CANDIDATE GENE ASSOCIATION STUDY OF BASELINE AND LONGITUDINAL BONE-QUALITY TRAITS IN A HEALTHY OLDER POPULATION

by

Ankur Mukherjee

B.Sc. Statistics, St. Xaviers' College, India, 2001

M.Sc. Statistics, Calcutta University, India, 2003

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Ankur Mukherjee

It was defended on

Aug 27, 2008

Dissertation Advisor:
Candace M. Kammerer, PhD
Associate Professor, Department of Human Genetics
Graduate School of Public Health, University of Pittsburgh

Daniel E. Weeks, PhD
Professor, Departments of Human Genetics and Biostatistics
Graduate School of Public Health, University of Pittsburgh

Robert E. Ferrell, PhD
Professor, Departments of Human Genetics and Pharmaceutical Sciences
Graduate School of Public Health and School of Pharmacy, University of Pittsburgh

Joseph M. Zmuda, PhD, MPH
Assistant Professor, Departments of Epidemiology and Human Genetics
Graduate School of Public Health, University of Pittsburgh

Candace M Kammerer, PhD

CANDIDATE GENE ASSOCIATION STUDY OF BASELINE AND LONGITUDINAL BONE-QUALITY TRAITS IN A HEALTHY OLDER POPULATION

Ankur Mukherjee, PhD

University of Pittsburgh, 2008

Motivation: Areal bone mineral density (BMD) is one of the major risk factors for osteoporosis, a public health concern in the US and other countries. The goal of our study was to identify genes that influenced areal BMD in a population of older, originally healthy, African American and Caucasian American individuals.

Methods: We used three genetic association methods (single SNP-single trait, single SNP-multiple trait, multiple SNP-single trait) to test for association between measures of BMD at three time points (2 years apart) and genotype data on 1439 single nucleotide polymorphisms (SNPs) in 138 candidate genes. We first developed a model to determine the genetic (SNP) coverage of our candidate genes, and then we assessed possible population within our populations, two factors that influence the power of association studies. We also investigated the effect of covariates on BMD traits using both cross-sectional as well as longitudinal methods, and used the BMD residuals from these analyses in our association studies.

Results: The SNP coverage of our candidate genes was reasonable, 52.99% compared to the theroetical HAPMAP coverage of 55.8%. We also classified our Caucasian American and African Americans based on genetic ancestry and controlled for subtle substructure. We detected several associations between candidate genes and BMD traits in all the four groups, but the most significant and consistent result was obtained in Caucasian American males. Five SNPs in the GNRHR locus were significantly associated

with hip BMD trait using both the single SNP association approaches, as well as the pathway based analysis. These results need to be followed up in additional populations.

Public Health Significance: As the world population ages, the cost, rate of mortality and morbidity of osteoporosis is also increasing. Identification of genes that influence risk of developing osteoporosis may help identify people at risk, as well as facilitate development of drugs and other measures to mitigate the effects of this disease.

