to BMI variation.[9,29,30] These associations are highly significant ($p<5x10^{-8}$) and replicate across multiple studies in non-Hispanic whites. The largest blood pressure GWAS consortium, the expanded Genetic Investigation of Anthropometric Traits (GIANT) consortium, includes a total of 249,796 individuals of European descent and identified associations between SNPs and BMI at18 novel loci.[9] They also replicated 2 known waist circumference loci [31,32] and 12 known BMI susceptibility loci reported from deCODE genetics and GIANT.[29,30] Together, the 32 confirmed BMI loci explained 1.45% of the interindividual variation in BMI. *FTO* gene accounted for the largest proportion of the variance (0.34%), followed by *TMEM18* (0.15%). Among these 32 risk

variants, all but one (rs13107325, *SLC39A8*) reside in non-coding regions and do not have any apparent functional effect (Appenidx 4.1). The current challenge is to elucidate how these risk variants influence the development of increased BMI. One promising approach is to evaluate how these variants and their proximal SNPs could regulate gene expression. This will allow us to better understand the potential functional mechanism at the molecular genetic level linking genetic variants and gene expression. Furthermore, examining whether the gene expression levels in these established genomic regions are associated with BMI could build our knowledge about the combined effects of multiple DNA variants that cannot be detected with a single SNP.

The central aim of this chapter is to investigate the functional genetic factors underlying body mass index variation in a three stage analysis (a) evaluating the associations between genetic variations in the significant and replicated genomic regions and body mass index, (b) investigating the association between these genetic variations and proximal gene expression, and (c) examining the relationship between gene expression and body mass index in the non-Hispanic whites of the Genetic Epidemiology Network of Arteriopathy (GENOA) (Figure 4.1).

Methods

Study Population

The study population for this dissertation consists of non-Hispanic, white individuals from Rochester, Minnesota (MN) that were initially enrolled in the GENOA study. GENOA is a multicenter, community-based study of hypertensive sibships collected to identify genes influencing blood pressure levels and development of target organ damage due to hypertension.[41] In Rochester, MN, the Mayo Clinic diagnostic index and medical record linkage system of the Rochester Epidemiology Project were used to identify non-Hispanic white residents of Olmsted County.[43] In the first phase (Phase I:1995-2000) of GENOA (N=1583 non-Hispanic whites),

sibships containing at least two individuals with clinically diagnosed essential hypertension before age 60 were invited to participate, including both hypertensive and normotensive siblings. Participants were considered to have hypertension if they had either 1) a previous clinical diagnosis of hypertension by a physician with current anti-hypertensive treatment, or 2) an average systolic blood pressure (SBP) ≥140 mm Hg or diastolic blood pressure (DBP) ≥90 mm Hg on the second and third blood pressure readings. Exclusion criteria were secondary hypertension, pregnancy, and insulin-dependent diabetes mellitus. There were 1583 people with phenotype data, and we removed people who had secondary hypertension (N=9), had gender reassignment (N=1), or lacked SNP data (N=138). Participants who did not self report as non-Hispanic whites (N=5) or were outliers (> mean + 4 standard deviation or < mean - 4 standard deviation) for the outcome (SBP, DBP, and BMI) (N=7) were also excluded. 1423 people with phenotype and genotype data were used for genetic association analysis. We then merged this sample with gene expression data and 789 people were retained in the sample for gene expression analysis. The GENOA study was approved by the institutional review board of the Mayo Clinic, Rochester, MN. Written informed consent was obtained from all participants.

Phenotype Measurement

Blood was drawn by venipuncture after an overnight fast of at least eight hours. Resting SBP and DBP were measured with random zero sphygmomanometers (Hawksley and Sons, West Sussex, England) and cuffs appropriate for arm size. Three readings were taken in the right arm after the participant rested in the sitting position for at least five minutes according to standards set by The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High BP (JNC-7) guidelines.[46] Of the three readings, the last two readings were averaged for the analyses. Height was measured by wall stadiometer, weight by

electronic balance, and BMI was calculated by definition as weight in kilograms divided by height in meters squared.

Genotyping and Quality Control of SNPs

In GENOA, a majority of non-Hispanic whites (N=1386) have been successfully genotyped on the Affymetrix® Genome-Wide Human SNP Array 6.0. There were 123 subjects whose genotypes were not successfully measured on the Affymetrix 6.0 platform, so they were re-genotyped using the Illumina Human 1M-DUO BeadChip.

Genotyping was carried out at the Mayo Clinic in Rochester, Minnesota. Briefly, 500ng genomic DNA at 50ng/ul in low ethylene-diamine-tetraacetic acid (EDTA)-Tris buffer was digested in two separate reaction mixtures using the appropriate restriction enzyme (StyI and NspI, 250ng of DNA for each mixture). This step was followed by ligation of an adaptor sequence containing a universal primer sequence. Samples were then subjected to polymerase chain reaction (PCR) (four PCR reactions per sample for the NspI mixture and three for StyI) with conditions designed to amplify 200-2,000 base pairs. The seven PCR products were then combined with Agencourt Ampure beads, passed over an E & K Scientific filter plate, and eluted with elution buffer. Agarose gel analysis of the PCR products and quantification of the amount of PCR product was performed. PCR product concentration were confirmed to be at least 5ug DNA in 1ul EB buffer. Product was then fragmented with DNase I and an agarose gel analysis of the fragmented DNA was used to confirm this step. Following fragmentation, DNA was labeled with Terminal Deoxynucleotidyl Transferase (TdT), hybridized to the appropriate GeneChip, and incubated overnight. The chip was stained and washed on the Affymetrix 450 Fluidics station and then scanned on the Affymetrix 3000 GeneChip scanner.

Preliminary SNP genotype calls were generated using the Dynamic Model (DM) algorithm.[58] The final SNP genotype calls were generated by Birdseed, an algorithm designed especially for the Affymetrix® Genome-Wide Human SNP Array 6.0, and based on the robust linear model with Mahalanobis distance classifier algorithm (RLMM).[59] In order to call genotypes while best accounting for experimental variability and population-specific allele frequencies, Birdseed utilizes information about variation across samples to modify pre-computed genotype calling models from Affymetrix for each SNP probe set. Birdseed has been shown to reduce the bias against heterozygous calls and boost call rates to over 99% while simultaneously increasing concordance rates.[60]

To obtain the final dataset, the following quality control (QC) thresholds were applied: sample call rates >95% and SNP call rates >95%. These thresholds are comparable to those used in the published GWAS.[8,9]

Imputation and Quality Control

Because only a portion of SNPs of interest have been genotyped, imputation methods were used to infer missing or untyped SNP genotypes based on known haplotype information from HapMap.[61] HapMap CEU (release 22 build 36) was used as a reference panel containing 30 trios that were collected from Utah residents with northern and western European descent by the Centre d'Etude du Polymorphisme Humain (CEPH).

Imputation was performed by using MACH v 1.0.16 which implements a Markov Chain based algorithm[62] to infer possible pairs of haplotypes for each individual's genotypes up to ~2.5 million SNPs. SNP imputation combined genotype data from the non-Hispanic white sample in Rochester, MN with the HapMap CEU samples and then inferred genotypes probabilistically according to shared haplotype stretches between study samples and HapMap. Imputation results

are summarized as an "allele dosage" defined as the expected number of copies of the alphabetically higher allele at the SNP (a fractional value between 0.0 and 2.0) for each genotype. The imputation accuracy rates of MACH have been shown to be similar to IMPUTE[63] and higher than several other imputation methods (e.g. fast PHASE, Beagle, and PLINK).[64] Quality control thresholds were applied as follows: SNPs with imputation quality score Rsq>0.3, and Hardy-Weiberg equilibrium p-value $>10^{-3}$ (tested in the GENOA unrelated sample).

Principal component analysis was conducted as another quality control step to identify and remove samples with outlying genotype profiles. First, we removed SNPs that had moderate to poor imputation quality as measured by the estimated R^2 between imputed and true genotypes ($R^2<0.8$) from MACH output. Because GENOA includes sibships, we obtained an unrelated sample of individuals (N=570) by randomly selecting one individual from each sibship. The first ten principal components (PCs) were then calculated on the set of SNPs that were common in both Affymetrix and Illumina platforms and were also available in HapMap (N=226,619 SNPs). The imputed genotypes were used to ensure no missing values for SNPs. An additive model was assumed for the SNPs with standardization of mean equals 0 and variance equals 1. We then used the loading matrix for these PCs to calculate the PC values in the full sample. Individuals with outlier values more than 6 standard deviations on any of the 10 PCs (N=45) were removed from the analysis sample.

Gene Expression Assessment and Quality Control

Blood samples for beta-lymphocyte extraction were collected during GENOA Phase I (1995-2000) study and GENOA Phase II (2000-2005) study. Beta-lymphocytes were isolated and transformed using Epstein-Barr virus at the time of collection. For the present study,

individuals who participated in an ancillary GENOA study: Genetics of Microangiopathic Brain Injury Study (GMBI) had their lymphocytes re-initiated for gene expression assessment. Specifically, cell lines of 237 subjects and 552 subjects were collected and transformed during 1995-2000 and 2001-2005, respectively. Immortalized lymphocytes were stored at -180°C in a freezing medium containing 10% dimethyl sulfoxide (DMSO) as a cryoprotectant. To re-initiate the lymphocytes, a vial of cells is retrieved from cryostorage, thawed at 37°C, rinsed twice in Roswell Park Memorial Institute (RPMI) 1640 culture medium (Irvine Scientific, Santa Ana CA), and transferred to a culture flask along with fresh culture medium supplemented with 15% fetal bovine serum (Irvine Scientific, Santa Ana CA). All cultured lymphocytes were maintained in the standardized culture medium, supplemented with 1% penicillin/streptomycin/L-glutamine (Sigma-Aldrich, St. Louis MO), at 37°C and 90% humidity in 5% CO2 for two to three weeks with regular changes of culture medium until a cell density of 5×10^5 cells/mL was achieved.

RNA samples were extracted using standard protocols. RNA quality was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Foster City, CA) and quantified by spectrophotometry using the Nanodrop ND-1000 (Nanodrop Inc., Wilmington, DE). All RNA samples used in the present study yielded both an A260/A280 absorbance ratio greater than 2.0 and a RNA Integrity Number (RIN) ≥ 8. One μg of RNA was labeled using the WT Expression labeling assay (Applied Biosystems/ Ambion, Foster City, CA) including the labeling controls from the GeneChip Eukaryotic Poly-A RNA Control Kit (Affymetrix, Santa Clara, CA). Each step of the labeling protocol was monitored using the Agilent 2100 Bioanalyzer or the Nanodrop spectrophotometer, as recommended by the manufacturer. Hybridization buffer, Eukaryotic Hybridization Controls, and OligoB2 controls were added to the cDNA fragments just prior to hybridization to the Affymetrix Human Exon 1.0 ST Array. Hybridization was performed at

45°C for 17 hours. Following hybridization, the chips were washed and stained with a phycoerythrin-strepavidin conjugate and were scanned at an excitation wavelength of 488 nm.

Array quality control was performed using Affymetrix Expression Console™ (v 1.1) at the transcript level using core-level probe sets. All array images passed visual inspection. Hybridization controls were all present with signal increases following concentration. Labeling control signal strengths followed the order Lys < Phe < Thr < Dap. Signal intensity plots were examined for raw and processed data to identify outliers.

Raw intensity data were processed using the Affymetrix Power Tool software. Probe summarization and probe set normalization were performed using Robust Multi-array Analysis (RMA), which included background correction, quantile normalization, log2-transformation, and probe set summarization.[47] Only core probesets were used to assess exon-level expression. Gene-level expression was assessed by averaging all the core probe sets for that gene.

Quality control of exon array data were done by examining the QC matrix to identify possible outliers or non-performing samples using the Partek Genomic Suite software (V.6.6). Analyses were restricted to core probe sets. Probe sets which are known to cross-hybridize and those with undetectable expression were also excluded.

Criteria to select SNPs, Genes, and Gene Expression Levels of Interest

Both deCODE genetics and GIANT studies were included in expanded GIANT consortium. Therefore, top SNPs that were identified to be associated with BMI from expanded GIANT were used as the index SNPs to select for proximal SNPs for study (Appendix 4.1).

Because only some of the index SNPs identified from previous GWAS fall within gene boundaries, we used four strategies to choose the set of SNPs and gene expression data to investigate the specific aims based on the location of index SNPs. First, if the index SNP is

inside of a gene, then all of the SNPs that are in a gene and ± 20 kb on either side of that specific gene were selected. Gene expression data for that specific gene was also selected. Second, if the index SNP is outside of a gene and the distance between the index SNP and the closest gene is within 100 kb, then SNPs that are within ± 20 kb of the index SNPs were selected. As for the gene expression data, the closest gene upstream and the closest gene downstream of the index SNP were selected. Third, in gene sparse regions, SNPs that are within ± 20 kb of the index SNPs were selected. As for the gene expression data, there is typically only one gene either upstream or downstream within ± 100kb of the index SNP. In this case, the single gene was chosen for gene expression data. Finally, in gene deserts where there is no gene within ± 100kb of the index SNP, we excluded the index SNP from our analysis. The detailed information of selected SNPs and genes for BMI are shown in Appendix 4.2.

Statistical Methods

All data management and statistical analyses were conducted using R version 2.14. Allele frequencies, genotype frequencies, and linkage disequilibrium (LD), as measured by r^2, was estimated for all pairs of SNPs using an expectation maximization (EM) algorithm. Before performing the statistical analysis, outcome variable distributions were examined for normality and outliers by using histograms and QQ-plots. Outlying values greater than four standard deviations from the mean were removed.

Quantitative variables include: SNP allele dosage (a fractional value between 0.0 and 2.0), age (years), BMI (kg/m^2), and gene expression data (gene level). Categorical variables include gender (0=male 1=female).

Association Testing

The central aim of this chapter is to investigate the functional genetic factors underlying

body mass index variation in a three stage analysis (a) evaluating the associations between genetic variations in the significant and replicated genomic regions and body mass index, (b) investigating the association between these genetic variations and proximal gene expression, and (c) examining the relationship between gene expression and body mass index in the non-Hispanic whites of the Genetic Epidemiology Network of Arteriopathy (GENOA) (Figure 4.1).

For model a (Figure 4.1), we evaluated the association between SNPs and BMI. The outcome variable (BMI) was first adjusted by the adjustment variables (age, and gender) in a linear regression model and the residual was saved. Then, this residual variable was treated as the outcome variable in the linear mixed model to test for the association with the predictor (SNP). Linear mixed effects modeling was used to control for family structure in the dataset. Family id was treated as random intercept to control for family correlation in this dataset. This two step approach was also performed in model b and c. In model b, gene expression data was first adjusted by age, gender and batch effect and the residual was treated as the outcome predicted by the proximal SNPs around that specific gene. For model c, we investigated the association between gene expression (adjustment for batch effect) and BMI (adjustment for age, and gender) by using the same statistical procedure.

The statistical models for three tests are listed as follows and statistical significance of the main effect of each respective SNP was determined based on model p-value (two-sided).

Model (a):

$$\text{BMI residual}_{ij} = \beta_{oi} + \beta_1(\text{SNP}_{ij}) + \varepsilon_{ij}$$

$$\mathbf{H_o}: \beta_1 = 0, \qquad \mathbf{H_a}: \beta_1 \neq 0 \qquad \sim\chi^2_{(df=1)}$$

Model (b):

$$\text{Gene Expression residual}_{ij} = \beta_{oi} + \beta_1(\text{SNP}_{ij}) + \varepsilon_{ij}$$

$$\mathbf{H_o}: \beta_1 = 0, \qquad \mathbf{H_a}: \beta_1 \neq 0 \qquad \sim\chi^2_{(df=1)}$$

Model (c):

$$\text{BMI residual}_{ij} = \beta_{oi} + \beta_1(\text{Gene Expression residual}_{ij}) + \varepsilon_{ij}$$

$$\mathbf{H_o}: \beta_1 = 0, \qquad \mathbf{H_a}: \beta_1 \neq 0 \qquad \sim\chi^2_{(df=1)}$$

In the linear mixed effect model, i represents family and j represents individual. β_{oi} is a random intercept for the i^{th} family, and it is normally distributed with mean β_o and variance σ_b^2. ε_{ij} is the residual variation within the j^{th} individual from the i^{th} family, and it is normally distributed with mean 0 and variance σ^2. ε_{ij} is assumed to be independent from $\beta_{oi.}$

Results

Descriptive statistics for covariates and outcome variables in GENOA non-Hispanic whites are shown in Table 4.1. For the sample with SNP data (N=1423), the average age was 55.3 years old, 55.3% of them were female, nearly half of them (45.26%) were obese, and around three-fourth of them (72.59%) had hypertension. For those subjects with both SNP and gene expression data (N=789), the demographic statistics were approximately the same except there were less people with obesity (41.44% vs. 45.26%). Compared to the sample who were removed (N=634) due to the lack of gene expression data, subjects with both SNP and gene expression data (N=789) had lower BMI (29.73 vs. 31.10, p=0.0001), there was a greater proportion of females (58.05% vs. 51.89% , p=0.023) and they were less obese (41.44 % vs. 50.00% , p=0.023).

As a preliminary step, we examined the association between the 25 Index SNPs reported

from the expanded GIANT with BMI and their effects on it's own gene's expression. Among the 25 index SNPs, only 3 SNPs were identified to be associated with BMI in 3 genetic regions: *FTO* (rs1558902, p= 3x10^{-5}), *TMEM18* (rs2867125, p=7x10^{-4}), and *SEC16B* (rs543874, p= 8x10^{-4}). In addition, 7 SNPs were found to have influence on the expression level around the following 7 genetic regions: *ADCY3* (p=8.16 x10^{-14}), *MTCH2* (p=3.44 x10^{-8}), *NUDT3* (p=6.00 x10^{-6}), *MTIF3* (p=1.16 x10^{-4}), *MAP2K5* (p=0.0057), *STK33* (p=0.0099), and *GPRC5B* (p=0.0182) (Table 4.2). We then compared the results from the association of index SNP and BMI between the expanded GIANT and GENOA studies (Appendix 4.3). In general, the allele frequency of coded allele between expanded GIANT and GENOA were very similar. The beta estimates of the three significant SNPs (rs543874, rs2867125, and rs1558902) associated with BMI in GENOA were larger than those in expanded GIANT (beta= -1.01 vs. -0.22; 1.09 vs. 0.31; 1.01 vs. 0.39, respectively). Because GENOA is a much smaller sample compared to expanded GIANT, we expected that only larger beta estimates would be identified to be significant. Overall, 76% (=19/25 β coefficients) of beta coefficients in GENOA had consistent direction of effect with those in expanded GIANT. Approximately 85% (=21/25 β coefficients) of the effect sizes in GENOA were larger than those in expanded GIANT.

Model a : Association between SNPs and BMI

Using 4213 SNPs proximal to 30 genes, we evaluated these proximal SNPs and their association with BMI in 1423 GENOA non-Hispanic whites. We found that 350 SNPs were associated with BMI based on alpha level of 0.05 (Table 4.3). Among these 350 SNPs with significant signals, 7 SNPs passed the Bonferroni corrected alpha-level (0.05/Number of proximal SNPs for each specific gene): 2 SNPs in *FTO*, 1 SNP in *SEC16B*, and 4 SNPs in

TMEM18. In addition, for *FTO* gene, the index SNP, rs1558902, had the strongest association with BMI of the 350 SNPs examined.

Model b : Association between Proximal SNPs and Gene Expression

We investigated the *cis*-effect of proximal SNPs on their proximal gene's expression of the 30 BMI related genes. Among 4213 association tests, 762 SNPs were significantly associated with gene expression levels and 198 of them passed the Bonferroni corrected alpha-level in the following 10 genes: 13 SNPs in *ADCY3*, 52 SNPs in *TNNI3K*, 17 SNPs in *STK33*, 15 SNPs in *MTCH2*, 54 SNPs in *MAP2K5*, 31 SNPs in *NUDT3*, 4 SNPs in *TMEM18*, 8 SNPs in *MTIF3*, 3 SNPs in *KCTD15*, and 1 SNP in *QPCTL* (Table 4.4). *ADCY3* was found to have the most significant signals and the top SNP, rs6737082, was highly associated with *ADCY3* gene expression (p=5.10x10 $^{-15}$). Out of 20 association tests for *ADCY3* gene, 13 SNPs were significant at the Bonferroni corrected alpha-level at 2.50 x10 $^{-3}$. The second most significant association with gene expression were with the *TNNI3K* gene. Among 260 association tests, 162 SNPs were identified (p<0.05) and 52 of them passed the Bonferroni corrected alpha-level (p<1.90x10^{-4}). Overall, 10 genes showed significant signals based on Bonferroni correction alpha-level and the top SNP for each gene is listed as follows: *ADCY3* (rs6737082, p=5.10 x10^{-15}), *TNNI3K* (rs11210427, p=1.74 x10^{-14}), *STK33* (rs11042008, p=1.94 x10^{-9}), *MTCH2* (rs12794570, p=3.07 x10^{-9}), *MAP2K5* (rs4776943, p=2.00x10^{-8}), *NUDT3* (rs10947494, p=5.25 x10^{-8}), *TMEM18* (rs10153797, p=3.25 x10^{-5}), *MTIF3* (rs9512702, p=5.91 x10^{-5}), *KCTD15* (rs4805059, p=1.10 x10^{-3}), and *QPCTL* (rs11672416, p=3.20 x10^{-3}) (Table 4.4). We then used LocusZoom to plot the regional association of proximal SNPs and gene expression for these 10 genes (Appendix 4.4 – Appendix 4.13). The LocusZoom plot for *ADCY3* shows that all of the 13 significant SNPs were in LD with the most significant SNP (r^2>0.6) except 1 SNP (Appendix

4.4). For *TNNI3K*, there were two clusters of significant SNPs and these two clusters were in different LD block (r^2<0.2) which suggested that different alleles in this locus could cause *TNNI3K* expression(i.e. allelic heterogeneity) (Appendix 4.5). Allelic heterogeneity was also shown in most of the remaining gene regions (Appendix 4.6- Appendix 4.13**)**

Model c : Association between Gene Expression and BMI

Last, gene expression levels of 30 genes were used to examine their associations with BMI. Only one gene was found to have gene expression level associated with BMI: *TFAP2B* (p=0.0042) and it did not pass Bonferroni correction (alpha level of 0.0017 (=0.05/30 genes). We then summarized the results from model a to model c based on alpha level of 0.05 in Table 4.5. In total, there were 33 SNPs that have significant association with both BMI and expression of 8 genes: 7 SNPs for *SLC39A8*, 1 SNP for *NUDT3*, 6 SNPs for *TNNI3K*, 8 SNPs for *FTO*, 1 SNPs for *NRXN3*, 1 SNP for *ETV5*, 8 SNPs for *SEC16B*, and 1 SNP for *TMEME18*. The information and association results for these 33 SNPs were listed in Appendix 4.14. All of these 33 SNPs were in introns and only one of these associations passed the Bonferroni corrected alpha level (association between rs10947494 and *NUTD3* expression). Rs10947494 was also the top SNP that has the strongest *cis*-effect on *NUTD3* expression (Table 4.4). None of these 30 genes had significant results in all three models (i.e. model a, b, c).

Sensitivity Analysis

We did sensitivity analysis by using the sample with 789 subjects to test for the SNP- BMI associations. Out of 4213 SNPs association tests, there were 24 SNPs associated with both BMI and gene expression levels (p<0.05) and none of them passed Bonferroni correction (Appendix 4.15). In the previous anlaysis with larger sample size (N=1423), we found 33 SNPs associated

with both BMI and gene expression and 24 SNPs of them were identified in the smaller sample . In addition, we did not identify any new findings in this smaller sample.

Discussion

Among 4213 SNPs, 33S NPs located in or near to 8 genes were identified to be simultaneously associated with BMI and gene expression. Only one signal (i.e. association between rs10947494 and *NUTD3* expression) of these 33 associations passed Bonferroni correction (Appendix 4.14). In addition, none of these 8 genes showed significant association of their expression level with BMI. The lack of identification of SNPs significantly associated with both BMI and gene expression after Bonferroni correction could be explained by several reasons.

First, many of the genes identified from previous BMI GWAS are expressed and known to act in the central nervous system.[29] Animal models and studies of monogenic human obesity have also shown that neuronal genes play a major role in susceptibility to obesity, [36,88,89] particularly for those genes that are expressed in the hypothalamus and involved in regulation of appetite or energy balance. In GIANT, they measured the expression of 8 genes (*FTO, MC4R, TMEM18, GNPDA2, MTCH2, KCTD15, NEGR1, and SH2B1*) in different human tissues (total brain, cerebellum, hypothalamus, cortext, pancreas, heart, adipocytes, spleen, liver, lung, kidney, and testes) and found all genes except *MTCH2* were highly expressed in the brain and/or hypothalamus.[29]

Second, expanded GIANT conducted gene expression quantitative trait loci (eQTL) analyses to examine the *cis*-effects of SNPs on gene expression level by using multiple tissues: human brain, lymphocyte, blood, subcutaneous and visceral adipose tissue, and liver. Among 32 BMI associated genetic loci, they observed significant *cis*-associations between 14 BMI-associated alleles and gene expression levels in various tissues for multiple genes.[9] For lymphocyte tissue,

the significant *cis*-associations were observed in the following 11 genes: *ADCY3, LGR4, FNBP4, NDUFS3, GTF3A, POLR1D, C16orf88, EIF3S8, SPNS1, TUFM,* and *TMEM160*. Among these 11 genes, only *ADCY3* expression was examined and found to have significant SNPs in GENOA non-Hispanic whites. The top SNP, rs6737082, associated with *ADCY3* expression is the same in both expanded GIANT (p=1.79x10^{-21}) and GENOA (p=5.10x10^{-15}, Table 4.4) and this SNP is in high LD with the index SNP rs713586 (LD r^2=0.78). Additionally, this index SNP, rs713586, was significantly associated with BMI in expanded GIANT (p=6.17x10^{-22}) but not in GENOA (p=0.79, Appendix 4.3). Based on the results of eQTL analyses from expanded GIANT, significant *cis*-associations were detected among various tissues for each gene. For example, the *cis*-association of *NEGR1* was only expressed in blood tissue while the *cis*-associations of *SH2B1* were expressed in both adipose and omental tissue. In GENOA, transformed beta-lymphocytes were used as the only tissue for gene expression data and thus this could explain why we did not detect significant *cis*-association for some genes.