TABLE OF CONTENTS

PREFACE ..... **XV**

1.0 INTRODUCTION **1**

1.1 PUBLIC HEALTH SIGNIFICANCE OF OSTEOPOROSIS **1**

1.2 ASSESSING BONE MASS **2**

1.3 RISK FACTORS FOR OSTEOPOROSIS **3**

1.4 GENETICS OF BONE MINERAL DENSITY AND/OR OSTEOPOROSIS SUSCEPTIBILITY **5**

1.5 STATISTICAL CONCERNS WITH ASSOCIATION STUDIES **6**

1.5.1 Population substructure **7**

1.5.2 Coverage of candidate genes **7**

1.5.3 Multiple testing **7**

1.5.4 Power of a study **8**

1.5.5 Longitudinal analyses **8**

1.5.6 Gene x Gene interaction **9**

1.6 HEALTHY AGING AND BODY COMPOSITION STUDY **9**

1.7 SPECIFIC AIMS **11**

2.0 PREDICTING GENE COVERAGE: HOW MANY SNPS ARE ENOUGH? **14**

2.1 ABSTRACT **14**

2.2 INTRODUCTION 14
2.3 MATERIAL AND METHODS 17
2.4 RESULTS 20
2.5 DISCUSSION 23
2.6 TABLES AND FIGURES 26
3.0 ASSESSING ANCESTRY AND POPULATION SUBSTRUCTURE AMONG THE HEALTH ABC COHORT 33
3.1 INTRODUCTION 33
3.2 DATA 34
3.2.1 Identification of Candidate Genes and SNPs 34
3.2.2 Data cleaning 35
3.3 METHODS 36
3.3.1 Eigen analysis methods 36
3.4 RESULTS 38
3.4.1 Exploring ancestry estimation using the eigen analysis method 38
3.4.2 Assigning Individuals to Ancestry Groups 39
3.4.3 Assessing Eigenvectors within Groups 40
3.5 DISCUSSION 41
3.6 TABLES AND FIGURES 43
4.0 ANALYZING THE EFFECTS OF COVARIATES ON BMD ACROSS DIFFERENT ETHNIC AS WELL AS GENDER GROUPS 49
4.1 INTRODUCTION 49
4.2 DATA AND METHODS 50

4.3 RESULTS 53

4.3.1 Results from analyses of covariates at baseline 54

4.3.2 Proportion of total phenotypic variation attributable to different covariate categories 57

4.3.3 Results from longitudinal analysis of BMD 58

4.4 DISCUSSION 58

4.5 TABLES AND FIGURES 61

5.0 ANALYSES OF POSSIBLE ASSOCIATIONS BETWEEN CANDIDATE GENES POLYMORPHISMS AND BONE MINERAL DENSITY TRAITS: SINGLE SNP-SINGLE TRAIT VERSUS SINGLE SNP -MULTIPLE TRAIT COMPARISONS. 69

5.1 INTRODUCTION 69

5.2 DATA AND ANALYTICAL METHODS 72

5.2.1 Health ABC populations 72

5.2.2 MrOS population 73

5.2.3 Single SNP-Single Trait Analyses 74

5.2.4 Single SNP-Multiple Traits Analyses: Principal Component of Heritability, PCH 75

5.2.5 Regression analyses 77

5.2.6 Haplotype Analyses 78

5.3 RESULTS 78

5.3.1 Single SNP-Single Trait analyses 78

5.3.2 Single SNP-Multiple Trait analyses 80

5.3.3 Relationship Between Individual GNRHR SNPs on BMD traits in Caucasian American Males 81
5.3.4 Haplotype Analyses of GNRHR SNPs 83
5.4 DISCUSSION 85
5.5 TABLES AND FIGURES 89
6.0 ASSOCIATION ANALYSIS OF BMD TRAITS WITH GENES IN COMMON PATHWAYS 119
6.1 INTRODUCTION 119
6.2 DATA 121
6.3 METHOD 122
6.3.1 Construction of the Pathway Annotation File 122
6.3.2 Construction of SNP to Gene Correspondence File 123
6.3.3 Gene-Enrichment or Pathway-based Association Analysis Statistic 124
6.4 RESULTS 127
6.4.1 Characteristics of the pathways 127
6.4.2 Pathway-based Associations for Six BMD Traits among the HABC Cohort 127
6.4.3 Exploration of Pathway-based Association Statistics 129
6.4.4 Pathway Association Analyses in MrOS – a Replicate Cohort 129
6.5 DISCUSSION 130
6.6 TABLES AND FIGURES 134
7.0 CONCLUSION 140
7.1 MOTIVATION 140

7.2 SUMMARY OF RESULTS 142

7.2.1 SNP coverage 142

7.2.2 Assessing population ancestry 142

7.2.3 Estimating effects of covariates 143

7.2.4 Association analysis results 143

7.3 SIGNIFICANCE OF THE STUDY 145

7.4 FUTURE WORK 145

APPENDIX: LIST OF ALL SNPS, COVARIATES AND SIGNIFICANT SNPS BY GROUPS 147

BIBLIOGRAPHY 169

LIST OF TABLES

Table 2.1:The number SNPs/gene 173 genes in common in the HAPMAP and SeattleSNP databases 26

Table 2.2:Description of the seven training models used in the cross-validation approach to predict dSNP and the mean square error obtained from the seven testing sets 26

Table 2.3: Description of the Final Model used to Estimate dSNP in the MrOS Data Set 26

Table 3.1: Details of genes studied in HABC cohort..... 43

Table 3.2: Table of self reported and genetically assigned ancestry groups 46

Table 4.1: Characteristics of the Covariates in HABC men and women by ethnic group measured at baseline..... 61

Table 4.2: Table of mean and standard deviation of raw BMD traits at baseline for the four sex by ancestry groups in HABC data 62

Table 4.3: Association of Covariates of BMD traits at baseline for four ancestry by sex groups 63

Table 4.4: Variation explained (%) by different covariate groups for BMD traits in four ancestry by sex groups 65

Table 4.5: Longitudinal Analysis of Covariates with BMD traits..... 66

Table 5.1: Results from single SNP single trait analysis in time points 1, 3 and 5 89

Table 5.2: Results from PCH analysis in time points 1, 3 and 5 90

Table 5.3: P-values of equality of slope test between three time points. Hip and Hip Neck BMD traits for SNPs from GNRHR gene..... 91

Table 5.4: Results from haplotype analysis of 5 significant SNPs from GNRHR gene on hip and femoral neck BMD in Caucasian male 92

Table 5.5: Results from MrOS analysis on baseline significant SNPs from Caucasian males in HABC cohort 93

Table 5.6: Allele frequency of GNRHR SNPs in the two populations Caucasian American Males in HABC and MrOS..... 94

Table 6.1: Characteristics of Ingenuity-Derived Pathways Using HABC Gene List 134

Table 6.2: Pathway Association with HABC BMD measures at baseline 135

Table 6.3: Pathway Association with HABC BMD measures at time point 3 136

Table 6.4: Pathway Association with HABC BMD measures at time point 5 137

Table 6.5: Pathway Association with MrOS BMD measures (using only 138 MrOS genes) 137

Table 6.6: Pathway Association with MrOS BMD measures (using all 393 MrOS genes) 137

Table 6.7: Table with association results for all the pathways for Hip BMD in HABC data for Caucasian Males at time point 1 and time point 3 138

Table A 1: Association of Covariates with the other nine BMD traits at baseline for four ancestry by sex groups. 147

Table A 2: Table with top 50 SNPs (according to p-value) for single trait single SNP analysis for the four ancestry by sex groups 151

Table A 3: SNPs used in the study..... 159

LIST OF FIGURES

Figure 2.1: Flow Chart of the approach used to develop the CART model 27

Figure 2.2: Distribution of Transformations of average R^2 28

Figure 2.3: Flow Chart of the cross validation approach to obtain estimates of density of total number of SNPs/ 10kbp (dsnp) 29

Figure 2.4: A plot of tag SNPs per gene selected from the HapMap database for $\eta = 0.5$ by the number of total SNPs present in HAPMAP database 30

Figure 2.5: Histogram of Average R^2 for predicting unmeasured SNPs within genes by the number of tag SNP per gene: 1-2 tag SNPs, 3-4 tag SNPs, and ≥ 5 tag SNPs 31

Figure 2.6: CART model for predicting coverage of HAPMAP tag SNPs 32

Figure 3.1: Histogram for minor allele frequencies in the two self-reported race groups 46

Figure 3.2: Plot of second and third eigen vectors when SNPs are correlated 47

Figure 3.3: Estimated genetic ancestry (represented by the first two eigen analysis components) versus self-reported ancestry 47

Figure 3.4: Pairwise plots of first three principal components for individuals in the African American ancestry group 48

Figure 3.5: Pairwise plots of first three principal components for individuals in the Caucasian American ancestry group 48