Lastly, in expanded GIANT, they selected all gene expression data that were within 1MB of the index SNPs while we only chose genes that were within 100 kb of the index SNPs. Their results also indicated that most BMI-associated variants were associated with the expression of multiple nearby genes but we only examined the expression of 1 or 2 nearby genes. This suggested that we might not be able capture the whole spectrum of the *cis*-associations for multiple genes around the index SNP and we could miss some important gene expression signals. However, due to the limited sample size, we chose 1 to 2 closest genes around the index SNPs to maintain the appropriate power in our study.

Out of 30 investigated associations between gene expression level and BMI, we found only 1 significant association (i.e. *TFAP2B* expression associated with BMI, p=0.042), but this

association did not pass Bonferroni correction (alpha level (0.05/30=0.00167)). Since we only used transformed beta-lymphocytes for gene expression data, we then searched for literature on associations between gene expression level and BMI in other tissues. Emilsson *et al.* used blood and adipose tissues for gene expression data and investigated the association between 23,720 gene transcripts and obesity related traits (BMI, percent body fat (PBF), and waist-to-hip ratio(WHR)) in a large population based sample of non-Hispanic whites.[51] In blood, they found 2,172 (9.2%) gene expression traits to be correlated with BMI, 1,098 (4.6%) with PBF, and 711 (3.0%) with WHR. For adipose tissue, 17,080 (72.0%) gene expression traits were found to be correlated with BMI, 16,977 (71.6%) with PBF, and 14,901 (62.8%) with WHR. They observed more gene transcripts associated with obesity related traits in adipose tissue than in blood. In addition, they found 2,784 of the gene transcripts in adipose tissue explained more than 10% of the BMI variation while none of the gene transcripts in blood achieved this level. Their results indicated that adipose tissue is better suited for estimation of the gene expression associated with BMI than blood. This implied that transformed lymphocyte cell lines might not be the most relevant cells for unraveling links between SNPs and BMI.

Zeller *et al.* examined whether monocyte gene expression might mediate the effects of loci recently identified in BMI GWAS in 1490 unrelated individuals in the Gutenberg Heart Study (GHS), a community based project conducted in a single centre in the region of Mainz (Germany).[52] Out of 10 top SNPs at 10 genetic loci found in another GWAS study by Thorleifsson et al,[30] only rs8050136 in the *FTO* gene was found to be significantly associated with BMI (p=0.0034) in the GHS. However, this SNP was not found to be correlated with gene expression. On the other hand, two SNPs (rs8049439 and rs29942) located around the *SH2B1* and *KCTD15* genes were found to have no association with BMI (p=0.71 and p=0.87,

respectively), but they were correlated with the gene expression of nearby genes (rs8049439: *CCDC101*, p=6.95x10^{-39}, *SPNS1*, p=6.55 x10^{-64}, *TUFM*, p=3.11x10^{-76}; rs29942: *KCTD151*, p=4.58x10^{-15}) in the GHS. In addition, two of these four gene expression traits were significantly associated with BMI (*CCDC101*, p=0.0004, *SPNS1*, p=0.0839, *TUFM*, p=0.2609; *KCTD151*, p=0.0030). Their results revealed that none of the SNP - BMI associations were mediated by gene expression. Our results also indicated that the top SNPs that were found to be strongly associated with BMI were not those SNPs that had strong *cis*-effects on gene expression in GENOA (Appendix 4.14). It is possible that SNPs with the strongest effects on gene expression level are different from those that are known to affect BMI. In addition, monocytes or lymphocytes might not be the most relevant cells linking the association between genome variants and BMI, which has been shown in the previous literature and discussed above.

Non-genetic factors, such as epigenetics, can also influence gene expression, including 1) DNA methylation and 2) post-translation modification of histone proteins and remodeling of chromatin. Another mechanism that could influence gene expression is the interaction between SNPs and risk factors on gene expression. Zeller *et al.* investigated the non-genetic risk factors that influence gene expression.[52] Out of 12,808 gene expression traits, 807 gene expression traits were associated with gender and 396 by age. For gene expression traits that were simultaneously associated with SNPs and risk factors (94 for age and 230 for gender), they examined the interaction between these two variables on gene expression and found significant interactions signals. Based on the alpha level of 0.05, they found 7 age by SNP interactions, and 30 gender by SNP interactions. The most significant age by SNP interaction was age by rs9415998 on *CDH23* expression with p-value of 0.000624. The strongest signal for gender by SNP interaction is rs4830487 by *ISCU* interaction (p= 0.000022). Understanding the non-genetic factors and their

interaction with genetic variants contributing to gene expression and protein level would help us to unravel the complex biological process linking gene expression to a health outcome, such as BMI.

There are several limitations of this study that need to be considered. 1) The linear mixed modeling we used in this study might be too simplistic to account for a much more complex biological reality that is associated with genetic variants, gene expression and BMI. 2) Compared to the largest BMI GWAS that has more than 200, 000 individuals,[9] our sample size is relatively small and thus the power to detect the moderate effects of the genetic variants is limited. We did power calculations by using a standardized phenotype (mean= 0, standardized deviation= 1) and an effective sample size of 989 individuals. At an alpha level of 0.05, the power to detect the association between SNPs and BMI varied from 0.52 (MAF=0.1, β=0.15) to 0.92 (MAF=0.5, β=0.15) (Appendix 4.16). 3) Due to the limited power, we only examined proximal SNPs that might have *cis*-effects on gene expression, but it is possible that SNPs that with *trans*-effects might actually have important influence on gene expression levels that affect BMI. However, we were not able to investigate their influence in a study with this sample size.

In summary, although we did not find any gene that has significant associations across the three models, highly significant *cis*-associations were detected in 10 genes *ADCY3*, *TNNI3K*, *STK33*, *MTCH2*, *MAP2K5*, *NUDT3*, *TMEM18*, *MTIF3, KCTD15,* and *QPCTL* (p=6x10^{-5}). Significant *cis* associations between the top SNP, rs6737082, and *ADCY3* expression was reported in both lymphocyte tissues and beta-transformed lymphocytes from the expanded GIANT and GENOA studies, respectively. The consistent findings between these two studies suggest that strong associations are more robust and more likely to be detected by different studies. *ADCY3* encodes adenylyl cyclase 3 which is a membrane-associated enzyme and

catalyzes the formation of the secondary messenger cyclic adenosine monophosphate (cAMP). This protein appears to be widely expressed in various human tissues (brain, heart, kidney, liver, lung, pancreas, placenta, and skeletal muscle). Yang *et al.* suggested that *ADCY3* expression may be involved with some physiological and pathophysiological metabolic processes.[90] The potential functions for the 30 BMI-related genes are listed in Appendix 4.17.1- 4.17.3.

Our results have shown the evidence of *cis*-effects of genetic variants on the expression levels of these 10 genes. Future studies focused on BMI-related tissue such as adipose tissue will aid in detecting more gene expression levels associated with BMI.

Figure 4.1. Association analysis among SNPs, gene expression and BMI.

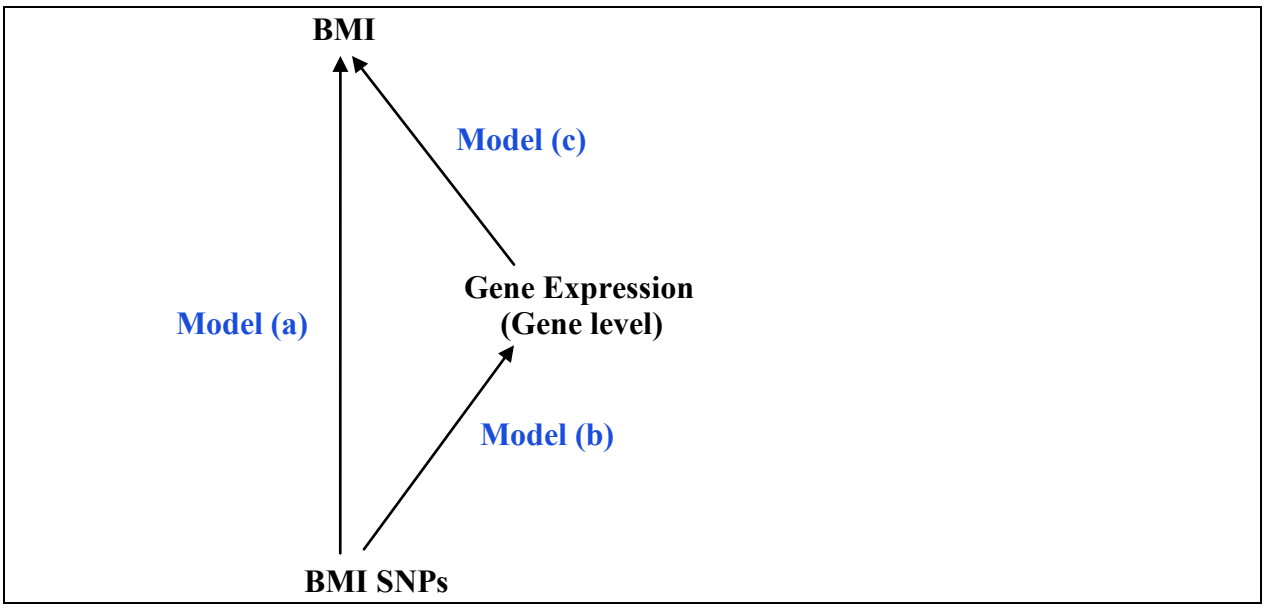


Table 4.1. Descriptive characteristics of non-Hispanic whites in GENOA.

	Genetic Association (N=1423)		With Gene Expression Data (N=789)		Without Gene Expression Data (N=634)		
Continuous Variables	Mean	SD	Mean	SD	Mean	SD	T-test P-value†
Age, years	55.30	10.76	55.65	10.36	54.87	11.23	0.173
Height, cm	168.75	9.24	168.41	9.09	169.17	9.41	0.125
Weight, kg	86.59	19.46	84.53	18.64	89.14	20.15	**0.0001**
BMI, kg/m^2	30.34	6.16	29.73	5.82	31.10	6.48	**0.0001**
Categorical Variables	N	%	N	%	N	%	χ^2 –test P-value†
Female	787	55.31	458	58.05	329	51.89	**0.023**
Obesity	644	45.26	327	41.44	317	50.00	**0.002**

† Statistical tests compared the subjects with both SNP and gene expression data (n=789) to subjects whithout gene expression data (n=634).

Table 4.2. Association of twenty-five index SNPs, BMI and gene expression in GENOA.

Index SNP	Chr	SNP Position	Nearest Gene	BMI		Gene Expression	
				Beta	P-value	Beta	P-value
rs713586	2	25,158,008	*ADCY3*	-0.064	0.7963	**-0.099**	**8.16E-14**
rs3817334	11	47,650,993	*MTCH2*	-0.227	0.3495	**0.055**	**3.44E-08**
rs206936	6	34,302,869	*NUDT3*	0.317	0.2999	**-0.048**	**6.00E-06**
rs4771122	13	28,020,180	*MTIF3*	0.159	0.5906	**0.042**	**0.00016**
rs2241423	15	68,086,838	*MAP2K5*	-0.453	0.1446	**-0.040**	**0.0057**
rs4929949	11	8,604,593	*STK33*	0.006	0.9804	**-0.048**	**0.0099**
rs12444979	16	19,933,600	*GPRC5B*	-0.191	0.5841	**-0.036**	**0.0182**
rs10767664	11	27,725,986	*BDNF*	0.491	0.1074	-0.018	0.0902
rs2287019	19	46,202,172	*QPCTL*	0.161	0.7032	-0.026	0.0958
rs9816226	3	185,834,499	*ETV5*	-0.608	0.0535	-0.066	0.1089
rs1514175	1	74,991,644	*TNNI3K*	0.442	0.0695	-0.013	0.1103
rs2890652	2	142,959,931	*LRP1B*	-0.182	0.5662	0.013	0.1880
rs13107325	4	103,188,709	*SLC39A8*	-0.330	0.5047	0.030	0.2299
rs9816226	3	185,834,499	*DGKG*	-0.608	0.0535	-0.020	0.2372
rs7138803	12	50,247,468	*FAIM2*	-0.006	0.9809	0.013	0.2609
rs2112347	5	75,015,242	*POC5*	-0.317	0.2116	0.012	0.2724
rs2867125	2	622,827	*TMEM18*	**1.092**	**0.0007**	0.012	0.2865
rs10150332	14	79,936,964	*NRXN3*	0.193	0.5242	0.036	0.3783
rs7359397	16	28,885,659	*SH2B1*	-0.206	0.4059	0.006	0.3833
rs987237	6	50,803,050	*TFAP2B*	-0.541	0.0797	0.025	0.4533
rs29941	19	34,309,532	*CHST8*	-0.382	0.1390	-0.007	0.4651
rs2815752	1	72,812,440	*NEGR1*	0.330	0.1833	-0.006	0.5475
rs29941	19	34,309,532	*KCTD15*	-0.382	0.1390	0.005	0.6484
rs7138803	12	50,247,468	*BCDIN3D*	-0.006	0.9809	-0.003	0.7446
rs10968576	9	28,414,339	*LINGO2*	-0.164	0.5252	-0.002	0.8284
rs13078807	3	85,884,150	*CADM2*	-0.187	0.5493	-0.002	0.8801
rs7359397	16	28,885,659	*ATP2A1*	-0.206	0.4059	0.001	0.8848
rs1558902	16	53,803,574	*FTO*	**1.007**	**3.50E-05**	0.0004	0.9536
rs543874	1	177,889,480	*SEC16B*	**-1.006**	**0.0008**	0.001	0.9584
rs713586	2	25,158,008	*DNAJC27*	-0.064	0.7963	0.0005	0.9638

Bold:$p<0.05$.

Table 4.3. Significant results from association of SNPs and BMI. †

Index SNP	Chr	Nearest Gene	SNP Selection Range		No. of Proximal SNPs	Bonferroni Corrected Alpha-Level	No. of Proximal SNPs associated with BMI (p<0.05)	No. of Significant SNPs based on Bonferroni Corrected Alpha-Level	Top SNP with lowest P-value		
			Star	End					SNP	Beta	P-value
Gene Region with Index SNP inside of it											
rs2287019	19	*QPCTL*	46,187,240	46,227,240	12	4.20E-03	2	0	rs2334255	-0.78	0.0411
rs4771122	13	*MTIF3*	28,004,711	28,044,711	55	9.10E-04	0	0	rs9581861	0.60	0.0579
rs3817334	11	*MTCH2*	47,644,206	47,684,206	20	2.50E-03	0	0	rs12794570	-0.26	0.3196
rs987237	6	*TFAP2B*	50,795,325	50,835,325	57	8.80E-04	8	0	rs16880682	-1.19	0.0088
rs13107325	4	*SLC39A8*	103,246,655	103,286,655	133	3.80E-04	26	0	rs4698845	-0.90	0.0017
rs206936	6	*NUDT3*	34,340,441	34,380,441	37	1.40E-03	1	0	rs10947494	-0.51	0.0449
rs4929949	11	*STK33*	8,595,503	8,635,503	133	3.80E-04	0	0	rs7942490	0.89	0.1355
rs2241423	15	*MAP2K5*	68,079,451	68,119,451	244	2.00E-04	1	0	rs7180716	-1.36	0.0123
rs13078807	3	*CADM2*	86,097,948	86,137,948	230	2.20E-04	56	0	rs9990223	-0.93	0.0119
rs1514175	1	*TNNI3K*	74,990,108	75,030,108	260	1.90E-04	10	0	rs1881029	-1.25	0.0221
rs1558902	16	*FTO*	54,128,378	54,168,378	454	1.10E-04	87	2	rs1558902	1.00	**3.50E-05**
rs10968576	9	*LINGO2*	28,699,303	28,739,303	955	5.24E-05	23	0	rs10812752	-0.64	0.0125
rs10150332	14	*NRXN3*	80,310,758	80,350,758	1229	4.07E-05	29	0	rs17107557	1.38	0.0112
Gene Dense Region (Multiple Genes within ±100 kb of Index SNP)											
rs7138803	12	*BCDIN3D*	50,227,468	50,267,468	31	1.60E-03	0	0	rs7134749	-0.81	0.1683
	12	*FAIM2*	50,227,468	50,267,468							
rs7359397	16	*SH2B1*	28,865,659	28,905,659	13	3.80E-03	0	0	rs4788101	-0.21	0.3973
	16	*ATP2A1*	28,865,659	28,905,659							
rs29941	19	*KCTD15*	34,289,532	34,329,532	26	1.90E-03	0	0	rs256335	-0.43	0.0867
	19	*CHST8*	34,289,532	34,329,532							
rs713586	2	*DNAJC27*	25,138,008	25,178,008	20	2.50E-03	0	0	rs17046742	-0.31	0.4026
	2	*ADCY3*	25,138,008	25,178,008							
rs9816226	3	*ETV5*	185,814,499	185,854,499	28	1.80E-03	2	0	rs16860471	0.89	0.0067
	3	*DGKG*	185,814,499	185,854,499							
rs12444979	16	*GPRC5B*	19,913,600	19,953,600	19	2.60E-03	0	0	rs6497430	0.53	0.2773
rs543874	1	*SEC16B*	177,869,480	177,909,480	53	9.40E-04	18	1	rs543874	-1.01	**8.30E-04**
Gene Sparse Region (Only One Gene within ±100 kb of Index SNP)											
rs2867125	2	*TMEM18*	602,827	642,827	101	5.00E-04	81	4	rs2867131	1.22	**1.70E-04**
rs2112347	5	*POC5*	74,995,242	75,035,242	28	1.80E-03	0	0	rs7734618	-0.55	0.0582
rs10767664	11	*BDNF*	27,705,986	27,745,986	14	3.60E-03	4	0	rs2030324	-0.49	0.0392
rs2815752	1	*NEGR1*	72,792,440	72,832,440	14	3.60E-03	2	0	rs2815749	-0.77	0.0170
rs2890652	2	*LRP1B*	142,939,931	142,979,931	47	1.10E-03	0	0	rs1523708	0.50	0.0629
Total	---	----------	----------	----------	4213	----------	350	7	----------	-----	----------

†Genes were sorted by gene size (See Appendix 4.2). Bold:*p*-value passed Bonferroni Corrected Alpha-Level.

* The Index SNP rs1558902 is the same as the top SNP with lowest p-value for *FTO* gene.

Table 4.4. Significant results from association of proximal SNPs and gene expression.

Gene	Chr	SNP Selection Range		No. of Proximal SNPs	Bonferroni Corrected Alpha-Level	No. of Proximal SNPs associated with Gene Expression (p<0.05)	No. of Significant SNPs based on Bonferroni Corrected Alpha-Level	Top SNP with lowest P-value		
		Star	End					SNP	Beta	P-value
ADCY3	2	25,138,008	25,178,008	20	2.50E-03	13	13	rs6737082	0.107	**5.10E-15**
TNNI3K	1	74,990,108	75,030,108	260	1.90E-04	162	52	rs11210427	-0.068	**1.74E-14**
STK33	11	8,595,503	8,635,503	133	3.80E-04	87	17	rs11042008	-0.129	**1.94E-09**
MTCH2	11	47,644,206	47,684,206	20	2.50E-03	15	15	rs12794570	0.062	**3.07E-09**
MAP2K5	15	68,079,451	68,119,451	244	2.00E-04	136	54	rs4776943	-0.077	**2.00E-08**
NUDT3	6	34,340,441	34,380,441	37	1.40E-03	35	31	rs10947494	0.048	**5.25E-08**
TMEM18	2	602,827	642,827	101	5.00E-04	6	4	rs10153797	-0.046	**3.25E-05**
MTIF3	13	28,004,711	28,044,711	55	9.10E-04	39	8	rs9512702	0.040	**5.91E-05**
LINGO2	9	28,699,303	28,739,303	955	5.24E-05	23	0	rs1197920	0.035	0.0007
KCTD15	19	34,289,532	34,329,532	26	1.90E-03	6	3	rs4805059	-0.033	**0.0011**
SEC16B	1	177,869,480	177,909,480	53	9.40E-04	18	0	rs570762	0.035	0.0012
FTO	16	54,128,378	54,168,378	454	1.10E-04	52	0	rs17218700	0.029	0.0013
NRXN3	14	80,310,758	80,350,758	1229	4.07E-05	93	0	rs17764884	-0.140	0.0016
SLC39A8	4	103,246,655	103,286,655	133	3.80E-04	17	0	rs17032436	0.087	0.0025
QPCTL	19	46,187,240	46,227,240	12	4.20E-03	2	1	rs11672416	0.047	**0.0032**
CHST8	19	34,289,532	34,329,532	26	1.90E-03	3	0	rs4239577	0.024	0.0075
BCDIN3D	12	50,227,468	50,267,468	31	1.60E-03	6	0	rs4898538	-0.027	0.0106
GPRC5B	16	19,913,600	19,953,600	19	2.60E-03	12	0	rs12444979	-0.036	0.0182
DGKG	3	185,814,499	185,854,499	28	1.80E-03	6	0	rs6783586	-0.034	0.0219
DNAJC27	2	25,138,008	25,178,008	20	2.50E-03	6	0	rs955855	-0.032	0.0316
CADM2	3	86,097,948	86,137,948	230	2.20E-04	19	0	rs6796402	0.026	0.0366
LRP1B	2	142,939,931	142,979,931	47	1.10E-03	3	0	rs13408409	0.027	0.0394
ETV5	3	185,814,499	185,854,499	28	1.80E-03	3	0	rs6802450	0.082	0.0467
FAIM2	12	50,227,468	50,267,468	31	1.60E-03	0	0	rs4898538	-0.029	0.0594
TFAP2B	6	50,795,325	50,835,325	57	8.80E-04	0	0	rs12206160	-0.071	0.0613
BDNF	11	27,705,986	27,745,986	14	3.60E-03	0	0	rs988748	0.018	0.0893
POC5	5	74,995,242	75,035,242	28	1.80E-03	0	0	rs6453143	0.018	0.1023
SH2B1	16	28,865,659	28,905,659	13	3.80E-03	0	0	rs6565259	-0.006	0.3595
ATP2A1	16	28,865,659	28,905,659	13	3.80E-03	0	0	rs6565259	-0.005	0.4526
NEGR1	1	72,792,440	72,832,440	14	3.60E-03	0	0	rs2815752	-0.006	0.5475
Total	----	----------	----------	4213	----------	762	198	----------	------	----------

Bold:*p*-value passed Bonferroni Corrected Alpha-Level.

Table 4.5. Summary results from association of SNPs, gene expression and BMI.†

Gene	Chr	SNP Selection Range		No. of Proximal SNPs	No. of SNPs associated with BMI & Gene Expression (p<0.05)	Association of BMI & Gene Expression	
		Star	End			Beta	P-value
Gene Region with Index SNP inside of it							
QPCTL	19	46,187,240	46,227,240	12	0	1.57	0.175
MTIF3	13	28,004,711	28,044,711	55	0	1.57	0.175
MTCH2	11	47,644,206	47,684,206	20	0	0.22	0.838
TFAP2B	6	50,795,325	50,835,325	57	0	-0.80	**0.042**
SLC39A8	4	103,246,655	103,286,655	133	7	0.31	0.690
NUDT3	6	34,340,441	34,380,441	37	1	-0.52	0.678
STK33	11	8,595,503	8,635,503	133	0	-0.41	0.481
MAP2K5	15	68,079,451	68,119,451	244	0	1.21	0.183
CADM2	3	86,097,948	86,137,948	230	0	-0.08	0.926
TNNI3K	1	74,990,108	75,030,108	260	6	-0.55	0.639
FTO	16	54,128,378	54,168,378	454	8	-0.27	0.868
LINGO2	9	28,699,303	28,739,303	955	0	0.51	0.606
NRXN3	14	80,310,758	80,350,758	1229	1	-0.19	0.544
Gene Dense Region (Multiple Genes within ±100 kb of Index SNP)							
BCDIN3D	12	50,227,468	50,267,468	31	0	0.15	0.904
FAIM2	12	50,227,468	50,267,468		0	-0.11	0.908
SH2B1	16	28,865,659	28,905,659	13	0	-0.32	0.834
ATP2A1	16	28,865,659	28,905,659		0	2.80	0.107
KCTD15	19	34,289,532	34,329,532	26	0	-0.04	0.968
CHST8	19	34,289,532	34,329,532		0	0.35	0.768
DNAJC27	2	25,138,008	25,178,008	20	0	0.47	0.644
ADCY3	2	25,138,008	25,178,008		0	-0.22	0.780
ETV5	3	185,814,499	185,854,499	28	1	0.19	0.583
DGKG	3	185,814,499	185,854,499		0	0.67	0.409
GPRC5B	16	19,913,600	19,953,600	19	0	-0.63	0.493
SEC16B	1	177,869,480	177,909,480	53	8	-0.36	0.713
Gene Sparse Region (Only One Gene within ±100 kb of Index SNP)							
TMEM18	2	602,827	642,827	101	1	0.40	0.739
POC5	5	74,995,242	75,035,242	28	0	0.09	0.929
BDNF	11	27,705,986	27,745,986	14	0	-1.16	0.337
NEGR1	1	72,792,440	72,832,440	14	0	-1.11	0.299
LRP1B	2	142,939,931	142,979,931	47	0	-0.52	0.687
Total	----	----------	----------	4213	33	----------	----------

†Genes were sorted by gene size (See Appendix 4.2). Bold: *p*-value <0.05

Appendix 4.1. Summary association results for 32 BMI associated SNPs from expanded Genetic Investigation of Anthropometric Traits (GIANT).

Index SNP	Nearest Gene	Function	Chr	SNP Position	Coded/ Non-Coded allele	Coded Allele Freq.	Beta	P-value
Novel loci identified by expanded GIANT consortium								
rs1514175	*TNNI3K*	Intron	1	74,991,644	A/G	0.43	0.07	8.16E-14
rs1555543	*PTBP2*	Intergenic	1	96,944,797	C/A	0.59	0.06	3.68E-10
rs713586	*ADCY3*	Intergenic	2	25,158,008	C/T	0.47	0.14	6.17E-22
rs887912	*FLJ30838*	Intergenic	2	59,302,877	T/C	0.29	0.10	1.79E-12
rs2890652	*LRP1B*	Intergenic	2	142,959,931	C/T	0.18	0.09	1.35E-10
rs13078807	*CADM2*	Intron	3	85,884,150	G/A	0.20	0.10	3.94E-11
rs13107325	*SLC39A8*	Missense	4	103,188,709	T/C	0.07	0.19	1.50E-13
rs2112347	*POC5*	Intergenic	5	75,015,242	T/G	0.63	0.10	2.17E-13
rs4836133	*ZNF608*	Intergenic	5	124,332,103	A/C	0.48	0.07	1.97E-09
rs206936	*NUDT3*	Intron	6	34,302,869	G/A	0.21	0.06	3.02E-08
rs10968576	*LINGO2*	Intron	9	28,414,339	G/A	0.31	0.11	2.65E-13
rs4929949	*STK33*	Intron	11	8,604,593	C/T	0.52	0.06	2.80E-09
rs4771122	*MTIF3*	Intron	13	28,020,180	G/A	0.24	0.09	9.48E-10
rs11847697	*PRKD1*	Intergenic	14	30,515,112	T/C	0.04	0.17	5.76E-11
rs2241423	*MAP2K5*	Intron	15	68,086,838	G/A	0.78	0.13	1.19E-18
rs12444979	*GPRC5B*	Intergenic	16	19,933,600	C/T	0.87	0.17	2.91E-21
rs2287019	*QPCTL*	Intron	19	46,202,172	C/T	0.80	0.15	1.88E-16
rs3810291	*ZC3H4*	Intron	19	47,569,003	A/G	0.67	0.09	1.64E-12
rs987237*	*TFAP2B*	Intron	6	50,803,050	G/A	0.18	0.13	2.90E-20
rs10150332*	*NRXN3*	Intron	14	79,936,964	C/T	0.21	0.13	2.75E-11
Loci that were identified in expanded GIANT consortium and also have been reported in GIANT and deCODE genetics								
rs2815752	*NEGR1*	Intergenic	1	72,812,440	A/G	0.61	0.13	1.16E-22
rs543874	*SEC16B*	Intergenic	1	177,889,480	G/A	0.19	0.22	3.56E-23
rs2867125	*TMEM18*	Intergenic	2	622,827	C/T	0.83	0.31	2.77E-49
rs9816226	*ETV5*	Intergenic	3	185,834,499	T/A	0.82	0.14	1.69E-18
rs10938397	*GNPDA2*	Intergenic	4	45,182,527	G/A	0.43	0.18	3.78E-31
rs10767664	*BDNF*	Intergenic	11	27,725,986	A/T	0.78	0.19	4.69E-26
rs3817334	*MTCH2*	Intron	11	47,650,993	T/C	0.41	0.06	1.59E-12
rs7138803	*FAIM2*	Intergenic	12	50,247,468	A/G	0.38	0.12	1.82E-17
rs7359397	*SH2B1*	Intergenic	16	28,885,659	T/C	0.40	0.15	1.88E-20
rs1558902	*FTO*	Intron	16	53,803,574	A/T	0.42	0.39	4.80E-120
rs571312	*MC4R*	Intergenic	18	57,839,769	A/C	0.24	0.23	6.43E-42
rs29941	*KCTD15*	Intergenic	19	34,309,532	G/A	0.67	0.06	3.01E-09

*Previously identified waist circumference loci.

Appendix 4.2. Summary of thirty-two BMI associated index SNPs and their closest genes. †

Index SNP	Nearest Gene	Chr	SNP Position	Gene Start	Gene End	Gene Size (bp)	SNP Selection Range		SNP Number (Total=4213)
							Star	End	
Gene Region with Index SNP inside of it									
rs2287019	*QPCTL*	19	46,202,172	46,195,741	46,207,240	11,499	46,187,240	46,227,240	12
rs4771122	*MTIF3*	13	28,020,180	28,009,783	28,024,711	14,928	28,004,711	28,044,711	55
rs3817334	*MTCH2*	11	47,650,993	47,638,859	47,664,206	25,347	47,644,206	47,684,206	20
rs987237	*TFAP2B*	6	50,803,050	50,786,439	50,815,325	28,886	50,795,325	50,835,325	57
rs13107325	*SLC39A8*	4	103,188,709	103,182,821	103,266,655	83,834	103,246,655	103,286,655	133
rs206936	*NUDT3*	6	34,302,869	34,256,002	34,360,441	104,439	34,340,441	34,380,441	37
rs4929949	*STK33*	11	8,604,593	8,413,418	8,615,503	202,085	8,595,503	8,635,503	133
rs2241423	*MAP2K5*	15	68,086,838	67,835,021	68,099,451	264,430	68,079,451	68,119,451	244
rs13078807	*CADM2*	3	85,884,150	85,775,632	86,117,948	342,316	86,097,948	86,137,948	230
rs1514175	*TNNI3K*	1	74,991,644	74,663,926	75,010,108	346,182	74,990,108	75,030,108	260
rs1558902	*FTO*	16	53,803,574	53,737,875	54,148,378	410,503	54,128,378	54,168,378	454
rs10968576	*LINGO2*	9	28,414,339	27,948,528	28,719,303	770,775	28,699,303	28,739,303	955
rs10150332	*NRXN3*	14	79,936,964	78,870,093	80,330,758	1,460,665	80,310,758	80,350,758	1229
Gene Dense Region (Multiple Genes within ±100 kb of Index SNP)									
rs7138803	*BCDIN3D*	12	50,247,468	50,229,826	50,236,912	7,086	50,227,468	50,267,468	31
	FAIM2	12	50,247,468	50,260,679	50,297,760	37,081	50,227,468	50,267,468	
rs7359397	*SH2B1*	16	28,885,659	28,875,078	28,885,533	10,455	28,865,659	28,905,659	13
	ATP2A1	16	28,885,659	28,889,809	28,915,830	26,021	28,865,659	28,905,659	
rs29941	*KCTD15*	19	34,309,532	34,287,751	34,306,665	18,914	34,289,532	34,329,532	26
	CHST8	19	34,309,532	34,112,861	34,264,414	151,553	34,289,532	34,329,532	
rs713586	*DNAJC27*	2	25,158,008	25,166,505	25,194,824	28,319	25,138,008	25,178,008	20
	ADCY3	2	25,158,008	25,042,039	25,142,055	100,016	25,138,008	25,178,008	
rs9816226	*ETV5*	3	185,834,499	185,764,108	185,826,901	62,793	185,814,499	185,854,499	28
	DGKG	3	185,834,499	185,864,990	186,080,023	215,033	185,814,499	185,854,499	
rs12444979	*GPRC5B**	16	19,933,600	19,870,295	19,896,151	25,856	19,913,600	19,953,600	19
rs543874	*SEC16B***	1	177,889,480	177,897,489	177,939,050	41,561	177,869,480	177,909,480	53
Gene Sparse Region (Only One Gene within ±100 kb of Index SNP)									
rs2867125	*TMEM18*	2	622,827	667,975	677,439	9,464	602,827	642,827	101
rs2112347	*POC5*	5	75,015,242	74,970,024	75,013,313	43,289	74,995,242	75,035,242	28
rs10767664	*BDNF*	11	27,725,986	27,676,442	27,722,600	46,158	27,705,986	27,745,986	14
rs2815752	*NEGR1*	1	72,812,440	71,868,626	72,748,405	879,779	72,792,440	72,832,440	14
rs2890652	*LRP1B*	2	142,959,931	140,988,996	142,889,270	1,900,274	142,939,931	142,979,931	47
Index SNPs Excluded (No gene within ±100 kb of Index SNP)									
rs571312	*MC4R*	18	57,839,769	58,038,564	58,040,001	1,437	NA	NA	NA
rs10938397	*GNPDA2*	4	45,182,527	44,704,169	44,728,612	24,443	NA	NA	NA
rs1555543	*PTBP2*	1	96,944,797	97,187,175	97,280,599	93,424	NA	NA	NA
rs4836133	*ZNF608*	5	124,332,103	123,972,610	124,080,865	108,255	NA	NA	NA
rs11847697	*PRKD1*	14	30,515,112	30,045,689	30,396,899	351,210	NA	NA	NA
SNPs Excluded (No gene expression data available)									
rs3810291	*ZC3H4*	19	47,569,003	47,567,447	47,617,009	49,562	NA	NA	NA
rs887912	*FLJ30838*	2	59,302,877	58,747,888	59,290,901	543,013	NA	NA	NA

* There is the other one gene ,"IQCK", upstream of the index SNP but no genes downstream.

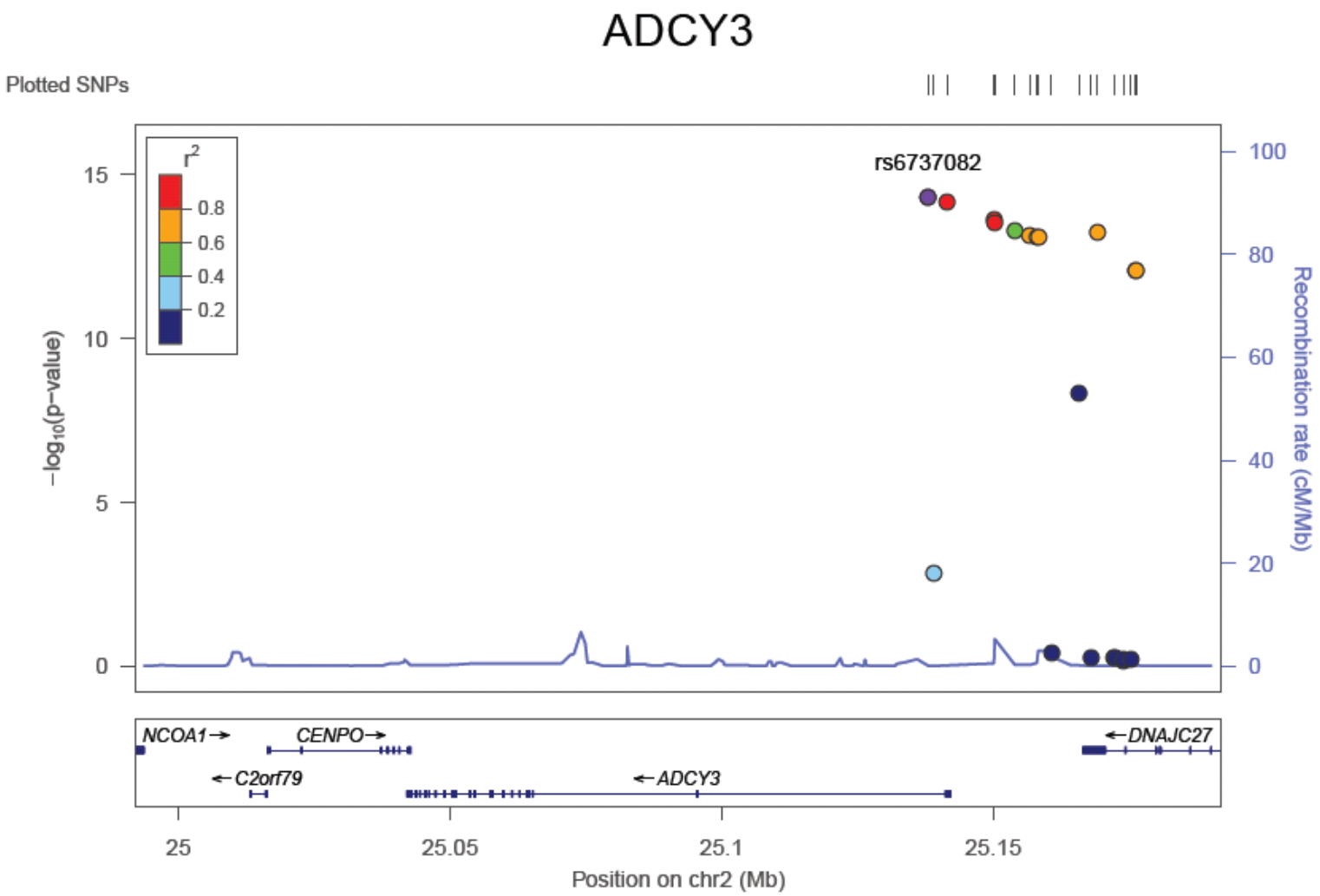
**There is the other one gene ,"LOC730102", downstream of the index SNP but no genes upstream. †Genes were sorted by gene size.

Appendix 4.3. Association of index SNP and BMI in expanded GIANT and GENOA.

Index SNP	Nearest Gene	Function	Chr	SNP Position	Coded/ Non-Coded allele	Expanded GIANT			GENOA		
						Coded Allele Freq.	Beta	P-value	Coded Allele Freq.	Beta	P-value
Novel loci identified by expanded GIANT consortium											
rs1514175	*TNNI3K*	Intron	1	74,991,644	A/G	0.43	0.07	8.16E-14	0.42	0.44	0.0695
rs713586	*ADCY3*	Intergenic	2	25,158,008	C/T	0.47	0.14	6.17E-22	0.50	-0.06	0.7963
rs2890652	*LRP1B*	Intergenic	2	142,959,931	C/T	0.18	0.09	1.35E-10	0.17	-0.18	0.5662
rs13078807	*CADM2*	Intron	3	85,884,150	A/G	0.80	-0.10	3.94E-11	0.82	-0.19	0.5493
rs13107325	*SLC39A8*	Missense	4	103,188,709	C/T	0.93	-0.19	1.50E-13	0.93	-0.33	0.5047
rs2112347	*POC5*	Intergenic	5	75,015,242	G/T	0.37	-0.10	2.17E-13	0.37	-0.32	0.2116
rs206936	*NUDT3*	Intron	6	34,302,869	A/G	0.79	-0.06	3.02E-08	0.81	-0.23	0.3495
rs10968576	*LINGO2*	Intron	9	28,414,339	A/G	0.69	-0.11	2.65E-13	0.68	-0.16	0.5252
rs4929949	*STK33*	Intron	11	8,604,593	C/T	0.52	0.06	2.80E-09	0.53	0.01	0.9804
rs4771122	*MTIF3*	Intron	13	28,020,180	A/G	0.76	-0.09	9.48E-10	0.75	0.16	0.5906
rs2241423	*MAP2K5*	Intron	15	68,086,838	A/G	0.22	-0.13	1.19E-18	0.20	-0.45	0.1446
rs12444979	*GPRC5B*	Intergenic	16	19,933,600	C/T	0.87	0.17	2.91E-21	0.86	-0.19	0.5841
rs2287019	*QPCTL*	Intron	19	46,202,172	C/T	0.80	0.15	1.88E-16	0.80	0.16	0.7032
rs987237	*TFAP2B*	Intron	6	50,803,050	A/G	0.82	-0.13	2.90E-20	0.81	-0.54	0.0797
rs10150332	*NRXN3*	Intron	14	79,936,964	C/T	0.21	0.13	2.75E-11	0.21	0.19	0.5242
Loci that were identified in expanded GIANT consortium and also have been reported in GIANT and deCODE genetics											
rs2815752	*NEGR1*	Intergenic	1	72,812,440	A/G	0.61	0.13	1.16E-22	0.61	0.33	0.1833
rs543874	*SEC16B*	Intergenic	1	177,889,480	A/G	0.81	-0.22	3.56E-23	0.80	**-1.01**	**0.0008**
rs2867125	*TMEM18*	Intergenic	2	622,827	C/T	0.83	0.31	2.77E-49	0.84	**1.09**	**0.0007**
rs9816226	*ETV5*	Intergenic	3	185,834,499	A/T	0.18	-0.14	1.69E-18	0.19	-0.61	0.0535
rs10767664	*BDNF*	Intergenic	11	27,725,986	A/T	0.78	0.19	4.69E-26	0.80	0.49	0.1074
rs3817334	*MTCH2*	Intron	11	47,650,993	C/T	0.59	-0.06	1.59E-12	0.58	-0.23	0.3495
rs7138803	*FAIM2*	Intergenic	12	50,247,468	A/G	0.38	0.12	1.82E-17	0.37	-0.01	0.9809
rs7359397	*SH2B1*	Intergenic	16	28,885,659	C/T	0.60	-0.15	1.88E-20	0.57	-0.21	0.4059
rs1558902	*FTO*	Intron	16	53,803,574	A/T	0.42	0.39	4.80E-120	0.44	**1.01**	**0.00003**
rs29941	*KCTD15*	Intergenic	19	34,309,532	A/G	0.33	0.06	3.01E-09	0.34	-0.38	0.139

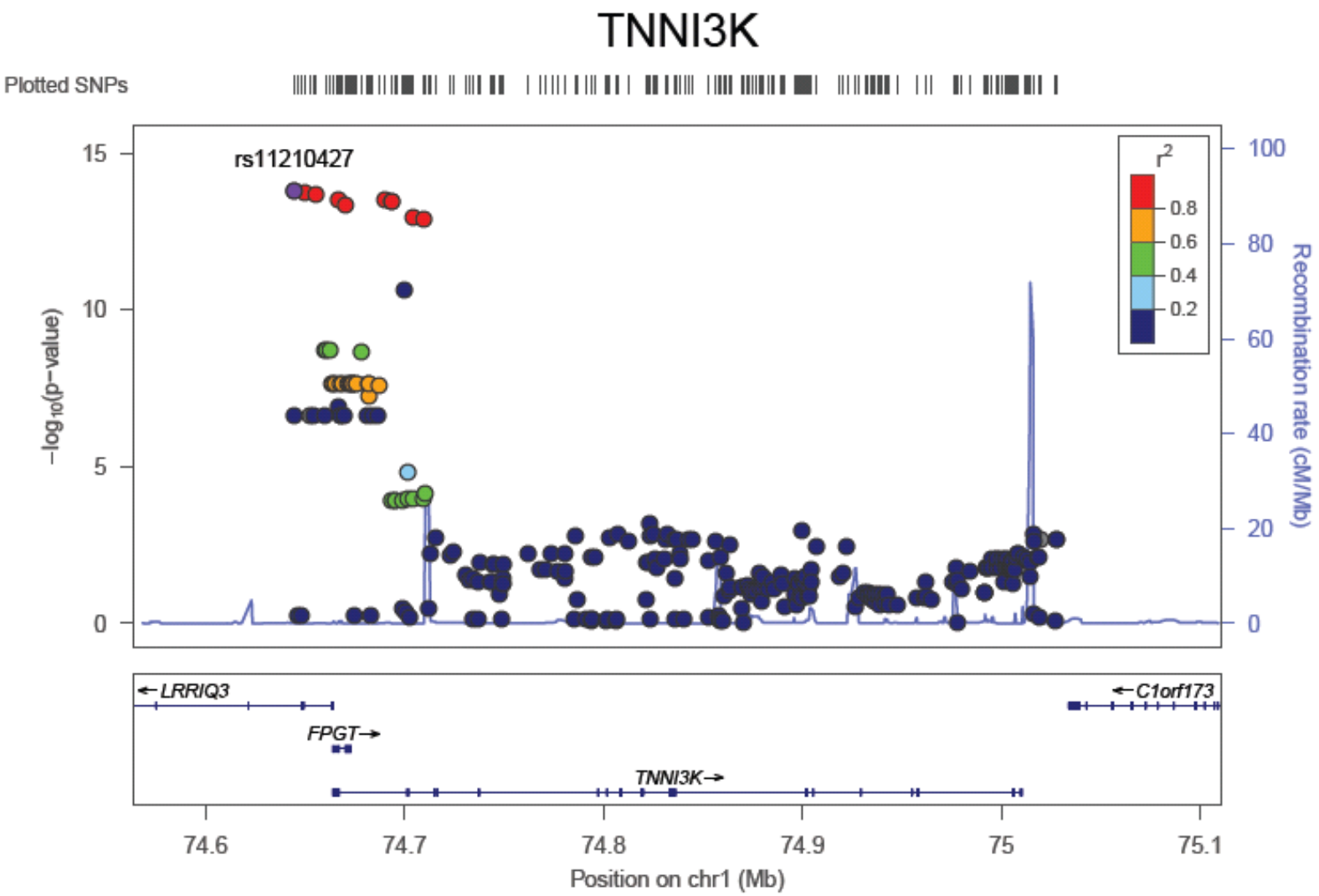
Bold:*p*-value<0.05 in GENOA.

Appendix 4.4. SNPs and their *cis*-effects on *ADCY3* expression.



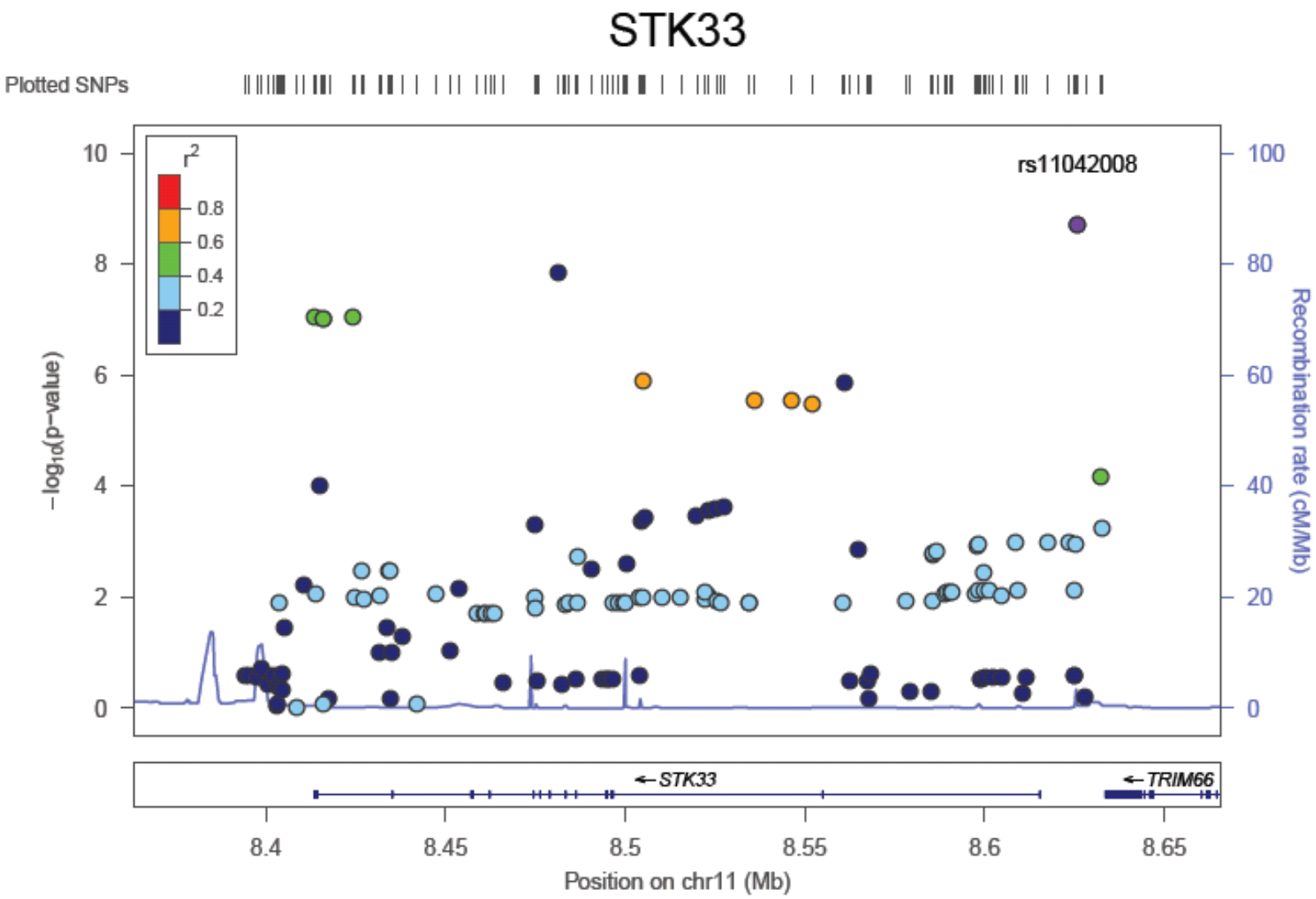
*Purple circle is the SNP with lowest p-value.