Figure 4.1: Mean BMD across six BMD sites for three time points in all four sex by ancestry group 68
Figure 5.1: Plot of –log10(P) value for each trait (single SNP single trait analysis time point 1) 98
Figure 5.2: Plot of –log10(P) value for each trait (single SNP single trait analysis time point 3) 102
Figure 5.3: Plot of –log10(P) value for each trait (single SNP single trait analysis time point 5) 106
Figure 5.4: Plot of –log10(PCH) value for each group (single SNP multiple trait analysis) 110
Figure 5.5: Plot of residual mean of Hip and Femoral Neck BMD for genotypes of each significant SNP in the GNRHR gene for Caucasian Males 115
Figure 5.6: LD structure of SNPs from GNRHR gene 116
Figure 5.7: Plot of mean BMD at hip and femoral neck skeletal sites across different groups of diplotypes (using the five significant SNPs in GNRHR) for Caucasian American males 117
Figure 5.8: Plot of residual mean of Hip BMD for genotypes of each significant SNP in the GNRHR gene for MrOS population 118
Figure 6.1: Pathway 2 (involving 18 genes from HABC list) 139

PREFACE

At the beginning of my dissertation I would like to express my sincere appreciation to all who made my research work successful. My sincere gratitude goes to my PhD committee chair and mentor Dr. Candace Kammerer, for her continuous encouragement and other committee members Dr. Daniel Weeks, Dr. Robert Ferrel and Dr. Joseph Zmuda. I was enriched by their wise advice and thoughtful guidance throughout my research work.

I would like to thank all the participants of Health ABC study, as well as NIH for their support in our training program.

I would like to thank Dr. Eleanor Feingold, Dr. Michael Barmada, Dr. Kathryn Roeder, Dr. Lambertus Klei and Dr. Bernie Devlin for their guidance and help in my research.

I am also thankful to my current and previous lab mates John Shaffer, Xiao Jing Wang, Amy Dressen, Margarate Kenny for their friendly support and assistance.

I would like to thank Chriss, Ryan for their support in computing and also Jeanette, Michale, Gloria and Debbie for their support in my paper works.

I would like to thank Indranil and Samsiddhi for their help. I would also like to thank all my friends in Pittsburgh.

I am deeply grateful to my family and friends, specially my wife Shaswati and my mother without whom my dissertation would not be possible.

1.0 INTRODUCTION

1.1 PUBLIC HEALTH SIGNIFICANCE OF OSTEOPOROSIS

Osteoporosis, a growing public health concern, is a skeletal condition that is characterized by decline in bone mass and microarchitectural deterioration of bone tissue. This decline may result in increased bone fragility and susceptibility to fractures (Consensus Development Conference: Diagnosis, prophylaxis and treatment of osteoporosis, 1993). Bone mineral density (BMD), which is a quantifier of bone mass, is an accurate indicator of bone strength. Based on the definition by World Health Organization [WHO], individuals with bone mineral density [BMD] values that lie between 1 and 2.5 Standard Deviation [SD] below the young adult mean value are classified as having osteopenia. Whereas if the BMD value is 2.5 SD below the young adult mean value, then the individuals are classified as having osteoporosis (1994). It is estimated that 10 million Americans will develop osteoporosis each year and another 34 million are at risk (Report of the Surgeon General, US department of Health and Human services, 2004). Osteoporosis can also be defined as the fracture incidence, especially hip fracture, the most common form of fractures. In most populations the incidence of hip fracture increases with age. In people over 50 years of age, the female to male ratio of hip fracture is 2:1 (Melton, 1988), and one in two women and one in five men will experience an osteoporotic fracture during the remaining years of their lives.

Osteoporotic fracture has a significant effect on one's health and can lead to mortality (Dennison et al. 2005). Hip fracture mortality is higher in men than in women, and increases with age (Cooper et al. 1993). In a population from Minnesota, the survival rate 5 years after hip or vertebral fracture was 80% of

the rate of similarly aged men and women having not experienced a fracture (Cooper et al. 1993). In the USA, 8% of men and 3% of women >50 years of age die during hospitalization for hip fractures, and 7% of survivors from any kind of fracture have some permanent disability (Chrischilles et al. 1991).

The financial costs regarding osteoporosis are also huge. The annual cost of hip fractures in the USA alone may reach $240 billion in about 40 years (Raisz 1997). The incidence of osteoporosis is expected to increase with time, due to the aging population, thus the costs regarding this major public health problem are also expected to increase.