Appendix 4.5. SNPs and their *cis*-effects on *TNNI3K* expression.



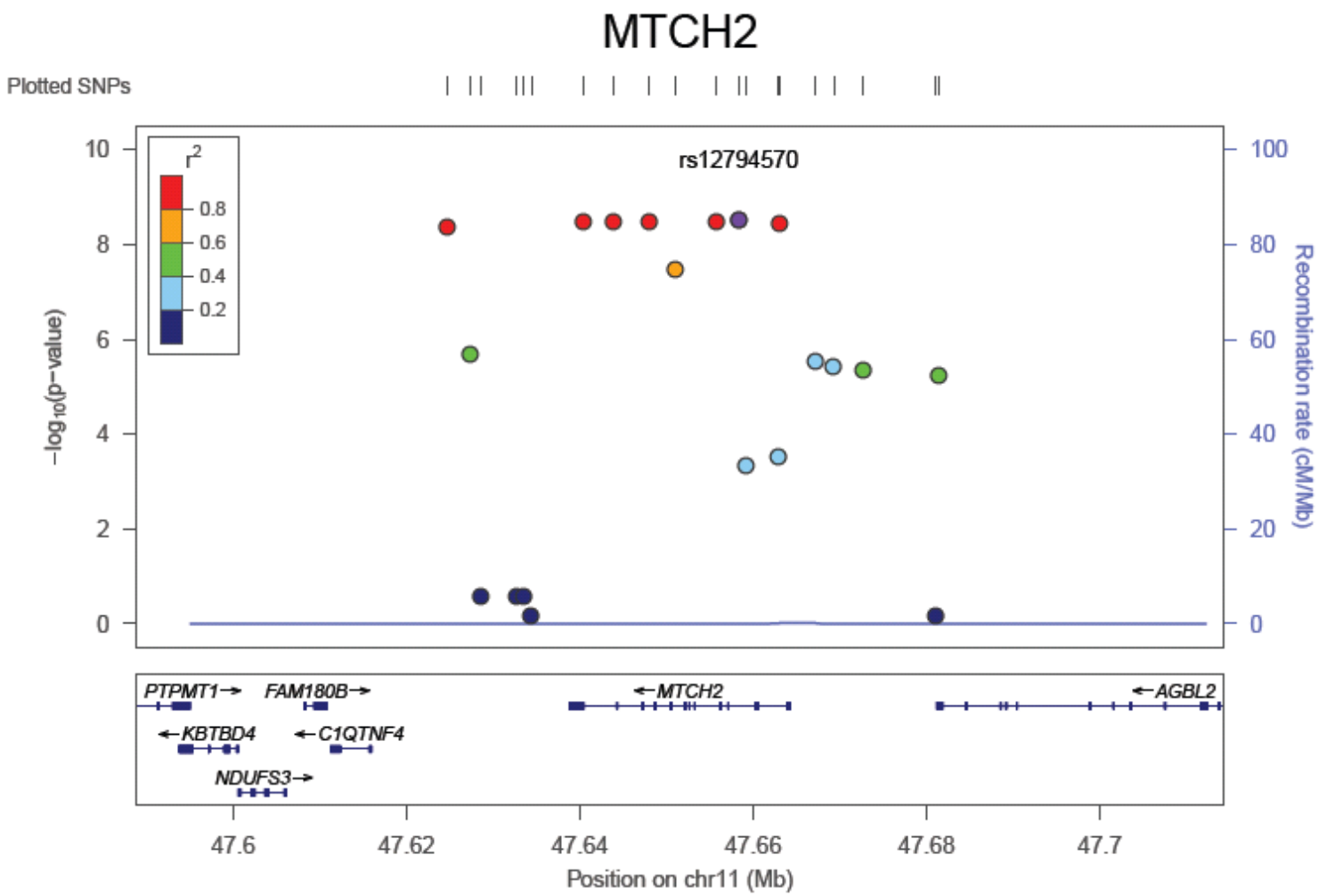
*Purple circle is the SNP with lowest p-value.

Appendix 4.6. SNPs and their *cis*-effects on *STK33* expression.



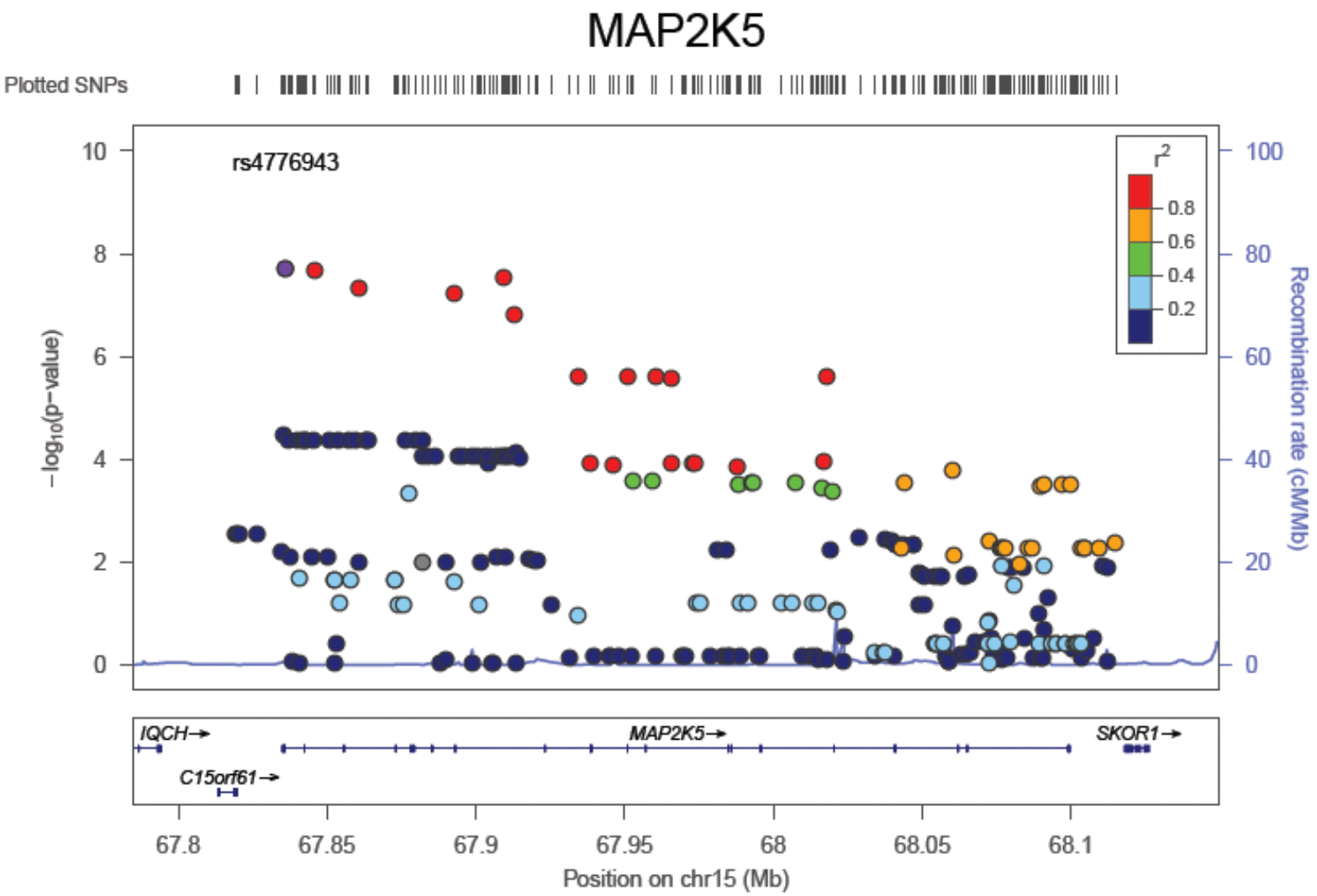
*Purple circle is the SNP with lowest p-value.

Appendix 4.7. SNPs and their *cis*-effects on *MTCH2* expression.



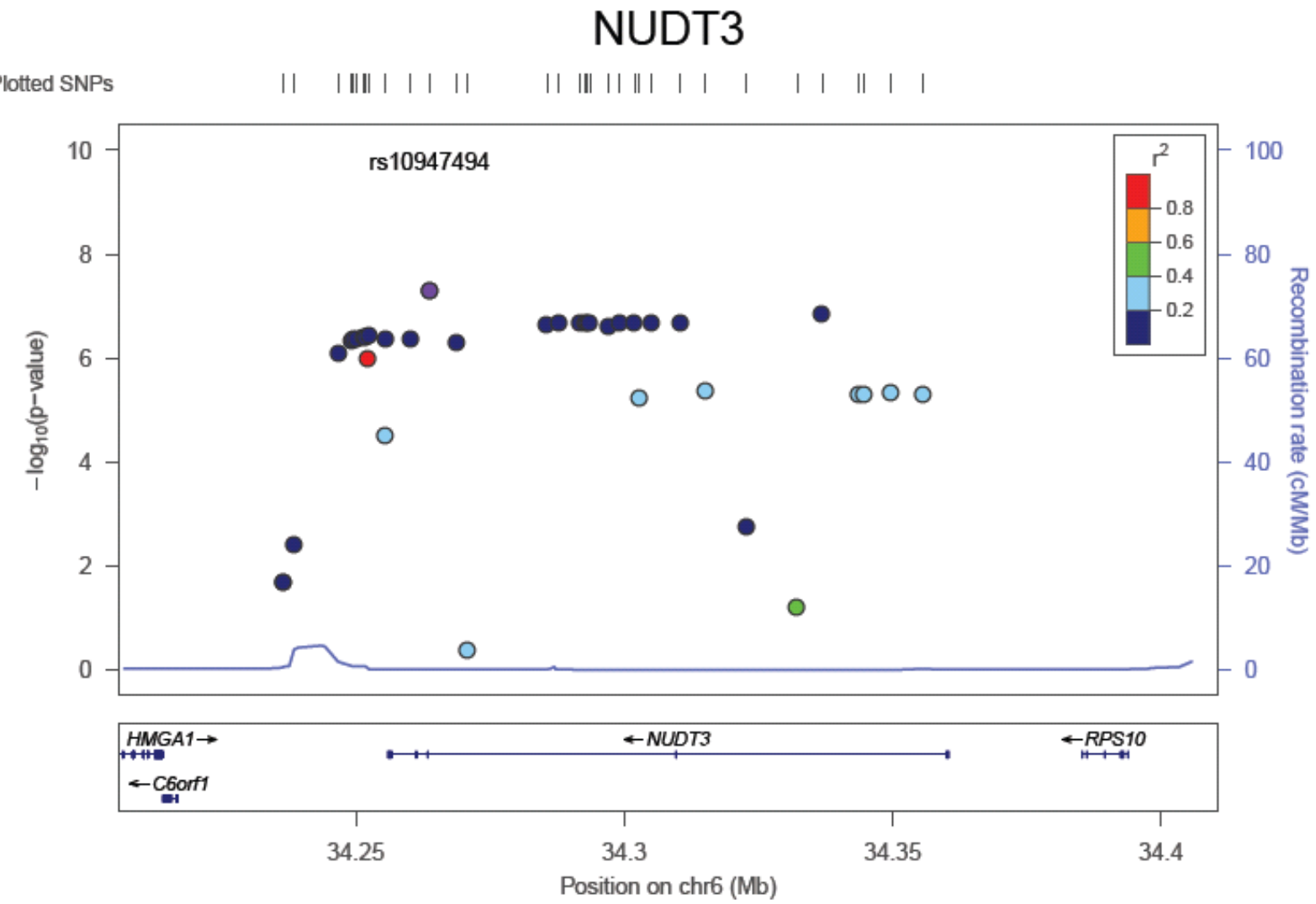
*Purple circle is the SNP with lowest p-value.

Appendix 4.8. SNPs and their *cis*-effects on *MAP2K5* expression.



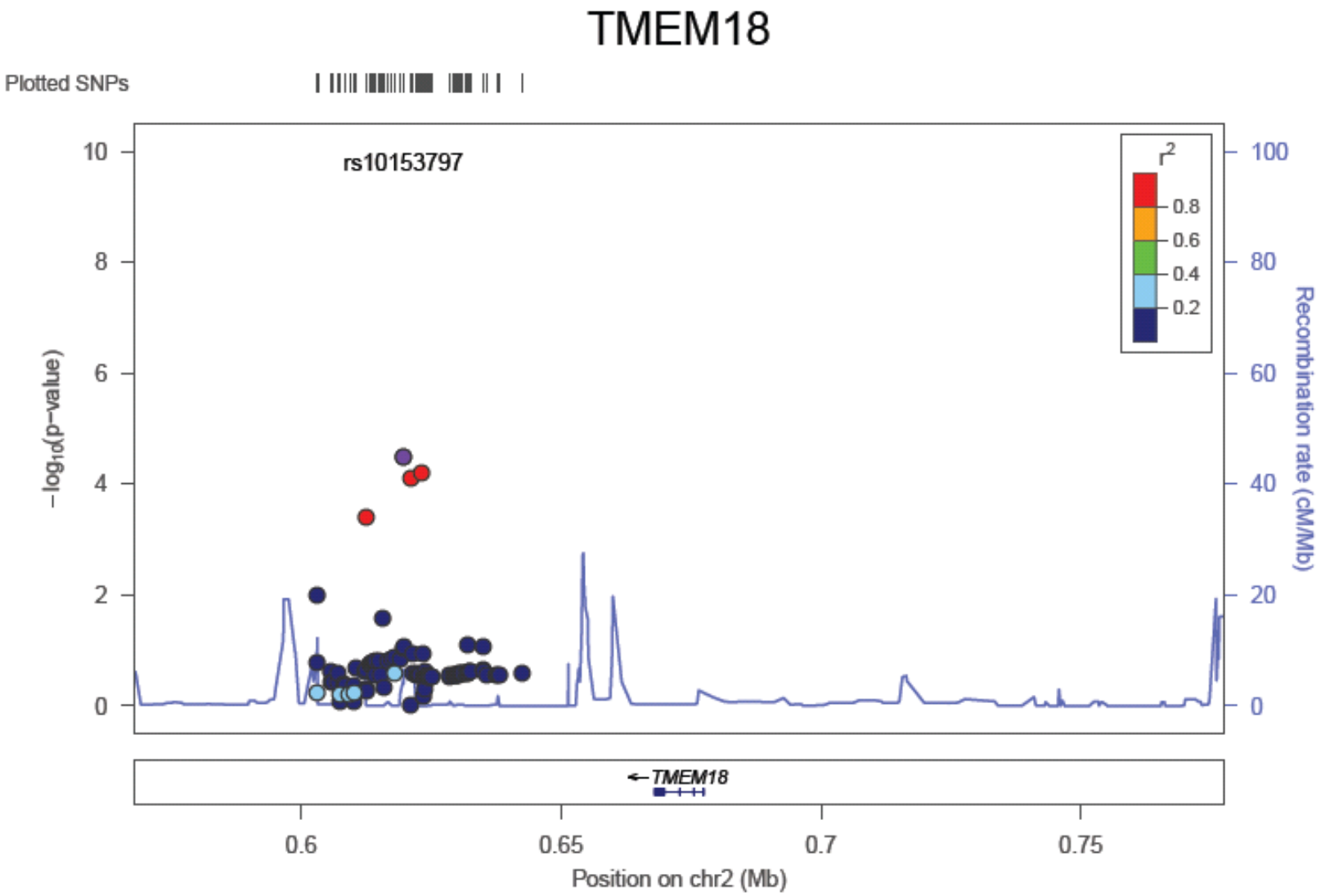
*Purple circle is the SNP with lowest p-value.

Appendix 4.9. SNPs and their *cis*-effects on *NUDT3* expression.



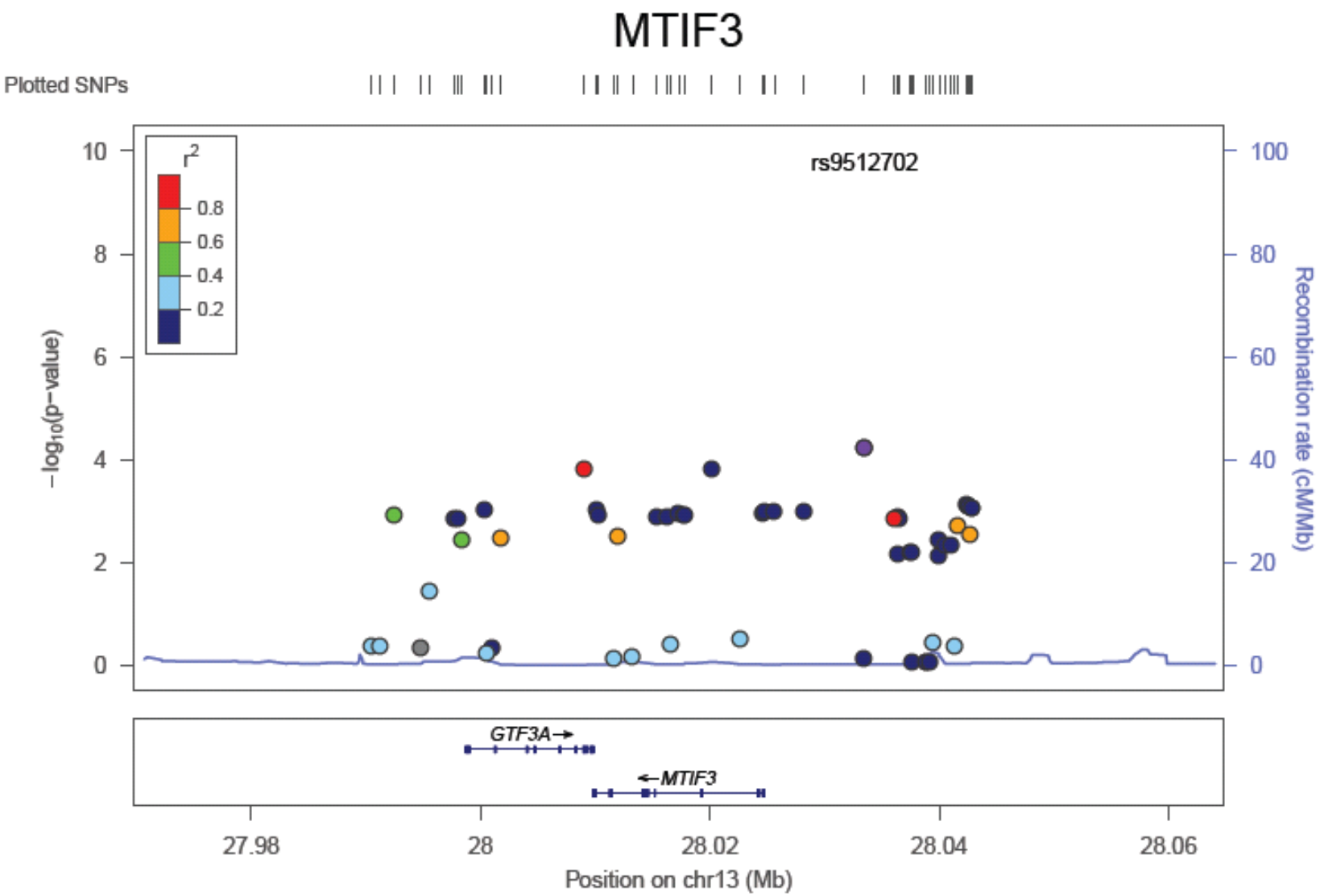
*Purple circle is the SNP with lowest p-value.

Appendix 4.10. SNPs and their *cis*-effects on *TMEM18* expression.



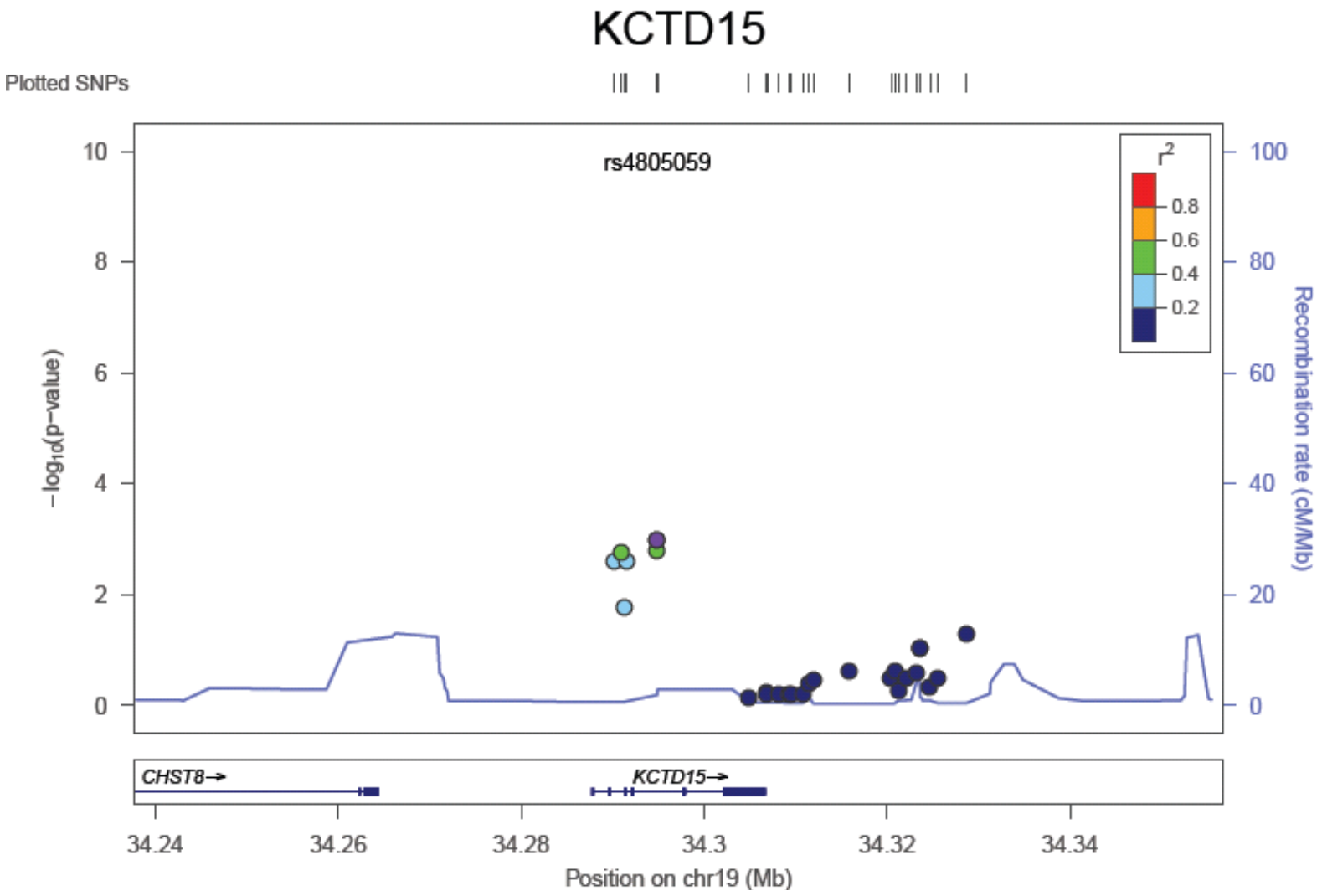
*Purple circle is the SNP with lowest p-value.

Appendix 4.11. SNPs and their *cis*-effects on *MTIF3* expression.



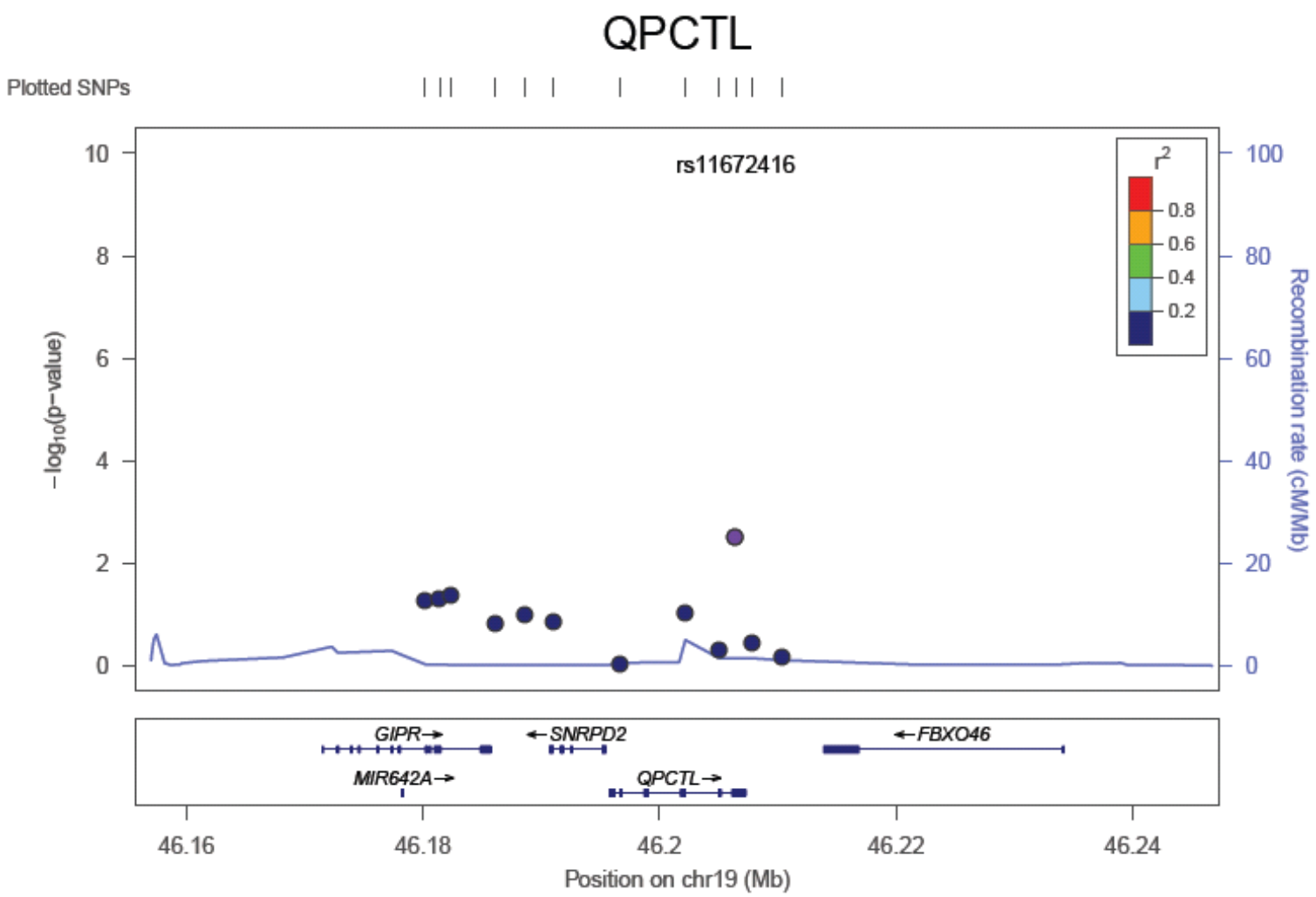
*Purple circle is the SNP with lowest p-value.

Appendix 4.12. SNPs and their *cis*-effects on *KCTD15* expression.



*Purple circle is the SNP with lowest p-value.

Appendix 4.13. SNPs and their *cis*-effects on *QPCTL* expression.



*Purple circle is the SNP with lowest p-value.

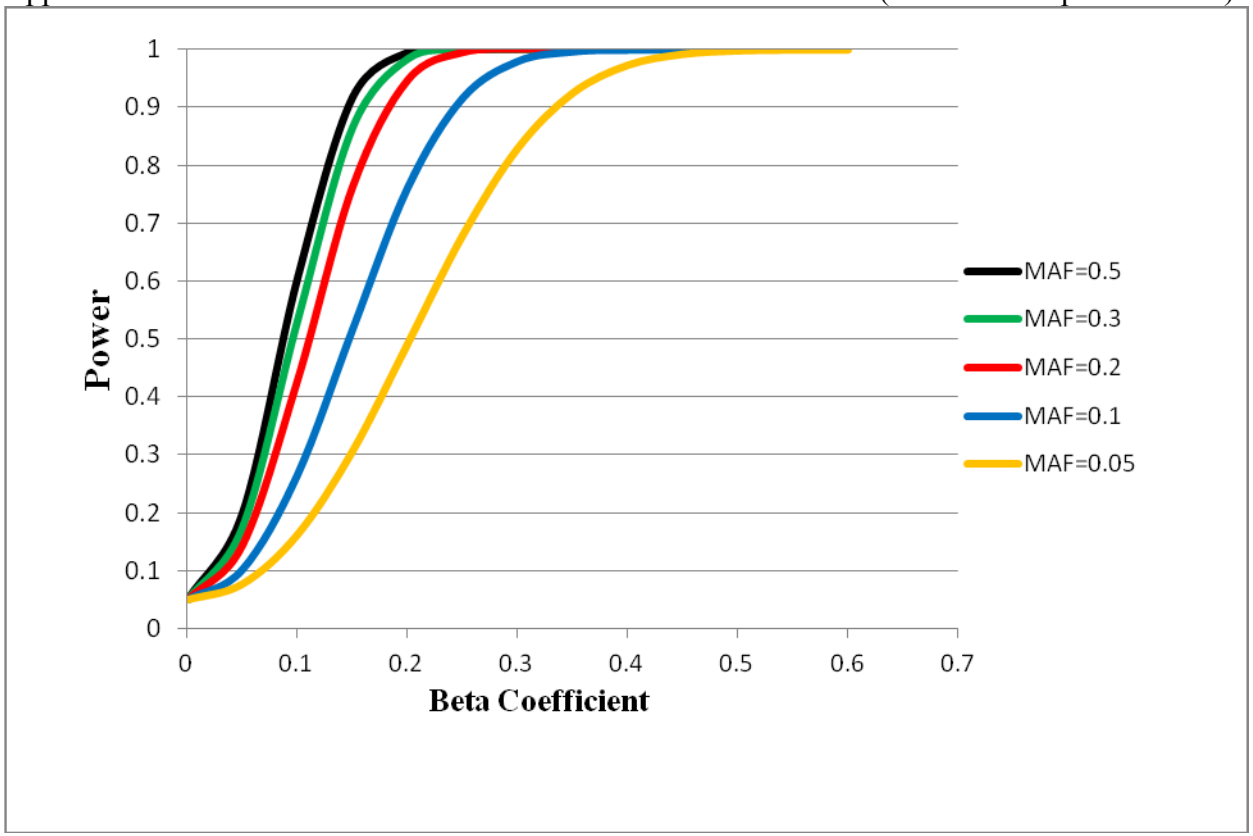
Appendix 4.14. The 33 SNPs associated with BMI and Gene Expression.

Chr	SNP	Gene	Position	Function	Coded/ Non Coded Allele	Coded Allele Freq.	Bonferroni Corrected Alpha-Level	BMI		Gene Expression	
								Beta	P-value	Beta	P-Value
1	rs576802	*TNNI3K*	74724296	Intron	A/T	0.13	1.90E-04	-0.80	0.026	-0.04	0.006
1	rs524252	*TNNI3K*	74732597	Intron	C/T	0.86	1.90E-04	0.70	0.048	0.03	0.045
1	rs569564	*TNNI3K*	74780335	Intron	A/C	0.83	1.90E-04	0.65	0.045	0.03	0.006
1	rs17544747	*TNNI3K*	74898246	Intron	A/T	0.06	1.90E-04	1.06	0.050	0.04	0.050
1	rs17095308	*TNNI3K*	74900212	Intron	A/C	0.95	1.90E-04	-1.21	0.026	-0.04	0.036
1	rs1881029	*TNNI3K*	74901440	Intron	A/G	0.95	1.90E-04	-1.25	0.022	-0.04	0.034
1	rs685233	*SEC16B*	177869941	Intron	A/G	0.68	9.40E-04	-0.63	0.015	0.02	0.038
1	rs623130	*SEC16B*	177871886	Intron	C/T	0.32	9.40E-04	0.63	0.016	-0.02	0.038
1	rs576101	*SEC16B*	177873383	Intron	A/T	0.32	9.40E-04	0.63	0.016	-0.02	0.038
1	rs604388	*SEC16B*	177877979	Intron	C/T	0.32	9.40E-04	0.64	0.015	-0.02	0.037
1	rs617763	*SEC16B*	177885238	Intron	C/T	0.68	9.40E-04	-0.69	0.008	0.02	0.032
1	rs618759	*SEC16B*	177885488	Intron	A/G	0.68	9.40E-04	-0.69	0.008	0.02	0.032
1	rs512898	*SEC16B*	177888429	Intron	C/T	0.32	9.40E-04	0.68	0.009	-0.02	0.034
1	rs570762	*SEC16B*	177890153	Intron	C/T	0.62	9.40E-04	0.52	0.041	0.03	0.001
2	rs7564851	*TMEM18*	615658	Intron	A/G	0.42	5.00E-04	-0.76	0.003	-0.02	0.026
3	rs9863591	*ETV5*	185841624	Intron	C/G	0.76	1.80E-03	0.57	0.041	0.07	0.047
4	rs2165265	*SLC39A8*	103203500	Intron	G/T	0.36	3.80E-04	-0.69	0.007	0.03	0.031
4	rs2119213	*SLC39A8*	103203853	Intron	C/T	0.35	3.80E-04	-0.74	0.004	0.03	0.040
4	rs10489122	*SLC39A8*	103206090	Intron	A/G	0.34	3.80E-04	-0.76	0.003	0.03	0.018
4	rs17032436	*SLC39A8*	103207765	Intron	C/T	0.07	3.80E-04	-1.27	0.029	0.09	0.002
4	rs4698844	*SLC39A8*	103217980	Intron	C/T	0.66	3.80E-04	0.89	0.002	-0.04	0.011
4	rs4698845	*SLC39A8*	103218327	Intron	C/T	0.33	3.80E-04	-0.90	0.002	0.04	0.012
4	rs7699390	*SLC39A8*	103218446	Intron	A/G	0.59	3.80E-04	0.66	0.020	-0.04	0.015
6	rs10947494	*NUDT3*	34263743	Intron	A/G	0.45	3.80E-04	-0.51	0.045	0.05	5.25E-08
14	rs12892370	*NRXN3*	78875607	Intron	C/T	0.08	4.07E-05	-1.13	0.012	0.13	0.030
16	rs12446228	*FTO*	53800387	Intron	A/G	0.36	1.10E-04	-0.66	0.011	-0.01	0.033
16	rs2058908	*FTO*	53806145	Intron	C/T	0.76	1.10E-04	0.76	0.017	0.02	0.026
16	rs1477196	*FTO*	53808258	Intron	A/G	0.36	1.10E-04	-0.65	0.011	-0.01	0.033
16	rs4783819	*FTO*	53816647	Intron	C/G	0.64	1.10E-04	0.64	0.013	0.01	0.035
16	rs3751813	*FTO*	53818708	Intron	G/T	0.42	1.10E-04	-0.87	0.0005	-0.02	0.010
16	rs7190492	*FTO*	53828752	Intron	A/G	0.35	1.10E-04	-0.60	0.020	-0.01	0.034
16	rs7185783	*FTO*	54007822	Intron	G/T	0.89	1.10E-04	-1.02	0.034	-0.03	0.023
16	rs708262	*FTO*	54114529	Intron	G/T	0.73	1.10E-04	0.95	0.010	0.02	0.043

Appendix 4.15. The twenty-Four SNPs that were associated with both BMI and gene expression levels in both sample.

Chr	SNP	Gene	Position	Function	Coded/ Non Coded Allele	Sample with SNP Data (N=1423)			Sample with both SNP and Gene Expreession Data (N=789)				
						BMI			BMI			Gene Expression	
						Coded Allele Freq.	Beta	P-value	Coded Allele Freq.	Beta	P-value	Beta	P-value
1	rs576802	*TNNI3K*	74724296	Intron	A/T	0.13	-0.80	0.026	0.14	-1.05	0.039	-0.04	0.006
1	rs17095308	*TNNI3K*	74900212	Intron	A/C	0.95	-1.21	0.026	0.94	-1.60	0.038	-0.04	0.036
1	rs1881029	*TNNI3K*	74901440	Intron	A/G	0.95	-1.25	0.022	0.94	-1.65	0.032	-0.04	0.034
1	rs685233	*SEC16B*	177869941	Intron	A/G	0.68	-0.63	0.015	0.67	-0.83	0.023	0.02	0.038
1	rs623130	*SEC16B*	177871886	Intron	C/T	0.32	0.63	0.016	0.33	0.83	0.023	-0.02	0.038
1	rs576101	*SEC16B*	177873383	Intron	A/T	0.32	0.63	0.016	0.33	0.83	0.023	-0.02	0.038
1	rs604388	*SEC16B*	177877979	Intron	C/T	0.32	0.64	0.015	0.33	0.84	0.022	-0.02	0.037
1	rs617763	*SEC16B*	177885238	Intron	C/T	0.68	-0.69	0.008	0.67	-0.91	0.012	0.02	0.032
1	rs618759	*SEC16B*	177885488	Intron	A/G	0.68	-0.69	0.008	0.67	-0.91	0.012	0.02	0.032
1	rs512898	*SEC16B*	177888429	Intron	C/T	0.32	0.68	0.009	0.33	0.89	0.014	-0.02	0.034
2	rs7564851	*TMEM18*	615658	Intron	A/G	0.42	-0.76	0.003	0.43	-1.00	0.005	-0.02	0.026
4	rs2165265	*SLC39A8*	103203500	Intron	G/T	0.36	-0.69	0.007	0.37	-0.91	0.010	0.03	0.031
4	rs2119213	*SLC39A8*	103203853	Intron	C/T	0.35	-0.74	0.004	0.36	-0.97	0.006	0.03	0.040
4	rs10489122	*SLC39A8*	103206090	Intron	A/G	0.34	-0.76	0.003	0.35	-1.01	0.005	0.03	0.018
4	rs4698844	*SLC39A8*	103217980	Intron	C/T	0.66	0.89	0.002	0.65	1.18	0.002	-0.04	0.011
4	rs4698845	*SLC39A8*	103218327	Intron	C/T	0.33	-0.90	0.002	0.34	-1.19	0.002	0.04	0.012
4	rs7699390	*SLC39A8*	103218446	Intron	A/G	0.59	0.66	0.020	0.58	0.88	0.030	-0.04	0.015
16	rs12446228	*FTO*	53800387	Intron	A/G	0.36	-0.66	0.011	0.37	-0.87	0.016	-0.01	0.033
16	rs2058908	*FTO*	53806145	Intron	C/T	0.76	0.76	0.017	0.75	1.00	0.025	0.02	0.026
16	rs1477196	*FTO*	53808258	Intron	A/G	0.36	-0.65	0.011	0.37	-0.86	0.017	-0.01	0.033
16	rs4783819	*FTO*	53816647	Intron	C/G	0.64	0.64	0.013	0.63	0.85	0.019	0.01	0.035
16	rs3751813	*FTO*	53818708	Intron	G/T	0.42	-0.87	0.0005	0.43	-1.14	0.001	-0.02	0.010
16	rs7190492	*FTO*	53828752	Intron	A/G	0.35	-0.60	0.020	0.36	-0.80	0.029	-0.01	0.034
16	rs708262	*FTO*	54114529	Intron	G/T	0.73	0.95	0.010	0.72	1.26	0.015	0.02	0.043

Appendix 4.16. Power calculation for association of SNPs and BMI (effective sample size=989).



*Power calculation was estimated by using standardized phenotype with mean of zero and standard deviation of one.

Appendix 4.17.1. Potential function for the 30 BMI-related genes.

Gene	Chr	Gene Name	Function
NEGR1	1	Neuronal growth regulator 1	*NEGR1* is a member of the immunoglobulin superfamily and it participates in the regulation of neurite outgrowth in the developing brain.[91]
SEC16B	1	SEC16 homolog B	*SEC16B* is a mammalian homolog of S. cerevisiae Sec16 that is required for organization of transitional endoplasmic reticulum sites and protein export.[92]
TNNI3K	1	TNNI3 interacting kinase	This gene encodes a protein that belongs to the MAP kinase kinase kinase family of protein kinases. The protein contains ankyrin repeat, protein kinase and serine-rich domains and is thought to play a role in cardiac physiology.[77]
ADCY3	2	Adenylate cyclase 3	This gene encodes adenylyl cyclase 3 which is a membrane-associated enzyme and catalyzes the formation of the secondary messenger cyclic adenosine monophosphate (cAMP). This protein appears to be widely expressed in various human tissues and may be involved in a number of physiological and pathophysiological metabolic processes.[77]
DNAJC27	2	DnaJ (Hsp40) homolog, subfamily C, member 27	The exact function of *DNAJC27* remains unknown.[77]
LRP1B	2	Low density lipoprotein receptor-related protein 1B	*LRP1B* belongs to the low density lipoprotein eceptor gene family. These receptors play a wide variety of roles in normal cell function and development due to their interactions with multiple ligands.[77]
TMEM18	2	Transmembrane protein 18	*TMEM18* is crucial for regulating general cell motility at its basal, physiologic level expression.[77]
CADM2	3	Cell adhesion molecule 2	This gene encodes a member of the synaptic cell adhesion molecule 1 (SynCAM) family which belongs to the immunoglobulin (Ig) superfamily.[77]
DGKG	3	Diacylglycerol kinase, gamma	This gene encodes an enzyme that is a member of the type I subfamily of diacylglycerol kinases, which are involved in lipid metabolism. These enzymes generate phosphatidic acid by catalyzing the phosphorylation of diacylglycerol, a fundamental lipid second messenger that activates numerous proteins, including protein kinase C isoforms.[77]
ETV5	3	Ets variant 5	*ETV5* is a transcription factor that plays a role in development and cancer. It is widely expressed, predominantly in brain and placenta, and to a lesser degree in lung, pancreas, and heart.[93]
SLC39A8	4	Solute carrier family 39 (zinc transporter), member 8	This gene encodes a member of the *SLC39* family of solute-carrier genes, which show structural characteristics of zinc transporters and it also transports cadmium and manganese. The encoded protein is glycosylated and found in the plasma membrane and mitochondria, and functions in the cellular import of zinc at the onset of inflammation.[77]

Appendix 4.17.2. Potential function for the 30 BMI-related genes.

Gene	Chr	Gene Name	Function
POC5	5	POC5 centriolar protein homolog (Chlamydomonas)	Essential for the assembly of the distal half of centrioles, required for centriole elongation.[77]
NUDT3	6	Nudix (nucleoside diphosphate linked moiety X)-type motif 3	Nudix proteins act as homeostatic checkpoints at important stages in nucleoside phosphate metabolic pathways.[77]
TFAP2B	6	Transcription factor AP-2 beta (activating enhancer binding protein 2 beta)	This gene encodes a member of the AP-2 family of transcription factors. Specific AP-2 family members differ in their expression patterns and binding affinity for different promoters. This protein functions as both a transcriptional activator and repressor.[77]
LINGO2	9	Leucine rich repeat and Ig domain containing 2	*LINGO2* have been reported as risk factors for developing Parkinson's disease in some Caucasian populations.[94]
BDNF	11	Brain-derived neurotrophic factor	The protein encoded by this gene is a member of the nerve growth factor family. This gene may play a role in the regulation of stress response and in the biology of mood disorders.
MTCH2	11	Mitochondrial carrier 2	*MTCH2* is highly expressed in human adipose tissue with increased levels in obese women. It suggested that *MTCH2* may play a role in cellular processes underlying obesity.[77]
STK33	11	Serine/threonine kinase 33	*STK33* contains one protein kinase domain and functions as a Ser/Thr protein kinase with a possible role in spermatogenesis.[77]
BCDIN3D	12	BCDIN3 domain containing	*BCDIN3D* is a relatively unknown gene with suggested methyltransferase activity.[95]
FAIM2	12	Fas apoptotic inhibitory molecule 2	*FAIM2* protects against Fas mediated apoptosis and is widely expressed with relatively higher expression observed in the hippocampus of the rat.[77]
MTIF3	13	Mitochondrial translational initiation factor 3	This gene encodes a translation initiation factor that is involved in mitochondrial protein synthesis. Polymorphism in this gene is associated with the onset of Parkinson's disease.[77]
NRXN3	14	Neurexin 3	Neurexins are a family of proteins that function in the vertebrate nervous system as cell adhesion molecules and receptors.[77]
MAP2K5	15	Mitogen-activated protein kinase kinase 5	The protein encoded by this gene is a dual specificity protein kinase that belongs to the *MAP* kinase kinase family. The signal cascade mediated by this kinase is involved in growth factor stimulated cell proliferation and muscle cell differentiation.[77]
ATP2A1	16	ATPase, Ca++ transporting, cardiac muscle, fast twitch 1	ATP2A1 is an ATPase which pumps Ca ions from the cytosol into the sarcoplasmic reticulum in order to maintain calcium stores needed for muscle excitation and contraction.[96]

Appendix 4.17.3. Potential function for the 30 BMI-related genes.

Gene	Chr	Gene Name	Function
FTO	16	Fat mass and obesity associated	*FTO* was shown to share sequence motifs with iron- and 2-oxoglutarate (2OG)-dependent oxygenases which are involved in a wide range of processes including fatty acid metabolism, DNA repair and posttranslational modifications. Mouse studies indicate that Fto is highly expressed in the hypothalamic nuclei, which governs energy balance, and that mRNA levels in the arcuate nucleus are regulated by feeding and fasting.[33]
GPRC5B	16	G protein-coupled receptor, family C, group 5, member B	The protein encoded by this gene is a member of the type 3 G protein-coupled receptor family.[77]
SH2B1	16	SH2B adaptor protein 1	This gene encodes a member of the SH2-domain containing mediators family. The encoded protein mediates activation of various kinases and may function in cytokine and growth factor receptor signaling and cellular transformation.[77]
CHST8	19	Carbohydrate (N-acetylgalactosamine 4-0) sulfotransferase 8	*CHST8*, carry out sulfations of carbohydrates which confer highly specific functions on glycoproteins, glycolipids, and proteoglycans and are critical for cell-cell interactions, signal transduction, and embryonic development. CHST8 is expressed in the pituitary where it sulfates the pituitary glycoprotein hormones.[97]
KCTD15	19	Potassium channel tetramerisation domain containing 15	*KCTD15* is a voltage gated potassium channel which has a BTB/POZ domain and although its function has not been elucidated it may be a transcription factor from sequence homology comparisons.[98]
QPCTL	19	Glutaminyl-peptide cyclotransferase-like	It belongs to the glutaminyl-peptide cyclotransferase family. The exact function of *QPCTL* remains unknown.[77]

Chapter 5

The Modifying Effects of Single Nucleotide Polymorphisms on the Association of Gene Expression Levels with Blood Pressure and Body Mass Index

Introduction

Hypertension and obesity affect about one-third of adults in United States and constitute two major risk factors for the nation's leading cause of death, heart disease.[4] Previous heritability studies have shown a substantial contribution (25% to 60%) of additive genetic factors to the inter-individual variation in blood pressure and BMI.[5-7] Given the availability of low-cost, high-throughput technologies, hundreds of thousands of DNA markers can be genotyped in a feasible way, and this enables researchers to examine the associations between genetic variants and diseases on a genome-wide scale. Recent meta-analysis of GWAS studies have successfully identified 29 and 32 genetic variants associated with blood pressure[16] and BMI[9], respectively. However, these variants only explained less than 2% of the variation in blood pressure (0.9%) and BMI (1.45%). In addition, the influences of genetic variants on phenotypic traits are often dependent on the context of environmental conditions or other genetic factors.[57,99] For example, Hines et al reported that the *ADH3* genotype modified the effect of alcohol consumption on the risk of myocardial infarction (MI). As the level of alcohol consumption increases (from low to moderate), men with *γ2 γ2* genotype have substantially decreased risk of MI compared to men with *γ1 γ1* genotype.[99] Recent evidence beyond the aforementioned genetic associations has

suggested that hypertension and obesity are complex traits which result from multiple genetic and environmental risk factors and the complex interactions between them.[57,99,100] Therefore, examining the influence of gene-gene interaction and gene-environment interaction on the phenotypic traits will allow us to better understand the underlying biological mechanisms and the extra phenotypic variation explained by these interactions.

Genetic variants can influence phenotypic traits by altering gene expression (i.e. protein quantity) or by altering the structure of encoded proteins. Mutation or polymorphism in the promoter regions or splicing sites have influence on the regulation of gene expression levels (Appendix 5.1). In addition, mutation or polymorphism at the coding regions (i.e. exons) could alter amino acid sequences and thus gene product activities (i.e. enzyme and receptor) (Appendix 5.1). The level of gene product activity might have an impact on the association between gene expression levels and phenotypic traits. For example, when the gene expression level is high and the level of gene product activity is good, their combined effects on the phenotypic traits would be high. When the gene expression level is high and the level of gene product activity is medium, their combined effects on the phenotypic traits would be medium. When the gene expression level is high, and the level of gene product activity is poor, their combined effects on the phenotypic traits would be low. Similarly, when the gene expression level is medium, its interaction with gene product activity would have medium, low, and very low impact on phenotypic traits depending on the level of gene product activity ranged from good, medium to low, respectively (Appendix 5.2). Because the association between gene expression levels and phenotypic traits depend on the level of gene product activity, and gene product activity is influenced by the mutation or polymorphism at the coding regions. Therefore, the goal of this study is to examine whether SNPs modify the association between gene expression levels and

three outcomes of interest, systolic blood pressure (SBP), diastolic blood pressure (DBP), and BMI. To achieve this goal, we developed a two stage approach that initially evaluates the marginal effects of SNPs and then investigates their interaction with gene expression level on blood pressure and BMI regardless of their marginal effects. This study will allow us to unravel the marginal and context-dependent genetic contributions to the inter-individual variation in blood pressure and BMI.