1.2 ASSESSING BONE MASS

There are two quantitative radiologic methods commonly used in assessing bone mass: dual-energy X-ray absorptiometry (DXA) and quantitative computed tomography (QCT).

DXA is a two-dimensional projection of 3D bone structure, and therefore is a function of the size, volume and mineral content of a bone (Carter et al. 1992). It gives good precision in measuring components of body composition, like fat and bone mineral density (Lohman et al. 1996; Hagino 2004), which makes DXA suitable for longitudinal studies. This non-invasive technique can be applied to both patients as well as healthy individuals and the radiation dose is very small. DXA can also be done very quickly, many epidemiological studies have obtained DXA measures of BMD over the years, including National Health and Nutrition Examination Survey (NHANES) (Looker et al. 1998); The Study of Osteoporotic Fractures (SOF) (Hosmer et al. 2002); The Osteoporotic Fractures in Men (MrOS) (Orwoll et al. 2005) and family studies (The Framingham Osteoporosis Study (Hannan et al. 2000); Indiana Sisters Study (Hui et al. 2006); The San Antonio Family Osteoporosis Study (SAFOS) (Mitchell et al. 2003).

QCT measurements are obtained using a standard CT scanner at any skeletal site. QCT assesses both the volume and the density of bone and is widely accepted and used for measuring vertebral cancellous bone (Guglielmi et al. 2002). QCT can determine true volumetric density and three-dimensional imaging of bone, thus providing insight into the true architecture of bone, which is an advantage over two-dimensional DXA. However the radiation level in QCT is 40-90 μSv compared to 5.4 μSv (Kalender 1992) for DXA in whole skeleton scan. Although QCT measures are likely to be better measures of bone architecture, DXA measures were analyzed in the current study because DXA measures were available on more individuals and at multiple time points.

1.3 RISK FACTORS FOR OSTEOPOROSIS

Low BMD is a key component of osteoporosis. The relationship between the ability to withstand trauma and BMD is well established for an individual who had previous fracture and/or fall. In other words much lower force is needed to get a fracture for an individual with low BMD whereas for individuals with higher BMD the level of force is relatively higher. 75-90% of the variation in bone strength is related to BMD (Jordan et al. 2002).

Advanced age is one of the most important risk factors for osteoporosis. Bone traits change over time because bone, as a dynamic trait, is constantly being remodeled, and the opposing processes of bone formation and bone resorption change over time (Blair et al. 1993). For example, in childhood, bone formation activity is higher than bone resorption activity, so bone mass is gained (Zhang et al. 2007). At advanced ages, it is reversed so bone mass is lost. Moreover, the alteration of bone formation differs in different bone compartments, such as cortical and trabecular bone (Judex et al. 2004; Riggs et al. 2004).

Sex and ethnicity are two additional important risk factors for osteoporosis. Women generally tend to lose bone mass more rapidly after menopause and Caucasian Americans are at increased risk (Wilkins et al. 2005). On average for women, older African-Americans have higher levels of BMD than whites (Evans et al. 2005), and lose BMD less rapidly with increasing age (Cauley et al. 2005). This leads to higher risk of fracture among Caucasian American women. Furthermore, average areal BMD in tibia differs among men and women as well as between age groups above and below 60 years (Khodadadyan-Klostermann et al. 2004). In the Healthy Aging and Body Composition [HABC] study population, both U.S. Caucasian American and African American men have higher volumetric BMD than their women counterparts (Taaffe et al. 2003).

Other environmental factors that increase bone loss, especially among elderly individuals (> 50 years), include smoking, low intake of calcium, low BMI, lack of physical activity, alcohol consumption, and presence of insulin dependent diabetes (Jordan et al. 2002). The effects of covariates may also differ by age and sex. For example, in 45-92 year old individuals, serum leptin levels are associated with BMD in women but not in men (Weiss et al. 2006). Likewise, postmenopausal women have a positive association of cyclooxygenase-2 use and BMD but the association is reversed in older men (Richards et al. 2006).