Methods

Study Population

The study population for this dissertation consists of non-Hispanic, white individuals from Rochester, Minnesota (MN) that were initially enrolled in the GENOA study. GENOA is a multicenter, community-based study of hypertensive sibships collected to identify genes influencing blood pressure levels and development of target organ damage due to hypertension.[41] In Rochester, MN, the Mayo Clinic diagnostic index and medical record linkage system of the Rochester Epidemiology Project were used to identify non-Hispanic white residents of Olmsted County.[43] In the first phase (Phase I:1995-2000) of GENOA (N=1583 non-Hispanic whites), sibships containing at least two individuals with clinically diagnosed essential hypertension before age 60 were invited to participate, including both hypertensive and normotensive siblings. Participants were considered to have hypertension if they had either 1) a previous clinical diagnosis of hypertension by a physician with current anti-hypertensive treatment, or 2) an average systolic blood pressure (SBP) ≥140 mm Hg or diastolic blood pressure (DBP) ≥90 mm Hg on the second and third blood pressure readings. Exclusion criteria were secondary hypertension, pregnancy, and insulin-dependent diabetes mellitus. There were 1583 people with phenotype data, and we removed people who had secondary hypertension (N=9), had gender

reassignment (N=1), or lacked SNP data (N=138). Participants who did not self report as non-Hispanic whites (N=5) or were outliers (> mean + 4 standard deviation or < mean - 4 standard deviation) for the outcome (SBP, DBP, and BMI) (N=7) were also excluded. 1423 people with phenotype and genotype data were used for genetic association analysis. We then merged this sample with gene expression data and 789 people were retained in the sample for gene expression analysis. The GENOA study was approved by the institutional review board of the Mayo Clinic, Rochester, MN. Written informed consent was obtained from all participants.

Phenotype Measurement

Blood was drawn by venipuncture after an overnight fast of at least eight hours. Resting SBP and DBP were measured with random zero sphygmomanometers (Hawksley and Sons, West Sussex, England) and cuffs appropriate for arm size. Three readings were taken in the right arm after the participant rested in the sitting position for at least five minutes according to standards set by The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High BP (JNC-7) guidelines.[46] Of the three readings, the last two readings were averaged for the analyses. Height was measured by wall stadiometer, weight by electronic balance, and BMI was calculated by definition as weight in kilograms divided by height in meters squared.

Genotyping and Quality Control of SNPs

In GENOA, a majority of non-Hispanic whites (N=1386) have been successfully genotyped on the Affymetrix® Genome-Wide Human SNP Array 6.0. There were 123 subjects whose genotypes were not successfully measured on the Affymetrix 6.0 platform, so they were re-genotyped using the Illumina Human 1M-DUO BeadChip.

Genotyping was carried out at the Mayo Clinic in Rochester, Minnesota. Briefly, 500ng

genomic DNA at 50ng/ul in low ethylene-diamine-tetraacetic acid (EDTA)-Tris buffer was digested in two separate reaction mixtures using the appropriate restriction enzyme (StyI and NspI, 250ng of DNA for each mixture). This step was followed by ligation of an adaptor sequence containing a universal primer sequence. Samples were then subjected to polymerase chain reaction (PCR) (four PCR reactions per sample for the NspI mixture and three for StyI) with conditions designed to amplify 200-2,000 base pairs. The seven PCR products were then combined with Agencourt Ampure beads, passed over an E & K Scientific filter plate, and eluted with elution buffer. Agarose gel analysis of the PCR products and quantification of the amount of PCR product was performed. PCR product concentration were confirmed to be at least 5ug DNA in 1ul EB buffer. Product was then fragmented with DNase I and an agarose gel analysis of the fragmented DNA was used to confirm this step. Following fragmentation, DNA was labeled with Terminal Deoxynucleotidyl Transferase (TdT), hybridized to the appropriate GeneChip, and incubated overnight. The chip was stained and washed on the Affymetrix 450 Fluidics station and then scanned on the Affymetrix 3000 GeneChip scanner.

Preliminary SNP genotype calls were generated using the Dynamic Model (DM) algorithm.[58] The final SNP genotype calls were generated by Birdseed, an algorithm designed especially for the Affymetrix® Genome-Wide Human SNP Array 6.0, and based on the robust linear model with Mahalanobis distance classifier algorithm (RLMM).[59] In order to call genotypes while best accounting for experimental variability and population-specific allele frequencies, Birdseed utilizes information about variation across samples to modify pre-computed genotype calling models from Affymetrix for each SNP probe set. Birdseed has been shown to reduce the bias against heterozygous calls and boost call rates to over 99% while simultaneously increasing concordance rates.[60]

To obtain the final dataset, the following quality control (QC) thresholds were applied: sample call rates >95% and SNP call rates >95%. These thresholds are comparable to those used in the published GWAS.[8,9]

Imputation and Quality Control

Because only a portion of SNPs of interest have been genotyped, imputation methods were used to infer missing or untyped SNP genotypes based on known haplotype information from HapMap.[61] HapMap CEU (release 22 build 36) was used as a reference panel containing 30 trios that were collected from Utah residents with northern and western European descent by the Centre d'Etude du Polymorphisme Humain (CEPH).

Imputation was performed by using MACH v 1.0.16 which implements a Markov Chain based algorithm[62] to infer possible pairs of haplotypes for each individual's genotypes up to ~2.5 million SNPs. SNP imputation combined genotype data from the non-Hispanic white sample in Rochester, MN with the HapMap CEU samples and then inferred genotypes probabilistically according to shared haplotype stretches between study samples and HapMap. Imputation results are summarized as an "allele dosage" defined as the expected number of copies of the alphabetically higher allele at the SNP (a fractional value between 0.0 and 2.0) for each genotype. The imputation accuracy rates of MACH have been shown to be similar to IMPUTE[63] and higher than several other imputation methods (e.g. fast PHASE, Beagle, and PLINK).[64] Quality control thresholds were applied as follows: SNPs with imputation quality score Rsq>0.3, and Hardy-Weiberg equilibrium p-value $>10^{-3}$ (tested in the GENOA unrelated sample).

Principal component analysis was conducted as another quality control step to identify and remove samples with outlying genotype profiles. First, we removed SNPs that had moderate to

poor imputation quality as measured by the estimated R^2 between imputed and true genotypes (R^2<0.8) from MACH output. Because GENOA includes sibships, we obtained an unrelated sample of individuals (N=570) by randomly selecting one individual from each sibship. The first ten principal components (PCs) were then calculated on the set of SNPs that were common in both Affymetrix and Illumina platforms and were also available in HapMap (N=226,619 SNPs). The imputed genotypes were used to ensure no missing values for SNPs. An additive model was assumed for the SNPs with standardization of mean equals 0 and variance equals 1. We then used the loading matrix for these PCs to calculate the PC values in the full sample. Individuals with outlier values more than 6 standard deviations on any of the 10 PCs (N=45) were removed from the analysis sample.

Gene Expression Assessment and Quality Control

Blood samples for beta-lymphocyte extraction were collected during GENOA Phase I (1995-2000) study and GENOA Phase II (2000-2005) study. Beta-lymphocytes were isolated and transformed using Epstein-Barr virus at the time of collection. For the present study, individuals who participated in an ancillary GENOA study: Genetics of Microangiopathic Brain Injury Study (GMBI) had their lymphocytes re-initiated for gene expression assessment. Specifically, cell lines of 237 subjects and 552 subjects were collected and transformed during 1995-2000 and 2001-2005, respectively. Immortalized lymphocytes were stored at -180°C in a freezing medium containing 10% dimethyl sulfoxide (DMSO) as a cryoprotectant. To re-initiate the lymphocytes, a vial of cells is retrieved from cryostorage, thawed at 37°C, rinsed twice in Roswell Park Memorial Institute (RPMI) 1640 culture medium (Irvine Scientific, Santa Ana CA), and transferred to a culture flask along with fresh culture medium supplemented with 15% fetal bovine serum (Irvine Scientific, Santa Ana CA). All cultured lymphocytes were maintained

in the standardized culture medium, supplemented with 1% penicillin/streptomycin/L-glutamine (Sigma-Aldrich, St. Louis MO), at 37°C and 90% humidity in 5% CO2 for two to three weeks with regular changes of culture medium until a cell density of 5 x 10^5 cells/mL was achieved.

RNA samples were extracted using standard protocols. RNA quality was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Foster City, CA) and quantified by spectrophotometry using the Nanodrop ND-1000 (Nanodrop Inc., Wilmington, DE). All RNA samples used in the present study yielded both an A260/A280 absorbance ratio greater than 2.0 and a RNA Integrity Number (RIN) ≥ 8. One μg of RNA was labeled using the WT Expression labeling assay (Applied Biosystems/ Ambion, Foster City, CA) including the labeling controls from the GeneChip Eukaryotic Poly-A RNA Control Kit (Affymetrix, Santa Clara, CA). Each step of the labeling protocol was monitored using the Agilent 2100 Bioanalyzer or the Nanodrop spectrophotometer, as recommended by the manufacturer. Hybridization buffer, Eukaryotic Hybridization Controls, and OligoB2 controls were added to the cDNA fragments just prior to hybridization to the Affymetrix Human Exon 1.0 ST Array. Hybridization was performed at 45°C for 17 hours. Following hybridization, the chips were washed and stained with a phycoerythrin-strepavidin conjugate and were scanned at an excitation wavelength of 488 nm.

Array quality control was performed using Affymetrix Expression Console™ (v 1.1) at the transcript level using core-level probe sets. All array images passed visual inspection. Hybridization controls were all present with signal increases following concentration. Labeling control signal strengths followed the order Lys < Phe < Thr < Dap. Signal intensity plots were examined for raw and processed data to identify outliers.

Raw intensity data were processed using the Affymetrix Power Tool software. Probe summarization and probe set normalization were performed using Robust Multi-array Analysis

(RMA), which included background correction, quantile normalization, log2-transformation, and probe set summarization.[47] Only core probesets were used to assess exon-level expression. Gene-level expression was assessed by averaging all the core probe sets for that gene.

Quality control of exon array data were done by examining the QC matrix to identify possible outliers or non-performing samples using the Partek Genomic Suite software (V.6.6). Analyses were restricted to core probe sets. Probe sets which are known to cross-hybridize and those with undetectable expression were also excluded.

Criteria to select SNPs, Genes, and Gene Expression Levels of Interest

Because only some of the index SNPs identified from previous GWAS fall within gene boundaries, we used four strategies to choose the set of SNPs and gene expression data to investigate the specific aims based on the location of index SNPs. First, if the index SNP is inside of a gene, then all of the SNPs that are in a gene and ± 20 kb on either side of that specific gene were selected. Gene expression data for that specific gene was also selected. Second, if the index SNP is outside of a gene and the distance between the index SNP and the closest gene is within 100 kb, then SNPs that are within ± 20 kb of the index SNPs were selected. As for the gene expression data, the closest gene upstream and the closest gene downstream of the index SNP were selected. Third, in gene sparse regions, SNPs that are within ± 20 kb of the index SNPs were selected. As for the gene expression data, there is typically only one gene either upstream or downstream within ± 100kb of the index SNP. In this case, the single gene was chosen for gene expression data. Finally, in gene deserts where there is no gene within ± 100kb of the index SNP, we excluded the index SNP from our analysis. Detailed information of selected SNPs and genes for blood pressure and BMI are shown in Chapter 3, Appendix 3.2 and Chapter 4, Appendix 4.2 respectively.

Statistical Methods

All data management and statistical analyses were conducted using R version 2.14. Allele frequencies, genotype frequencies, and linkage disequilibrium (LD), as measured by r^2, were estimated for all pairs of SNPs using an expectation maximization (EM) algorithm. Before performing the statistical analysis, variable distributions were examined for normality and outliers by using histograms and QQ-plots. Outliers greater than four standard deviations from the mean were removed.

Quantitative variables of interest include age (years), BMI (kg/m^2), SBP (mm Hg), DBP (mm Hg), and gene expression data. Categorical variables include gender (0=male, 1=female). Before analysis, 15 mm Hg was added to observed SBP values and 10 mm Hg was added to DBP values for participants who were taking antihypertensive medication to adjust for the effect of blood pressure lowering medications.[65] This adjustment was consistent with the antihypertensive drug adjustment in ICBP. The outcome variables (SBP, DBP, and BMI) were first adjusted (adjusted age, age^2, gender and BMI for blood pressure; adjusted age and gender for BMI) in a linear regression model, and the residual was saved. Then, this residual variable was treated as the outcome variable in the linear mixed effects modeling to test for association with the predictor (SNP).

Association Testing

A two-stage approach was used in order to identify both marginal and interactive genetic effects associated with blood pressure and BMI. In stage 1, we examine each SNP for association with blood pressure and BMI. Specifically, 3537 BP-related SNPs were tested for their association with blood pressure (Chapter 3, Appendix 3.2), and 4213 BMI-related SNPs were tested for BMI (Chapter 4, Appendix 4.2). In this chapter, we used the genotypic model because

the additive model is a limited model where the beta estimate (β) of the interaction term is constrained. When SNP is coded as 0,1,2 and is treated as one variable, the beta estimate (β) of the interaction term (SNP by gene expression) represents the difference of the slopes between each genotype for the association of gene expression level and phenotype. For example, the difference of slopes between SNP=0 and SNP=1 is 1β, the difference of slopes between SNP=1 and SNP=2 is 1β, and the difference of slopes between SNP=0 and SNP=2 is 2β. In other words, when we coded the SNPs as the additive model, this interaction term will only be detected when the difference of the slopes for each genotype is additive. Therefore, by using genotypic model, we were able to detect more SNP by gene expression interaction on the phenotypic traits that is not restricted to the additive model. In genotypic model, these selected SNPs were coded as dummy variables where SNP_{bb} genotype is the reference genotype, where b represents the alphabetically lower allele. The other two genotypes were coded as SNP_{ab}=1 if heterozygous, and 0 if else; SNP_{aa}=1 if homozygous for the alphabetically higher allele and 0 if else. The statistical model is shown as below. Statistical significance of the marginal effect of each respective SNP was determined based on model p-value.

$$\text{Outcome residual}_{ij} = \beta_{oi} + \beta_1(SNP_{ab\ ij}) + \beta_2(SNP_{aa\ jj}) + \varepsilon_{ij}$$

$$\mathbf{H_o}: \beta_1 = 0 \text{ and } \beta_2 = 0, \qquad \mathbf{H_a}: \beta_1 \neq 0 \text{ or } \beta_2 \neq 0 \qquad \sim\chi^2_{(df=2)}$$

In stage 2, proximal SNPs and gene expression data from the same genes were used to test for their interaction effect on blood pressure (Chapter 3, Appendix 3.2), and BMI (Chapter 4, Appendix 4.2) regardless of the marginal effects of the SNPs tested in stage 1. Statistical significance of the interaction was assessed using a likelihood ratio (LR) test statistic by

comparing the full model (including interaction terms and main effects) to a reduced model that only includes main effects.

Full model : Outcome residual$_{ij}$ = β_{oi} + β_1(SNP$_{ab\ ij}$) + β_2(SNP$_{aa\ jj}$) + β_3(Gene Expression$_{ij}$) +

β_4(SNP$_{ab\ ij}$) (Gene Expression$_{ij}$) +

β_5(SNP$_{aa\ ij}$) (Gene Expression$_{ij}$) + ε_{ij}

Reduced model : Outcome residual$_{ij}$ = β_{oi} + β_1(SNP$_{ab\ ij}$) + β_2(SNP$_{aa\ jj}$) + β_3(Gene Expression$_{ij}$)

$\mathbf{H_o}$: $\beta_4 = 0$ and $\beta_5 = 0$, $\mathbf{H_a}$: $\beta_4 \neq 0$ or $\beta_5 \neq 0$

LR = -2 (log likelihood reduced model – log likelihood full model) $\sim\chi^2_{(df=2)}$

In the linear mixed effect model, i represents family and j represents individual. β_{oi} is a random intercept for the ith family, and it is normally distributed with mean β_o and variance σ_b^2. ε_{ij} is the residual variation within the jth individual from the ith family, and it is normally distributed with mean 0 and variance σ^2. ε_{ij} is assumed to be independent from β_{oi}.

Results

Descriptive statistics of the outcome variables and covariates can be found in Table 5.1. The mean age of the subjects was 55.65 years old, 58.05% were female, 41.44% were obese, around three-fourth of them (74.14%) had hypertension, and more than half of them (66.16%) were taking antihypertensive medication. Participants had a mean SBP of 132.80 mm-Hg and DBP of 78.52 mm Hg. The mean, standard deviation, minimum, and maximum statistics of gene expression data for the 27 BP- related genes and the 30 BMI-related genes are shown in chapter 2, Table 2.3 and Table 2.4, respectively.

Marginal Effects of SNPs and Interaction of SNPs*Gene Expression Levels on Outcomes

3537 BP-related SNPs and 4213 BMI-related SNPs were tested for their marginal effects and their interaction with gene expression level on blood pressure and BMI, respectively. Out of 3537 SNPs for the outcome variable of SBP, 322 SNPs had significant marginal effects, 326 SNPs showed significant interaction with gene expression level, and 23 of them had both marginal and interaction effects (p<0.05) (Table 5.2). After applying Bonferroni correction, only four SNPs (rs1779244, rs11598236, rs1779241, and rs2482103) located in *CACNB2* showed significant interaction with *CACNB2* expression on systolic blood pressure ($p<5.41x10^{-5}$) (Table 5.3). These four SNPs were in high LD ($r^2>0.95$) and their interaction with *CACNB2* expression explained around 2.04%-2.28% variation on systolic blood pressure. However, the marginal effects of these four SNPs were not statistically significant (p-values ranged from 0.099 to 0.2543) (Table 5.4). In addition, *CACNB2* expression was not associated with SBP (chapter 3, Table 3.6). The relationship of the most significant interaction term (i.e. rs1779244 by *CACNB2* expression) associated with systolic blood pressure is shown in Figure 5.1. The regression line indicates that as *CACNB2* expression level increases, individuals with genotype CA have increased predicted systolic blood pressure while those have genotype CC have decreased predicted systolic blood pressure.

There were 239 SNPs and 326 SNPs with significant marginal effects and interaction with gene expression level on diastolic blood pressure, respectively (Table 5.2). Only 9 of them had both marginal and interaction effects (p<0.05). None of these associations were significnat after applying Bonferroni correction. Similarly, 362 SNPs and 299 SNPs were associated with marginal and interaction effects on BMI, respectively. Of these, 23 SNPs had both marginal and interaction effects (p<0.05) (Table 5.5). None of these association tests passed Bonferroni

correction for multiple testing. In summary, only 7.1% (=23/326), 2.5% (=9/351), and 7.7% (=23/299) of those SNPs that had significant interaction with gene expression also had marginal effects (p<0.05) for the outcome variable of SBP, DBP and BMI, respectively.

Discussion

One notable feature of the results from SNP by gene expression interaction analysis is that the majority of SNPs (~92%) that had significant interaction with gene expression did not have significant marginal effects (Table 5.2 and Table 5.5). Previous literature has also reported that SNPs involved in an interaction will not always exhibit significant marginal effects[101], and some studies haves shown this phenomenon in coronary artery disease[102] and atrial fibrillation.[103] Our results provide evidence in support of the approach to search for SNPs with interaction effects, not based on initially significant marginal effects. For the four interactions in *CACNB2* gene we found, none of them had significant marginal effects (p-values ranges from 0.099 to 0.2543) (Table 5.4).

We found four significant statistical interactions of genetic variants by *CACNB2* expression on systolic blood pressure. A crucial question regarding these interaction results is: how might these SNPs interact biologically with *CACNB2* expression to influence systolic blood pressure? Before figuring out the answer, it is important to recognize the difference between "biological interaction"[104] and "statistical interaction"[105]. Biological interaction was first reported by William Bateson in 1909, and it can be described as the situation in which the causes of diseases are dependent on the presence of other factors.[104] A good example of biological interaction is phenylkeonutia (PKU), a autosomal recessive metabolic genetic disorder that requires the combination of two sufficient elements: a genetic mutation and an environmental factor. People with PKU are missing an enzyme called phenylalanine hydroxylase (PAH), and thus, they cannot

process one of the amino acids called phenylalanine which is commonly found in foods that contain protein. Without this enzyme, levels of phenylalanine build up in the body, and these substances can cause severe mental retardation.[106] To prevent the development of this disease, infants with two copies of a mutated *PAH* gene are put on a life-long restricted diet. On the other hand, statistical interaction refers to the deviation from additivity in a linear regression model by Fisher's 1981 definition.[105] The extent to which biological interaction can be inferred from statistical interaction might be limited and has been extensively debated in both the genetic[107,108] and epidemiological literature.[109,110]

CACNB2 (chromosome 10p12) encodes the beta-2 subunit of a voltage-gated calcium channel protein and is one member of a family of voltage-gated calcium channel genes. This gene family influences blood pressure regulation and serves as a target of calcium channel blockers. *CACNB2* may alter blood pressure through the interaction of the beta-2 subunit with alpha-1 calcium channels (CaV1.2).[81] Mutations in *CACNB2* are associated with a type of heart disease, Brugada syndrome type 4, which can cause the ventricles to beat so fast that the blood is prevented from circulating efficiently in the body.[80] Evidence that genetic variants in *CACNB2* act as context-dependent modifiers has been shown in hypertensive patients with adverse cardiovascular outcomes.[111] Yuxin *et al.* reported that among patients with the rs2357928 GG genotype, individuals who were randomized to a calcium channel blocker were more likely to experience an adverse cardiovascular outcome than those randomized to a beta-blocker (Hazard ratio=2.35, p=0.0143). No evidence was found for treatment difference among patients with the AA or AG genotype.[111]

One of the major limitations of this current study is the lack of a replication sample with adequate data including genetic variants, gene expression, and the relevant phenotype

information in non-Hispanic whites. Replication of the results in an independent sample will be critical to inform future studies attempting to investigate the functional mechanism associated with the modifying influence of the genetic variants we found. Second, due to the heterogeneity (i.e. population differences in allelic distribution), our results may not be generalizable to other ethnic groups. Third, as has been discussed above, inferences of biological interaction are difficult to make from tests of statistical interaction. In addition, the lymphocyte cell line we used has been transformed, and it may not reflect the natural biological pathway in the *in vivo* cells.

Despite these limitations, this study contributed to the developing knowledge of gene by environment interaction by showing that the relationship between *CACNB2* expression and systolic blood pressure may be modified by some intrinsic genetic variants in *CNCAB2*. To our knowledge, no one has investigated the influence of SNP*gene expression interaction on blood pressure. Our results present a new avenue of investigation into the potential link between gene expression and blood pressure.

Table 5.1 Descriptive statistics of covariates and outcome variables.

Continuous Variables	Mean	SD
Age, years	55.65	10.36
Height, cm	168.41	9.09
Weight, kg	84.53	18.64
BMI, kg/m^2	29.73	5.82
SBP, mm Hg	132.80	16.66
SBP+15mm Hg for Hypertension Drug	142.73	19.27
DBP, mm Hg	78.52	9.43
DBP+10mm Hg for Hypertension Drug	85.14	10.69
Categorical Variables	**Number**	**%**
Female	458	58.05
Obesity	327	41.44
Hypertension	585	74.14
Antihypertensive Medication	522	66.16

Table 5.2. Summary of the number of SNP marginal effects and SNP*gene expression interactions associated with SBP and DBP (p<0.05).

Gene	Chr	No. of Proximal SNPs	SNP Marginal effects			SNP*Gene Expression Interaction			SNP Marginal effects & SNP*Gene Expression Interaction		
			SBP	DBP	Both SBP and DBP	SBP	DBP	Both SBP and DBP	SBP	DBP	Both SBP and DBP
CACNB2	10	542	28	24	9	76	71	45	4	3	0
MECOM	3	563	34	44	24	90	40	31	11	2	0
ARHGAP42	11	497	35	41	21	30	96	24	0	0	0
PLEKHA7	11	299	32	20	5	24	28	21	1	0	0
ULK4	3	512	69	8	1	14	34	5	7	0	0
MTHFR	1	66	0	4	0	6	7	5	0	0	0
BAT3	6	65	1	3	0	15	8	4	0	3	0
CNNM2	10	49	0	1	0	3	8	3	0	0	0
ZNF652	17	66	1	1	0	3	3	3	0	0	0
NT5C2	10	49	0	1	0	3	0	3	0	0	0
GUCY1A3	4	103	14	1	0	5	5	2	0	0	0
FES	15	25	4	5	4	6	1	1	0	0	0
PLCD3	17	51	0	3	0	1	3	1	0	1	0
PLCE1	10	255	54	52	30	20	0	0	0	0	0
CSK	15	28	0	0	0	19	0	0	0	0	0
PRDM8	4	40	1	2	2	4	1	0	0	0	0
SLC39A8	4	133	8	4	4	3	24	0	0	0	0
FGF5	4	40	1	2	2	2	0	0	0	0	0
ADM	11	35	6	5	4	1	6	0	0	0	0
CAPZA1	1	24	0	4	0	1	0	0	0	0	0
GOSR2	17	48	23	0	0	0	15	0	0	0	0
MOV10	1	24	0	4	0	0	1	0	0	0	0
SH2B3	12	20	5	2	2	0	0	0	0	0	0
ATP2B1	12	14	4	5	4	0	0	0	0	0	0
NPR3	5	25	1	2	1	0	0	0	0	0	0
SLC4A7	3	24	0	0	0	0	0	0	0	0	0
HFE	6	53	1	1	0	0	0	0	0	0	0
Total	---	3537	322	239	113	326	351	148	23	9	0

Table 5.3. Four significant SNP by *CACNB2* expression interactions on SBP that passed Bonferroni correction ($p<9.23x10^{-5}$)

SNP	Function	Allele A	Allele B	Allele B Freq.	Beta			P-value			Interaction R^2	LD r^2
					Genotype AB	GE	AB* GE	Genotype AB	GE	AB* GE		
rs1779244	Intron	C	A	0.089	2.16	-3.66	13.59	0.195617	0.0075	$2.64x10^{-5}$	2.28%	Ref
rs11598236	Intron	T	C	0.091	1.48	-3.71	13.35	0.368482	0.0071	$2.86x10^{-5}$	2.20%	0.95
rs1779241	Intron	G	A	0.090	1.78	-3.56	13.12	0.284263	0.0095	$5.31x10^{-5}$	2.08%	0.98
rs2482103	Intron	G	A	0.090	2.41	-3.63	13.12	0.149198	0.0081	$5.41x10^{-5}$	2.04%	0.99

Note: GE=gene expression. Genotype AA is the reference for dummy variable. There were less than 10 people in genotype BB thus were removed from analysis.
LD=Linkage disequilibrium.

Table 5.4. Marginal effects of the four SNPs that had significant interaction with *CACNB2* expression on systolic blood pressure.

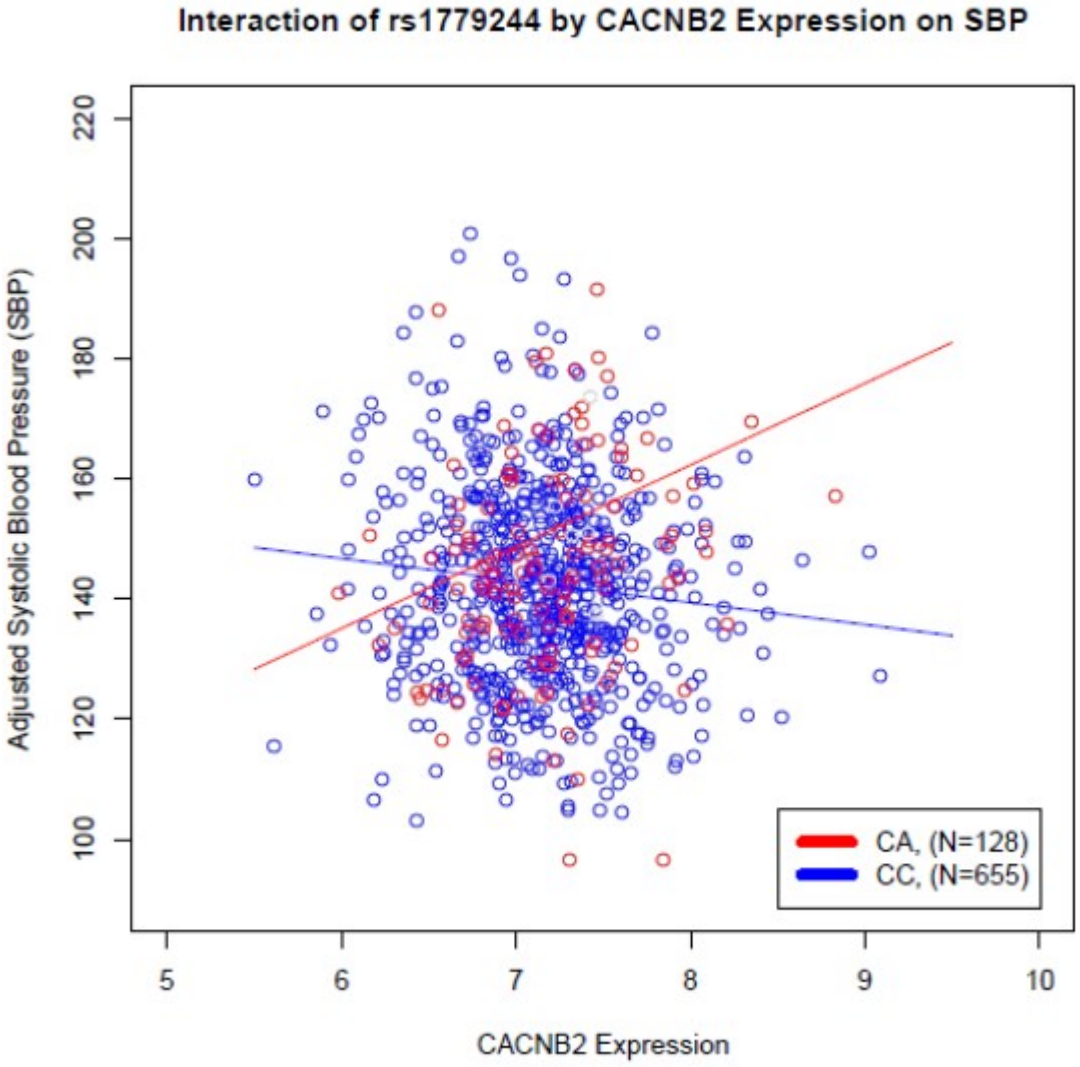
SNP	Position	Genotype AA	Genotype AB	Genotype BB	No. of Genotype AA	No. of Genotype AB	No. of Genotype BB	Genotype AB		LD r^2
								Beta	P-value	
rs1779244	18477625	CC	CA	AA	655	128	0	2.52	0.1334	Ref
rs11598236	18451884	TT	TC	CC	651	132	0	1.89	0.2543	0.95
rs1779241	18476322	GG	GA	AA	653	130	0	2.15	0.1975	0.98
rs2482103	18479948	GG	GA	AA	654	128	0	2.78	0.0990	0.99

Note. Genotype AA is the reference for dummy variable.
LD= linkage disequilibrium.

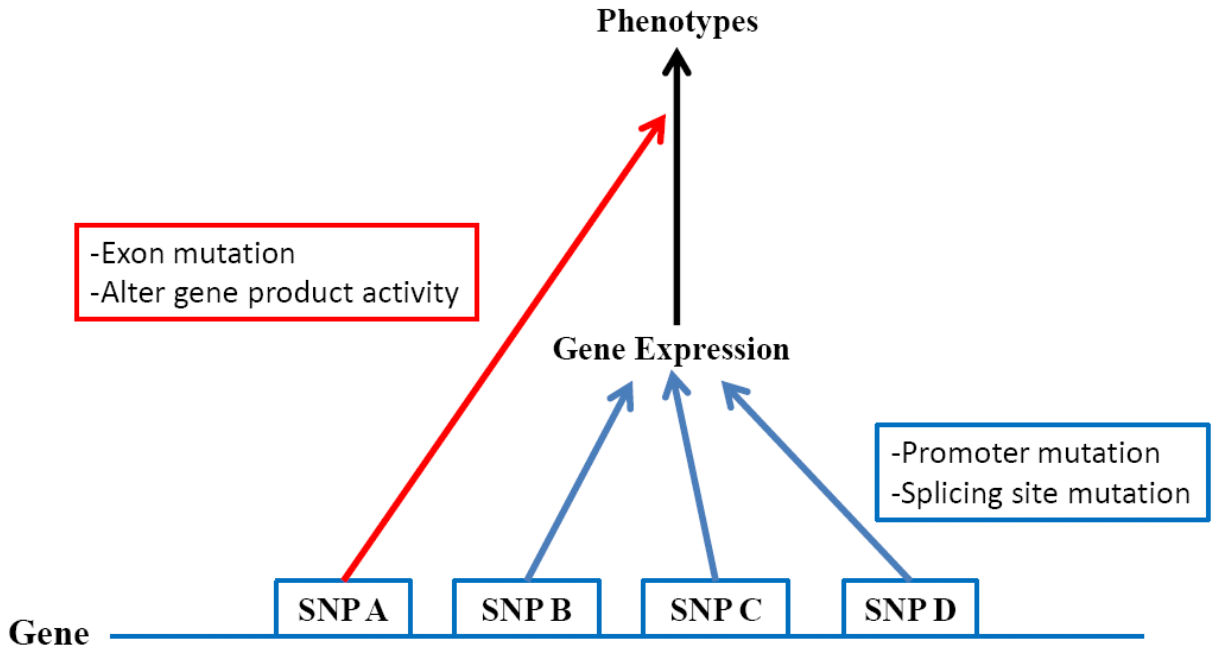
Table 5.5. Summary of the number of SNP marginal effects and SNP*gene expression interactions associated with BMI (p<0.05).

Gene	Chr	No. of Proximal SNPs	SNP Marginal effects	SNP*Gene Expression Interactions	SNP Marginal effects &SNP*Gene Expression Interactions
NRXN3	14	1229	88	169	3
LINGO2	9	955	21	49	7
CADM2	3	230	41	28	9
TNNI3K	1	260	18	9	0
LRP1B	2	47	1	9	0
TMEM18	2	101	59	7	2
STK33	11	133	6	6	2
DNAJC27	2	20	0	5	0
MTIF3	13	55	0	4	0
BCDIN3D	12	31	0	4	0
QPCTL	19	12	0	3	0
TFAP2B	6	57	1	2	0
FTO	16	454	70	2	0
MTCH2	11	20	0	1	0
DGKG	3	28	1	1	0
SLC39A8	4	133	6	0	0
NUDT3	6	37	0	0	0
MAP2K5	15	244	0	0	0
FAIM2	12	31	0	0	0
SH2B1	16	13	7	0	0
ATP2A1	16	13	7	0	0
KCTD15	19	26	4	0	0
CHST8	19	26	4	0	0
ADCY3	2	20	0	0	0
ETV5	3	28	1	0	0
GPRC5B	16	19	1	0	0
SEC16B	1	53	14	0	0
POC5	5	28	8	0	0
BDNF	11	14	4	0	0
NEGR1	1	14	0	0	0
Total	--	4213	362	299	23

Figure 5.1. Interaction of rs1779244 by *CACNB2* expression on systolic blood pressure.



Appendix 5.1. Mechanism of genetic variants associated with gene expression levels and phenotypic traits.



Appendix 5.2. The influence of gene product activity on the association between gene expression levels and phenotypic traits.

	Gene Product Activity			
Gene Expression Level		Good	Medium	Poor
	High	Phenotype = ++++	++	-
	Medium	+++	+	-
	Low	++	-	-

Chapter 6

Conclusion and Future Directions

Summary of Findings

In this dissertation, I presented an integrative approach of combining information about genetic variants, gene expression levels and phenotypic data to take the next step in investigating the potential functional mechanisms underlying the etiology of increased blood pressure and body mass index (BMI) among non-Hispanic whites in the Genetic Epidemiology Network of Arteriopathy (GENOA) study.

At the beginning of this dissertation (Chapter 2), we were interested in exploring how much variation in blood pressure, BMI, and gene expression levels is due to genetic factors. The heritability results indicated that both blood pressure and BMI were attributable, at least in part, to additive genetic factors (h^2: systolic blood pressure (SBP) =0.35, diastolic blood pressure (DBP) =0.31, and BMI=0.47) after adjusting for known risk factors in the statistical model in the SOLAR analysis. In addition, the estimates of h^2 in gene expression levels at previously identified genetic loci revealed that around 67% and 47% of the expression levels in, respectively, the 27 BP-related genes and the 30 BMI-related genes were significantly heritable ($p<0.05$). Furthermore, the Pearson correlation analysis of gene expression levels suggested that 75% and 80% of the correlation coefficients were significant ($p<0.05$) for BP-related genes and BMI-related genes, respectively.

After demonstrating that there were substantial variations in the phenotypes of interest and gene expression levels influenced by genetic factors in our study sample, we then conducted a genetic epidemiology study to investigate the potential functional genetic factors underlying blood pressure (Chapter 3) and BMI (Chapter 4). Then, we examined whether SNPs modify the association between gene expression levels and three complex traits (i.e. SBP, DBP, and BMI) (Chapter 5). The main finding from chapter 3 revealed that there were 35 genetic variants in *ULK4* that simultaneously contributed to the variation in DBP ($p<0.05$) and *ULK4* expression ($p<10^{-45}$). In addition, a significant association was found between *ULK4* expression and DBP ($p=0.0004$). All of these association tests passed Bonferroni correction except for the associations between SNPs and DBP. Similarly, 28 SNPs were identified to be associated with both SBP and *ULK4* expression and a significant association was found between *ULK4* expression and SBP ($p=0.0148$). However, none of these association tests passed Bonferroni correction except for the association between rs9311289 and *ULK4* expression. Interestingly, in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, investigators also reported that three missense SNPs (rs1716975, rs2272007 and rs1052501, $p<10^{-7}$) were associated with altered expression of *ULK4* among HapMap CEU in lymphoblastoid cell lines were found to be significantly associated with DBP ($p<5x10^{-5}$) but not with SBP ($p=0.84$-0.94) in their study.[16] As has been discussed in Chapter 3, our study only has moderate power to detect an association between SNPs and blood pressure levels, and thus, none of these associations were statistically significant after Bonferroni correction in our study. Besides *ULK4*, 7 and 3 proximal SNPs were strongly associated with *MTHF* expression ($p<1x10^{-8}$) and *PLEKHA7* expression ($p<1x10^{-5}$), respectively, but they were only moderately associated with blood pressure levels ($p<0.05$). Although no significant associations were found

between their expression levels and blood pressure, the strong signals in these SNPs suggested that the variation in these gene expression levels were significantly attributable to these SNPs.

Chapter 4 focused on studying the potential functional genetic factors contributing to BMI. Among 4213 genetic variants, 198 SNPs had strong *cis*-effects on the expression levels of 10 genes (13 SNPs in *ADCY3*, 52 SNPs in *TNNI3K*, 17 SNPs in *STK33*, 15 SNPs in *MTCH2*, 54 SNPs in *MAP2K5*, 31 SNPs in *NUDT3*, 4 SNPs in *TMEM18*, 8 SNPs in *MTIF3*, 3 SNPs in *KCTD15*, and 1 SNP in *QPCTL*). However, only one (rs10947494 in *NUTD3*) of them had a moderate association with BMI (p=0.045), and none of genes mentioned above had its expression level associated with BMI. As discussed in chapter 4, beta-transformed lymphocytes might not be the most relevant tissue to use for examination of gene expression levels associated with BMI, and this could explain why we did not find any significant associations between them. For example, Emilsson *et al.* found that there were more gene expression levels (72.0%, 17,080 out of 23,720 gene transcripts) correlated with BMI in adipose tissue than those in blood tissue (9.2%, 2,172 gene transcripts).[51] Except for this, our results indicated that there were some genetic variants contributing to the variation in gene expression levels at previously identified genetic regions reported from large genome-wide association studies (GWAS).

There were two major findings in chapter 5. First, we found that there were four SNPs (rs1779244, rs11598236, rs1779241, and rs2482103) located in *CACNB2* modifying the relationship between *CACNB2* expression level and SBP ($p<5.41\text{x}10^{-5}$). These four SNPs were in high LD ($r^2>0.95$), and their interaction with *CACNB2* expression explained around 2.04%-2.28% variation in systolic blood pressure. None of these four SNPs had significant marginal effects on SBP. Second, we observed that majority of the SNPs that had significant interaction with gene expression levels did not have obvious marginal effects on the phenotypic traits (SBP:

93%, DBP: 97%; BMI: 92%). This result was consistent with what previous studies have reported. SNPs involved in an interaction will not always exhibit significant marginal effects.[101] Investigating the context dependent genetic contribution to phenotypic traits not based on marginal effects will allow us to detect more significant interaction signals.

Comparison of Blood Pressure and BMI Results

In Chapter 2, the heritability of of BMI (h^2=0.47) was found to be higher than blood pressure (h^2: SBP=0.35, DBP=0.31) after adjusting for the risk factors. In addition, around 67% and 47% of the expression levels in the 27 BP-related genes and the 30 BMI-related genes were significantly heritable, respectively (p<0.05). In Chapter 3,we found that 21 SNPs and 22 SNPs in the *ULK4* gene were associated with *ULK4* expression and its expression was associated with SBP and DBP, respectively. After adjusting for *ULK4* expression, the associations between these SNPs and blood pressure were no longer significant (p>0.05) which suggests that *ULK4* expression might mediate the association between SNPs and blood pressure. However, in Chapter 4, only *TFAP2B* expression was associated with BMI (p=0.042) and we did not find any gene expression levels mediated the association between genetic variants and BMI. As discussed in chapter 4, beta-transformed lymphocytes might not be the most relevant tissue associated with BMI. For example, Emilsson *et al.* reported that there were more transcripts in adipose tissues (72.0%, 17,080 out of 23,720 gene transcripts) than those in blood (72.0%, 17,080 out of 23,720 gene transcripts) that were found to be associated with BMI. Thus, adipose tissues were suggested to be the relevant tissue for studying gene expression levels associated with BMI and obesity in future studies. Lastly, in Chapter 5, four SNPs located in the *CACNB2* gene were found to modify the association between *CACNB2* expression and SBP. However, we did not find any SNPs that modify the association between gene expression levels and BMI.

Public Health Significance

One of the significant findings from this dissertation was that *ULK4* expression levels might mediate the association between genetic variants and blood pressure. After adjusting for epidemiological risk factors and SNPs that were associated with *ULK4* expression, every 1 standard deviation increment of *ULK4* expression was associated with ~ 2.0 mm Hg increment of SBP and ~ 1.5 mm Hg increment of DBP (Appendix 3.24 and Appendix 3.25, respectively). This increment has been shown to produce meaningful population changes in cardiovascular and stroke risk. Previous studies have reported that 2 mm Hg lower SBP was associated with 6% less stroke and 5% less coronary heart disease.[112,113]

Another important finding from this dissertation was four SNPs located in the *CACNB2* gene were found to modify the association between *CACNB2* expression and SBP. *CACNB2* (chromosome 10p12) encodes the beta-2 subunit of a voltage-gated calcium channel protein and calcium channels mediate Ca^{+} entry into cells in response to action potentials and subthreshold depolarizing signals. Previous literature found that genetic variation within *CACNB2* may influence treatment-related adverse cardiovascular outcomes in high risk patients with hypertension.[111] Among patients with the rs2357928 GG genotype, individuals who were randomized to a calcium channel blocker were more likely to experience an adverse cardiovascular outcome than those randomized to a beta-blocker (Hazard ratio=2.35, p=0.0143). Their results suggested that personalized medicine could be suggested based on people's genetic background of the *CACNB2* gene to control for adverse cardiovascular outcome. In addition, it is promising to investigate whether personalized medicine could lower blood pressure for those individuals who were at higher risk for increased blood pressure when they have higher *CACNB2* expression levels (e.g. genotype CA of rs1779244 in *CACNB2* gene, Chapter 5, Figure 5.1).

Limitations

There were several limitations in this dissertation that need to be considered. First, beta transformed lymphocytes were the only cell type that we used in these studies. It is possible that because these cell lines were transformed and cultured under artificial conditions, they may not represent the natural gene expression state *in vivo.* Therefore, although we did not detect many statistically significant associations between gene expression levels and phenotypic traits in this dissertation, we cannot exclude the possibility that gene expression levels *in vivo* have an influence on phenotypic traits. However, one of the advantages of using a transformed cell line is that standardization of the environment in cell culture removes individual-level environmental variation and thus the measured gene expression levels could optimally reflect individual DNA variations.

Currently, the genotyping platforms for GWAS, such as those from Affymetrix and Illumina, are often designed based on the "common disease common variant" hypothesis which posits that the genetic etiology of common diseases is mostly the result of susceptibility alleles which have moderate frequency. Thus, rare variants are not tagged well through the LD structure in GWAS.[114] Previous studies have already shown that rare variants were associated with both blood pressure and BMI.[115,116] Due to the lack of adequate power, single rare variants were less likely to be detected in this dissertation. However, new analytical strategies have been developed, such as sequence kernel association test (SKAT), which will allow us to test for association between genetic variants (both common and rare variants) in a region and disease traits in future studies.[117]

One of the limitations in this dissertation is the strategy to adjust for the effects of antihypertensive drugs. We adjusted for drug treatment by adding 15 mm Hg and 10 mm Hg to

SBP and DBP, respectively, to be consistent with the adjustment method in the International Consortium for Blood Pressure Genome-Wide Association Studies (ICBP). Tobin *et al.* compared multiple adjustment methods in simulated data to assess the estimation bias and the loss of power and then compared with real data. Based on the true beta estimate of a genetic factor's association with SBP (β=3.00), the least biased genetic estimate was attained when adding 15 mm Hg to SBP (β=3.00), followed by using a censored normal regression model (β=3.04), adding 10 mm Hg to SBP (β=2.90), ignoring antihypertensive treatment (β=2.68), and excluding individuals on therapy (β=2.55). In addition, the largest power was reached when using a censored normal regression model (82.6), followed by adding 15 mm Hg to SBP (82.5), adding 10 mm Hg to SBP (82.3), ignoring antihypertensive treatment (79.8), excluding individuals on therapy (64.8) and fitting a regression model with treatment as a covariate.[65] There are two adjustment methods that performed quite well and have been advocated in previous literatures.[65,118,119] The first one is what we applied here: the addition of a constant reasonable value to the observed blood pressure in treated subjects. The second one is a censored normal regression model with treated blood pressure considered as censored.[65] Although the first method is commonly used in most blood pressure studies, other factors such as medication categories, drug dosage, and variation in prescription patterns could influence blood pressure regulation and this adjustment scheme might be an oversimplification. Nevertheless, this drug adjustment should generally lead the bias toward the null. An alternative method for treatment adjustment was also suggested by Cui *et al.* (the addition of 10/5 mm Hg to SBP and DBP, respectively) which gave very similar results in an interim analysis in our study.[75]

Replication of these results in an independent dataset is crucial for validation. There are several criteria needed for an independent study to qualify as a replication sample. This sample

will need to have similar demographic distributions (i.e. age, gender percentage, blood pressure, and BMI) as ours, have genotype data available and gene expression levels measured in transformed beta-lymphocytes. Because we have not found an adequate replication sample that meets all of the criteria mentioned above, inference from this study needs to be carefully addressed, and replication is required in future studies.

The study population of non-Hispanic whites in GENOA were generally older and around 75% of them were hypertensive. Therefore the inferences from this dissertation may not be generalizable to individuals who are younger, normotensive or belong to other ethnic groups.

Last, cell lines for gene expression data were collected and transformed during 1995-2005 and only 30% (=237/789) of the cell lines were collected during 1995-2000 when phenotype data were collected. Therefore, establishing causality between gene expression levels and phenotypic traits from this dissertation is not possible at this moment because the majority of the gene expression levels were measured after collecting the phenotype data. However, this is not an issue for the individual genotypes because they are relatively constant within individuals. As we have rationalized in the introduction, gene expression levels can represent the cumulative influences of multiple DNA variations. Given that the gene expression levels in this dissertation were mostly significantly heritable and attributable substantially to genetic factors, we cannot exclude the possibility that the link between DNA variants and our phenotypic traits might be mediated through gene expression levels. Specifically, *ULK4* expression was highly heritable (0.71) and it had a significant association with DBP (p=0.0004). A longitudinal study could help us to unravel the puzzle from DNA variants to gene expression and then to these phenotypes.

Potential Follow-up Studies in GENOA

Owing to the emergence of advanced sequencing technologies (i.e. next-generation

sequencing), the dense sequence data in the 1000 Genomes project offers a great opportunity for researchers to impute up to 22 million SNPs (both rare and common variants).[120] Imputation procedures using 1000 Genomes project as a reference panel have been completed both in non-Hispanic whites and African Americans in GENOA. With this imputed dataset, we will be able to investigate more common variants associated with the complex traits for each genetic region of interest. GWAS studies with imputed SNPs based on the 1000 Genomes project have identified novel signals associated with diabetes[121] and coronary artery disease.[122] For example, in Welcome Case Control Consortium (WTCCC), investigators detected two novel signals associated with diabetes. One genetic variant within the *IL2RA* gene is associated with type I diabetes, and the other one proximal to *CDKN2B* gene is associated with type II diabetes.[121]

In addition to common variants, rare variants can be examined through using a remarkably flexible and powerful analytic method called SKAT which uses a score-based variance component test.[117] This new analytic strategy will allow us to evaluate the aggregate impact of multiple rare variants on phenotypic variation in GENOA. There are several advantages of SKAT over the other rare variants analysis methods such as the C-alpha test,[123] Combined Multivariate and Collapsing (CMC),[124] Cumulative Minor Allele Test (CMAT),[125] Variable Threshold (VT),[126] and Weighted Sum Statistic (WSS).[127] The advantages of SKAT are: 1) rare variants are not assumed to have similar direction and magnitude of effects and these values are estimated from the data instead, 2) covariates and population stratification (principal components) can be easily adjusted in the model, 3) both dichotomous and continuous traits can be tested, 4) multiple genetic models (dominance, recessive, epistasis) can be evaluated, 5) SKAT's p-values are analytically derived without the need for intensive permutation, and 6) both unrelated sample and family data can be handeled by SKAT. Examining rare variants in the

genetic regions identified from previous GWAS will aid in evaluating their association with these two complex traits in GENOA.