Family history of osteoporosis is also a risk factor (Wilkins et al. 2005). More than 30 years ago it was recognized that measures of bone mass (e.g. bone density) were more similar between monozygotic twins than between dizygotic twins (Dequeker et al. 1987; Pocock et al. 1987). This result indicates the heritability of bone mass traits. Few studies have been done to analyze the heritability of osteoporotic fractures. The risk of wrist fracture has been shown to be significantly heritable in 2500 Midwestern U.S. women (Deng et al. 2000) and in >6500 twins (Andrew et al. 2005). Likewise, osteoporotic fracture has been shown to be highly heritable [h^2= 0.48 for hip fracture] in a Swedish twin cohort born before 1944 (Michaelsson et al. 2005). BMD is also a highly heritable trait. At multiple skeletal sites in the peripheral and axial skeleton, 66-75% of variation in BMD is explained by genetic factors (Dequeker et al. 1987; Pocock et al. 1987). In a study on Amish

population, 37-54% of total variation in post-menopausal women at hip and spine BMD sites is explained by genetic factors (Brown et al. 2005). The loss in BMD level is also found to be heritable (80%) at spine among adults aged 23-75 years (Kelly et al. 1993).

Taken together, these studies clearly show that osteoporosis is a genetic disorder and that additional work is needed to elucidate the specific genetic variants causing this disease.

1.4 GENETICS OF BONE MINERAL DENSITY AND/OR OSTEOPOROSIS SUSCEPTIBILITY

Because of the importance of family history and ethnicity, many studies, including genome-wide linkage scans, have been performed to identify genes that influence BMD and/or risk of developing osteoporosis. Several studies have reported a QTL on chromosome 11q for femoral neck or spine BMD in Caucasian Americans (Johnson et al. 1997; Koller et al. 1998; Shen et al. 2004). Additionally, QTL signals have been found on chromosome 13 in Caucasian Americans as well as Mexican Americans for spine, femoral neck and trochanter BMD (Kammerer et al. 2003). Also, QTLs for wrist and femoral neck BMD have been found on the X chromosome (Shen et al. 2004). Several other QTLs across the genome have also been reported (Econs et al. 2004; Huang et al. 2004; Devoto et al. 2005; Rubin et al. 2007; Wang et al. 2007; Willaert et al. 2008; Zhang et al. 2008). The apparent inconsistencies of the linkage studies may be due to the complexity of the genetic nature of BMD traits and also differences between the studies, such as study population, ascertainment criteria, sample size, and family size. Overall, these results support the notion that osteoporosis is a complex multifactorial disease controlled by several genetic and environmental factors.

Association studies to determine the genetic effects of candidate genes on bone phenotypes have focused mainly on genes that are involved in bone regulation. Growth factors/receptors, cytokines, sex

hormones/receptors, metabolic pathways, calciotropic hormone/receptors, adhesion molecules and ligands, collagenic and non-collagenic proteins have been at the center of such association studies (Gennari et al. 2002).

Specific genes that have been considered in previous association studies are: estrogen receptor (Gennari et al. 1998; Duncan et al. 1999; Ioannidis et al. 2002), vitamin D receptor (Gennari et al. 1998; Duncan et al. 1999; Gennari et al. 1999; Zmuda et al. 1999; Brown et al. 2001; Masi et al. 2002), PTH receptor type 1 (Duncan et al. 1999), collagen type 1 (Hustmyer et al. 1999; Duncan et al. 1999; Gennari et al. 1999; Zmuda et al. 1999; Brown et al. 2001), interleukin 6 (Ota et al. 1999; Zmuda et al. 1999; Takacs et al. 2000; Ota et al. 2001), tumor necrosis factor- alpha (Ota et al. 2000; Ota et al. 2002), low density lipoprotein receptor-related protein 5 (Koay et al. 2004) insulin like growth factor (Takacs et al. 1999), and alpha 2HS glycoprotein (Zmuda et al. 1998). The findings of these association studies are inconsistent as some of the studies found association and some did not, possibly due to population differences, population substructure/admixture, poorly selected controls that may have led to false positives, or selection bias (Gennari et al. 2002). A recent genome wide association study identified two SNPs in LRP5 and TNFRSF11B at femoral neck and lumbar spine among Caucasian American females (Richards et al. 2008). Another genome wide association study found association between SNPs at ESR1, RANKL and OPG gene among 5861 Icelandic subjects at lumbar spine and hip (Styrkarsdottir et al. 2008).

1.5 STATISTICAL CONCERNS WITH ASSOCIATION STUDIES

There are several concerns regarding genetic association studies and a lot of research is being done in these areas. Some of the concerns are:

1.5.1 Population substructure

Some results of association studies may be incorrect because of population substructure; for example, there is a possibility that we may detect a false positive or miss a true association due to population substructure. Several methods are available to control for population substructure including eigen value estimation and the Transmission Disequilibrium Test (TDT). TDT uses family-based pseudo-controls, so that the issue of population substructure never arises (Spielman et al. 1993). However not all traits are amenable to using family-based approaches and this methodology would preclude use of data on large population-based studies. Therefore, researchers have recently developed methods such as eigen analysis, which uses data on all genotypes, and enables the estimation of possible population substructure of a particular study population. It also allows for a comparison of possible substructure between different groups, e.g. cases versus controls. The estimated substructure may then be incorporated into the association analysis (Patterson et al. 2006; Yu et al. 2008).

1.5.2 Coverage of candidate genes

If much of the genetic variation within a gene is not assessed, that is the gene is poorly covered, or the SNPs being tested are not in LD with the 'causal' SNP, it will be difficult to detect a true genetic association. Therefore, the genes of interest should be well covered with genotyped SNPs. HapMap (www.hapmap.org) and SeattleSNPs (http://pga.gs.washington.edu/) are common sources that can be used to obtain sufficient coverage of the genes of interest.

1.5.3 Multiple testing

Another main concern in association analysis is multiple testing. Testing each SNP individually with each phenotype results in lots of tests and many false positives. The Bonferroni method (Simes 1986) of

adjusting for multiple testing is one of the popular methods, but it is overly conservative. So we can adjust for false positives but we might miss some true positives also (Yongchao et al. 2003), The same is true for Sidak (Sidak 1967), an additional multiple testing correction method. Another method to counter the multiple testing problem is to perform dimension reduction techniques, such as principal components. By testing the principal components of a group of related phenotypes, we can reduce the total number of tests, and thereby lessen the effect of multiple testing (Nyholt 2004). Moreover, other methods of analyzing the significance of the findings of an association study, such as permutation tests (Ritchie et al. 2001) and false discovery rates (FDR) (Benjamini et al. 1995) are available. However, the gold standard for determining whether associations are likely to be real is to perform a replication analysis. However, there may be questions regarding what constitutes a true 'replication'.

1.5.4 Power of a study

Power of an association study depends on several aspects. The sample size of the study, allele frequency of the markers tested, LD between the markers genotyped and the actual susceptibiltiy locus (Fisher et al. 2008).

1.5.5 Longitudinal analyses

The main concern in longitudinal analysis is that the correlation structure of the data must be taken into account before doing the analysis. Exploratory analysis like lowess curve and smoothing spline fitting are used to gain an idea about the correlation structure of the data. The generalized estimating equations (GEE) method takes into account the correlation and the missingness of the data over different time points.

1.5.6 Gene x Gene interaction

Gene x gene interaction is an important part in genetic analysis of a complex trait. Simple regression based methods (both linear and logistic) can be used for analysis, but testing the interaction effect of several SNPs from multiple genes that are affecting common biological pathways is problematic because of the number of tests being performed. However, there are other problems in detecting the gene x gene and gene x environment interactions. Because variation associated with the interaction is also subsumed into the variation explained by the main effects in the model, traditional linear modeling strategies have very low power to detect interactions (Cheverud et al. 1995). When testing for two or more factor interactions, small or empty cells (for specific combinations) may result, and these cells can also affect the estimation and robustness of the model (Searle 1987). Because of a large increase in the number of analyses to be done and simultaneous increase in type I error, analyzing all important interaction factors is not feasible (e.g. for 10 SNPs, there are 45 possible two-factor interactions and for 15 SNPs there are 105 possible two-factor interactions).