Future Directions

Given the complex nature of many common chronic diseases involved with multiple genetic and environmental factors, unraveling the biological mechanism underlying these diseases is important but challenging. Integration of comprehensive information from genomic, transcriptomic, proteomic, and epigenetic data will help researchers to elucidate biological pathways and identify novel pathophysiological mechanisms contributing to common, complex multifactorial diseases. For example, the integration of genomic and transcriptomic data in humans and mice have identified *SORT1* and *CELSR2* to be associated with coronary artery disease and plasma low-density lipoprotein cholesterol levels in the process.[128] In addition, the role of genetic variants in determining protein levels has been assessed at a genome-wide level on 42 proteins.[129] Eight *cis*-effects were identified in or near the *IL6R, CCL4, IL18, LPA, GGT1, SHBG, CRP* and *IL1RN* genes, and all of them were associated with their respective protein products. It is unknown whether the altered gene expression levels and protein levels are involved with disease etiology or are just a result of disease process.[130,131] The identification of genetic variants that influence gene expression levels and alter protein levels may help us to dissect their relationship with the complex diseases.

Conclusion

The work in this dissertation presented an integrative approach to investigating the potential functional mechanisms of common genetic variants associated with blood pressure and BMI through studying their marginal effects and interaction with gene expression levels. Significant findings from this research highlight the importance of mutually considering different aspects of

genetic and biological information linking from DNA to gene expression level and then to phenotypic traits. Replication of our findings in an independent sample is required before making any strong conclusions about SNPs that simultaneously influence both gene expression levels and disease traits. In summary, the results from this dissertation and future studies in GENOA will continuously provide a deeper knowledge of the potential functional role of genetic variants underlying two major public health issues: increased blood pressure and BMI.

References

1. MacMahon S, Peto R, Cutler J, Collins R, Sorlie P, Neaton J *et al.* Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* 1990; **335**(8692)**:** 765-774.

2. Mittendorfer B, Peterson LR. Cardiovascular Consequences of Obesity and Targets for Treatment. *Drug Discov Today Ther Strateg* 2008; **5**(1)**:** 53-61.

3. Turner ST, Boerwinkle E. Genetics of hypertension, target-organ complications, and response to therapy. *Circulation* 2000; **102**(20 Suppl 4)**:** IV40-45.

4. Lloyd-Jones D, Adams R, Carnethon M, De Simone G, Ferguson TB, Flegal K *et al.* Heart disease and stroke statistics--2009 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* 2009; **119**(3)**:** 480-486.

5. Levy D, DeStefano AL, Larson MG, O'Donnell CJ, Lifton RP, Gavras H *et al.* Evidence for a gene influencing blood pressure on chromosome 17. Genome scan linkage results for longitudinal blood pressure phenotypes in subjects from the framingham heart study. *Hypertension* 2000; **36**(4)**:** 477-483.

6. Elks CE, den Hoed M, Zhao JH, Sharp SJ, Wareham NJ, Loos RJ *et al.* Variability in the heritability of body mass index: a systematic review and meta-regression. *Front Endocrinol (Lausanne)* 2012; **3:** 29.

7. Maes HH, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet* 1997; **27**(4)**:** 325-351.

8. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011; **478**(7367)**:** 103-109.

9. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010; **42**(11)**:** 937-948.

10. Klabunde R. *Cardiovascular Physiology Concepts*. Lippincott Williams & Wilkins, 2005, 93-94pp.

11. Understanding blood pressure readings. American Heart Association: http://www.heart.org/HEARTORG/Conditions/HighBloodPressure/AboutHighBloodPressure/Understanding-Blood-Pressure-Readings_UCM_301764_Article.jsp.

12. Beevers G, Lip GY, O'Brien E. ABC of hypertension: The pathophysiology of hypertension. *BMJ* 2001; **322**(7291)**:** 912-916.

13. van Rijn MJ, Schut AF, Aulchenko YS, Deinum J, Sayed-Tabatabaei FA, Yazdanpanah M *et al.* Heritability of blood pressure traits and the genetic contribution to blood pressure variance explained by four blood-pressure-related genes. *J Hypertens* 2007; **25**(3)**:** 565-570.

14. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet* 2009.

15. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol* 2008; **32**(4)**:** 381-385.

16. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A *et al.* Genome-wide association study of blood pressure and hypertension. *Nat Genet* 2009.

17. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI *et al.* Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011; **478**(7367)**:** 103-109.

18. Matsukawa N, Grzesik WJ, Takahashi N, Pandey KN, Pang S, Yamauchi M *et al.* The natriuretic peptide clearance receptor locally modulates the physiological effects of the natriuretic peptide system. *Proc Natl Acad Sci U S A* 1999; **96**(13)**:** 7403-7408.

19. Ishimitsu T, Ono H, Minami J, Matsuoka H. Pathophysiologic and therapeutic implications of adrenomedullin in cardiovascular disorders. *Pharmacol Ther* 2006; **111**(3)**:** 909-927.

20. Martin RM, Lin CJ, Costa EM, de Oliveira ML, Carrilho A, Villar H *et al.* P450c17 deficiency in Brazilian patients: biochemical diagnosis through progesterone levels confirmed by CYP17 genotyping. *J Clin Endocrinol Metab* 2003; **88**(12)**:** 5739-5746.

21. Geller DH, Auchus RJ, Mendonca BB, Miller WL. The genetic and functional basis of isolated 17,20-lyase deficiency. *Nat Genet* 1997; **17**(2)**:** 201-205.

22. Mussig K, Kaltenbach S, Machicao F, Maser-Gluth C, Hartmann MF, Wudy SA *et al.* 17alpha-hydroxylase/17,20-lyase deficiency caused by a novel homozygous mutation (Y27Stop) in the cytochrome CYP17 gene. *J Clin Endocrinol Metab* 2005; **90**(7)**:** 4362-4365.

23. Larsson SC., Alicja W. *Epidemiology of Obesity and Diabetes*, Mantzoros CS. edn. Humana Press: New Jersey, 2006, 15-30pp.

24. Keith SW, Redden DT, Katzmarzyk PT, Boggiano MM, Hanlon EC, Benca RM *et al.* Putative contributors to the secular increase in obesity: exploring the roads less traveled. *Int J Obes (Lond)* 2006; **30**(11)**:** 1585-1594.

25. Atwood LD, Heard-Costa NL, Cupples LA, Jaquish CE, Wilson PW, D'Agostino RB. Genomewide linkage analysis of body mass index across 28 years of the Framingham Heart Study. *Am J Hum Genet* 2002; **71**(5)**:** 1044-1050.

26. Patel SR, Larkin EK, Redline S. Shared genetic basis for obstructive sleep apnea and adiposity measures. *Int J Obes (Lond)* 2008; **32**(5)**:** 795-800.

27. Coady SA, Jaquish CE, Fabsitz RR, Larson MG, Cupples LA, Myers RH. Genetic variability of adult body mass index: a longitudinal assessment in framingham families. *Obes Res* 2002; **10**(7)**:** 675-681.

28. Wu X, Cooper RS, Borecki I, Hanis C, Bray M, Lewis CE *et al.* A combined analysis of genomewide linkage scans for body mass index from the National Heart, Lung, and Blood Institute Family Blood Pressure Program. *Am J Hum Genet* 2002; **70**(5)**:** 1247-1256.

29. Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 2009; **41**(1)**:** 25-34.

30. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* 2009; **41**(1)**:** 18-24.

31. Lindgren CM, Heid IM, Randall JC, Lamina C, Steinthorsdottir V, Qi L *et al.* Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PLoS Genet* 2009; **5**(6)**:** e1000508.

32. Heard-Costa NL, Zillikens MC, Monda KL, Johansson A, Harris TB, Fu M *et al.* NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. *PLoS Genet* 2009; **5**(6)**:** e1000539.

33. Gerken T, Girard CA, Tung YC, Webby CJ, Saudek V, Hewitson KS *et al.* The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* 2007; **318**(5855)**:** 1469-1472.

34. Wahlen K, Sjolin E, Hoffstedt J. The common rs9939609 gene variant of the fat mass- and obesity-associated gene FTO is related to fat cell lipolysis. *J Lipid Res* 2008; **49**(3)**:** 607-611.

35. Walley AJ, Asher JE, Froguel P. The genetic contribution to non-syndromic human obesity. *Nat Rev Genet* 2009; **10**(7)**:** 431-442.

36. Ren D, Zhou Y, Morris D, Li M, Li Z, Rui L. Neuronal SH2B1 is essential for controlling energy and glucose homeostasis. *J Clin Invest* 2007; **117**(2)**:** 397-406.

37. Grinberg M, Schwarz M, Zaltsman Y, Eini T, Niv H, Pietrokovski S *et al.* Mitochondrial carrier homolog 2 is a target of tBID in cells signaled to die by tumor necrosis factor alpha. *Mol Cell Biol* 2005; **25**(11)**:** 4579-4590.

38. Schafer M, Brauer AU, Savaskan NE, Rathjen FG, Brummendorf T. Neurotractin/kilon promotes neurite outgrowth and is expressed on reactive astrocytes after entorhinal cortex lesion. *Mol Cell Neurosci* 2005; **29**(4)**:** 580-590.

39. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC *et al.* A genome-wide association study of global gene expression. *Nat Genet* 2007; **39**(10)**:** 1202-1207.

40. Chen Y, Zhu J, Lum PY, Yang X, Pinto S, MacNeil DJ *et al.* Variations in DNA elucidate molecular networks that cause disease. *Nature* 2008; **452**(7186)**:** 429-435.

41. Daniels PR, Kardia SL, Hanis CL, Brown CA, Hutchinson R, Boerwinkle E *et al.* Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. *Am J Med* 2004; **116**(10)**:** 676-681.

42. O'Meara JG, Kardia SL, Armon JJ, Brown CA, Boerwinkle E, Turner ST. Ethnic and sex differences in the prevalence, treatment, and control of dyslipidemia among hypertensive adults in the GENOA study. *Arch Intern Med* 2004; **164**(12)**:** 1313-1318.

43. Melton LJ, 3rd. History of the Rochester Epidemiology Project. *Mayo Clin Proc* 1996; **71**(3)**:** 266-274.

44. Falconer DS, Mackay TFC. *Introduction to Quantitative Genetics*: Longman, Harlow, 1996.

45. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet* 2008; **9**(4)**:** 255-266.

46. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, Jr. *et al.* The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA* 2003; **289**(19)**:** 2560-2572.

47. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003; **4**(2)**:** 249-264.

48. Mather WB. *Principles of Quantitative Genetics*. Minneapolis, Minnesota: Burgess Publishing Company, 1964.

49. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998; **62**(5)**:** 1198-1211.

50. Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA *et al.* Individuality and variation in gene expression patterns in human blood. *Proc Natl Acad Sci U S A* 2003; **100**(4)**:** 1896-1901.

51. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J *et al.* Genetics of gene expression and its effect on disease. *Nature* 2008; **452**(7186)**:** 423-428.

52. Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R *et al.* Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* 2010; **5**(5)**:** e10693.

53. Chesler EJ, Lu L, Shou S, Qu Y, Gu J, Wang J *et al.* Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat Genet* 2005; **37**(3)**:** 233-242.

54. Brem RB, Kruglyak L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc Natl Acad Sci U S A* 2005; **102**(5)**:** 1572-1577.

55. Cheung VG, Conlin LK, Weber TM, Arcaro M, Jen KY, Morley M *et al.* Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat Genet* 2003; **33**(3)**:** 422-425.

56. Lynch M, Walsh B. *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates, Inc., 1998.

57. Lanktree MB, Hegele RA. Gene-gene and gene-environment interactions: new insights into the prevention, detection and management of coronary artery disease. *Genome Med* 2009; **1**(2)**:** 28.

58. Cutler DJ, Zwick ME, Carrasquillo MM, Yohn CT, Tobin KP, Kashuk C *et al.* High-throughput variation detection and genotyping using microarrays. *Genome Res* 2001; **11**(11)**:** 1913-1925.

59. Rabbee N, Speed TP. A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics* 2006; **22**(1)**:** 7-12.

60. Affymetrix. Affymetrix® Genome-Wide Human SNP Nsp/Sty 6.0 User Guide. 2007.

61. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics Hum Genet* 2009; **10:** 387-406.

62. Li Y, Ding J, GR. A. Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. *Am J Hum Genet* 2006;(S79)**:** 2290.

63. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; **39**(7)**:** 906-913.

64. Pei YF, Li J, Zhang L, Papasian CJ, Deng HW. Analyses and comparison of accuracy of different genotype imputation methods. *PLoS One* 2008; **3**(10)**:** e3551.

65. Tobin MD, Sheehan NA, Scurrah KJ, Burton PR. Adjusting for treatment effects in studies of quantitative traits: antihypertensive therapy and systolic blood pressure. *Stat Med* 2005; **24**(19)**:** 2911-2935.

66. Scott AJ, Holt D. The Effect of Two-Stage Sampling on Ordinary Least Squares Methods. *Journal of the American Statistical Association* 1982; **77**(380)**:** 848-854.

67. Gamazon ER, Zhang W, Konkashbaev A, Duan S, Kistner EO, Nicolae DL *et al.* SCAN: SNP and copy number annotation. *Bioinformatics* **26**(2)**:** 259-262.

68. Cox NJ. Human genetics: an expression of interest. *Nature* 2004; **430**(7001)**:** 733-734.

69. Qian X, Lu Z, Tan M, Liu H, Lu D. A meta-analysis of association between C677T polymorphism in the methylenetetrahydrofolate reductase gene and hypertension. *Eur J Hum Genet* 2007; **15**(12)**:** 1239-1245.

70. Boushey CJ, Beresford SA, Omenn GS, Motulsky AG. A quantitative assessment of plasma homocysteine as a risk factor for vascular disease. Probable benefits of increasing folic acid intakes. *JAMA* 1995; **274**(13)**:** 1049-1057.

71. Meng W, Mushika Y, Ichii T, Takeichi M. Anchorage of microtubule minus ends to adherens junctions regulates epithelial cell-cell contacts. *Cell* 2008; **135**(5)**:** 948-959.

72. Lin Y, Lai X, Chen B, Xu Y, Huang B, Chen Z *et al.* Genetic variations in CYP17A1, CACNB2 and PLEKHA7 are associated with blood pressure and/or hypertension in She ethnic minority of China. *Atherosclerosis* **219**(2)**:** 709-714.

73. Hong KW, Jin HS, Lim JE, Kim S, Go MJ, Oh B. Recapitulation of two genomewide association studies on blood pressure and essential hypertension in the Korean population. *J Hum Genet* **55**(6)**:** 336-341.

74. Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG. Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet* 2007; **39**(2)**:** 226-231.

75. Cui JS, Hopper JL, Harrap SB. Antihypertensive treatments obscure familial contributions to blood pressure variation. *Hypertension* 2003; **41**(2)**:** 207-210.

76. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B *et al.* The structure of haplotype blocks in the human genome. *Science* 2002; **296**(5576)**:** 2225-2229.

77. http://www.ncbi.nlm.nih.gov/gene/. National Center for Biotechnology Information.

78. Vatner SF. FGF induces hypertrophy and angiogenesis in hibernating myocardium. *Circ Res* 2005; **96**(7)**:** 705-707.

79. He L, Wang B, Hay EB, Nebert DW. Discovery of ZIP transporters that participate in cadmium damage to testis and kidney. *Toxicol Appl Pharmacol* 2009; **238**(3)**:** 250-257.

80. Antzelevitch C, Pollevick GD, Cordeiro JM, Casis O, Sanguinetti MC, Aizawa Y *et al.* Loss-of-function mutations in the cardiac calcium channel underlie a new clinical entity characterized by ST-segment elevation, short QT intervals, and sudden cardiac death. *Circulation* 2007; **115**(4)**:** 442-449.

81. Lao QZ, Kobrinsky E, Harry JB, Ravindran A, Soldatov NM. New Determinant for the CaVbeta2 subunit modulation of the CaV1.2 calcium channel. *J Biol Chem* 2008; **283**(23)**:** 15577-15588.

82. Hinkes B, Wiggins RC, Gbadegesin R, Vlangos CN, Seelow D, Nurnberg G *et al.* Positional cloning uncovers mutations in PLCE1 responsible for a nephrotic syndrome variant that may be reversible. *Nat Genet* 2006; **38**(12)**:** 1397-1405.

83. Arnett DK, Hong Y, Bella JN, Oberman A, Kitzman DW, Hopkins PN *et al.* Sibling correlation of left ventricular mass and geometry in hypertensive African Americans and whites: the HyperGEN study. Hypertension Genetic Epidemiology Network. *Am J Hypertens* 2001; **14**(12)**:** 1226-1230.

84. Monteith GR, Kable EP, Kuo TH, Roufogalis BD. Elevated plasma membrane and sarcoplasmic reticulum Ca2+ pump mRNA levels in cultured aortic smooth muscle cells from spontaneously hypertensive rats. *Biochem Biophys Res Commun* 1997; **230**(2)**:** 344-346.

85. Sayeski PP, Ali MS. The critical role of c-Src and the Shc/Grb2/ERK2 signaling pathway in angiotensin II-dependent VSMC proliferation. *Exp Cell Res* 2003; **287**(2)**:** 339-349.

86. Suh PG, Park JI, Manzoli L, Cocco L, Peak JC, Katan M *et al.* Multiple roles of phosphoinositide-specific phospholipase C isozymes. *BMB Rep* 2008; **41**(6)**:** 415-434.

87. Lewis CE, McTigue KM, Burke LE, Poirier P, Eckel RH, Howard BV *et al.* Mortality, health outcomes, and body mass index in the overweight range: a science advisory from the American Heart Association. *Circulation* 2009; **119**(25)**:** 3263-3271.

88. Huszar D, Lynch CA, Fairchild-Huntress V, Dunmore JH, Fang Q, Berkemeier LR *et al.* Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* 1997; **88**(1)**:** 131-141.

89. O'Rahilly S, Farooqi IS. Human obesity as a heritable disorder of the central control of energy balance. *Int J Obes (Lond)* 2008; **32 Suppl 7:** S55-61.

90. Yang B, He B, Abdel-Halim SM, Tibell A, Brendel MD, Bretzel RG *et al.* Molecular cloning of a full-length cDNA for human type 3 adenylyl cyclase and its expression in human islets. *Biochem Biophys Res Commun* 1999; **254**(3)**:** 548-551.

91. Marg A, Sirim P, Spaltmann F, Plagge A, Kauselmann G, Buck F *et al.* Neurotractin, a novel neurite outgrowth-promoting Ig-like protein that interacts with CEPU-1 and LAMP. *J Cell Biol* 1999; **145**(4)**:** 865-876.

92. Watson P, Townley AK, Koka P, Palmer KJ, Stephens DJ. Sec16 defines endoplasmic reticulum exit sites and is required for secretory cargo export in mammalian cells. *Traffic* 2006; **7**(12)**:** 1678-1687.

93. Monte D, Baert JL, Defossez PA, de Launoit Y, Stehelin D. Molecular cloning and characterization of human ERM, a new member of the Ets family closely related to mouse PEA3 and ER81 transcription factors. *Oncogene* 1994; **9**(5)**:** 1397-1406.

94. Vilarino-Guell C, Wider C, Ross OA, Jasinska-Myga B, Kachergus J, Cobb SA *et al.* LINGO1 and LINGO2 variants are associated with essential tremor and Parkinson disease. *Neurogenetics* 2010; **11**(4)**:** 401-408.

95. Zhu W, Hanes SD. Identification of drosophila bicoid-interacting proteins using a custom two-hybrid selection. *Gene* 2000; **245**(2)**:** 329-339.

96. Hovnanian A. SERCA pumps and human diseases. *Subcell Biochem* 2007; **45:** 337-363.

97. Hiraoka N, Misra A, Belot F, Hindsgaul O, Fukuda M. Molecular cloning and expression of two distinct human N-acetylgalactosamine 4-O-sulfotransferases that transfer sulfate to GalNAc beta 1-->4GlcNAc beta 1-->R in both N- and O-glycans. *Glycobiology* 2001; **11**(6)**:** 495-504.

98. Aravind L, Koonin EV. Fold prediction and evolutionary analysis of the POZ domain: structural and evolutionary relationship with the potassium channel tetramerization domain. *J Mol Biol* 1999; **285**(4)**:** 1353-1361.

99. Hines LM, Stampfer MJ, Ma J, Gaziano JM, Ridker PM, Hankinson SE *et al.* Genetic variation in alcohol dehydrogenase and the beneficial effect of moderate alcohol consumption on myocardial infarction. *N Engl J Med* 2001; **344**(8)**:** 549-555.

100. Cooper RS. Gene-environment interactions and the etiology of common complex disease. *Ann Intern Med* 2003; **139**(5 Pt 2)**:** 437-440.

101. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet* 2009; **10**(6)**:** 392-404.

102. Nelson MR, Kardia SL, Ferrell RE, Sing CF. A combinatorial partitioning method to identify multilocus genotypic partitions that predict quantitative trait variation. *Genome Res* 2001; **11**(3)**:** 458-470.

103. Tsai CT, Lai LP, Lin JL, Chiang FT, Hwang JJ, Ritchie MD *et al.* Renin-angiotensin system gene polymorphisms and atrial fibrillation. *Circulation* 2004; **109**(13)**:** 1640-1646.

104. Bateson W. *Mendel's principles of heredity*. Cambridge University Press: Cambridge, 1909.

105. Fisher RA. The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Roy al Society of Edinburgh* 1981;(52)**:** 399-433

106. Filiano JJ. Neurometabolic diseases in the newborn. *Clin Perinatol* 2006; **33**(2)**:** 411-479.

107. Alvarez-Castro JM, Carlborg O. A unified model for functional and statistical epistasis and its application in quantitative trait Loci analysis. *Genetics* 2007; **176**(2)**:** 1151-1167.

108. McClay JL, van den Oord EJ. Variance component analysis of polymorphic metabolic systems. *J Theor Biol* 2006; **240**(1)**:** 149-159.

109. Siemiatycki J, Thomas DC. Biological models and statistical interactions: an example from multistage carcinogenesis. *Int J Epidemiol* 1981; **10**(4)**:** 383-387.

110. Greenland S. Interactions in epidemiology: relevance, identification, and estimation. *Epidemiology* 2009; **20**(1)**:** 14-17.

111. Niu Y, Gong Y, Langaee TY, Davis HM, Elewa H, Beitelshees AL *et al.* Genetic variation in the beta2 subunit of the voltage-gated calcium channel and pharmacogenetic association with adverse cardiovascular outcomes in the INternational VErapamil SR-Trandolapril STudy GENEtic Substudy (INVEST-GENES). *Circ Cardiovasc Genet* 2010; **3**(6)**:** 548-555.

112. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* 2002; **360**(9349)**:** 1903-1913.

113. Stamler J, Rose G, Stamler R, Elliott P, Dyer A, Marmot M. INTERSALT study findings. Public health and medical care implications. *Hypertension* 1989; **14**(5)**:** 570-577.

114. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 2008; **40**(6)**:** 695-701.

115. Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB *et al.* Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet* 2008; **40**(5)**:** 592-599.

116. Ahituv N, Kavaslar N, Schackwitz W, Ustaszewska A, Martin J, Hebert S *et al.* Medical sequencing at the extremes of human body mass. *Am J Hum Genet* 2007; **80**(4)**:** 779-791.

117. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 2011; **89**(1)**:** 82-93.

118. Law MR, Wald NJ, Morris JK, Jordan RE. Value of low dose combination treatment with blood pressure lowering drugs: analysis of 354 randomised trials. *BMJ* 2003; **326**(7404)**:** 1427.

119. Delaney JA, Oddson BE, Kramer H, Shea S, Psaty BM, McClelland RL. Baseline depressive symptoms are not associated with clinically important levels of incident hypertension during two years of follow-up: the multi-ethnic study of atherosclerosis. *Hypertension* 2010; **55**(2)**:** 408-414.

120. Consortium TGP. A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**(7319)**:** 1061-1073.

121. Huang J, Ellinghaus D, Franke A, Howie B, Li Y. 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. *Eur J Hum Genet* 2012; **20**(7)**:** 801-805.

122. Davies RW, Wells GA, Stewart AF, Erdmann J, Shah SH, Ferguson JF *et al.* A genome-wide association study for coronary artery disease identifies a novel susceptibility locus in the major histocompatibility complex. *Circ Cardiovasc Genet* 2012; **5**(2)**:** 217-225.

123. Neale BM, Rivas MA, Voight BF, Altshuler D, Devlin B, Orho-Melander M *et al.* Testing for an unusual distribution of rare variants. *PLoS Genet* 2011; **7**(3)**:** e1001322.

124. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* 2008; **83**(3)**:** 311-321.

125. Zawistowski M, Gopalakrishnan S, Ding J, Li Y, Grimm S, Zollner S. Extending rare-variant testing strategies: analysis of noncoding sequence and imputed genotypes. *Am J Hum Genet* 2010; **87**(5)**:** 604-617.

126. Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, Wei LJ *et al.* Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet* 2010; **86**(6)**:** 832-838.

127. Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet* 2009; **5**(2)**:** e1000384.

128. Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* 2008; **6**(5)**:** e107.

129. Melzer D, Perry JR, Hernandez D, Corsi AM, Stevens K, Rafferty I *et al.* A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet* 2008; **4**(5)**:** e1000072.

130. Timpson NJ, Lawlor DA, Harbord RM, Gaunt TR, Day IN, Palmer LJ *et al.* C-reactive protein and its role in metabolic syndrome: mendelian randomisation study. *Lancet* 2005; **366**(9501)**:** 1954-1959.

131. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; **32**(1)**:** 1-22.