We assessed possible gene x gene interactions from slightly different angle. We first constructed some common pathways using Ingenuity pathway analysis, using all our candidate genes. The pathways are developed using some common biological functions. Instead of just looking at single SNP/gene, we looked at the effect of those pathways as a whole (Wang et al. 2007). We wanted to see if the genes within a specific pathway are overrepresented among the significant set of genes. We used a variation of the gene-set enrichment analysis algorithm (GSEA) (Subramanian et al. 2005) for these analyses.

1.6 HEALTHY AGING AND BODY COMPOSITION STUDY

The Health, Aging and Body Composition Study (HABC) is a large (N=3075) prospective cohort study of changes in body composition in well-functioning older Caucasian American and African-American men

and women. The data were collected on different body composition phenotypes in bone, muscle and fat. We analyzed data on DXA measures of bone phenotypes in this dissertation. HealthABC is a longitudinal cohort study; measurements were obtained once per year over 6 years starting in 1996-97. The primary goal of Health ABC study was to identify factors associated with incident disability and decline in mobility in healthier older persons, with an emphasis on the role of changes in body composition. The HABC cohort is aged between 68-79 years at baseline.

The data was collected from two sites Memphis, Tennessee and Pittsburgh, Pennsylvania. Data on several anthropometric, demographic, medications and lifestyle covariates were also collected and are described in more detail below.

Anthropometry: Height and weight were estimated using a stadiometer (Harpenend, Wales, UK) and balance beam scale, respectively, without wearing any shoes. Body mass index (kg/m^2) was also calculated as one of the measures of body composition. Waist circumference was also measured in cm. The difference between sitting height and 30 cm seat was denoted as trunk length and the difference between height and trunk length was defined as leg length.

Physical Activity: was assessed by self-report of leisure activities and exercise. Exercise level (kcal/kg/week) was estimated based on a questionnaire administered by the interviewer.

Prevalent Diseases and Medications: Data on different medications and prevalent disease status were obtained from self-reports by the participants. Data on medications like vitamin D supplement, oral estrogen, calcium supplement, any osteoporosis drug, were obtained. Data on several prevalent diseases like cancer, osteoporoses by T-score, high blood pressure and cardiovascular disease were obtained by self-report. Prevalent diabetes status was also obtained.

Life Style Activities: Measures on several lifestyle activities like drinking and smoking history were obtained. Drinking and smoking history were classified as never, current and former. For smoking history pack years were also estimated.

For all the participants self reported race was recorded.

Genetic Data: Genotypes were assayed on 1536 SNPS among 138 candidate genes (including gene clusters, in which more than two genes are present).

Phenotype Data: BMD data from the DXA scans were obtained on 15 skeletal sites at all six time points. However, in this project we analyzed data collected at baseline (time point 1), time point 3 and time point 5. Data was available for bone mineral density, content and area separately.

1.7 SPECIFIC AIMS

Osteoporosis is one of the major health problems in America and osteoporotic fractures are a major cause of morbidity and mortality among older individuals. However, among older individuals (> 65 years), very little is known regarding the effect of different covariates on BMD traits over different time points or the effect of polymorphisms at several candidate genes on bone traits. Even less is known about possible gene-gene interactions on baseline and longitudinal measures of bone traits among originally healthy individuals who are > 68 years old. The overall goal of the current proposal was to investigate whether specific candidate genes are associated with DXA BMD measures at baseline and at two other timepoints in the elderly. Bone phenotype data were available on 2979 unrelated men and women (68-80 years of age) of both Caucasian and African American ancestry at baseline and at four additional timepoints. Genotypes were available on a total of 1439 SNPs across 138 candidate genes that fall into three general categories: (1) genes involved in steroid hormone production, metabolism and action, (2) growth factor related genes, and (3) cytokine related genes.

To determine whether variation in specific candidate genes was associated with BMD in older Caucasian American men and women, as well as older African American men and women, we performed the following specific aims.

AIM 1: Assess SNP coverage of candidate genes and ancestry of individuals.

The accuracy and applicability of the results of the association studies depend critically on genotype quality, SNP coverage of candidate genes, and possible population substructure.

1A. SNP coverage (Chapter 2)

To assess whether a candidate gene is, or is not, associated with a particular bone trait, a large proportion of the total genetic variation in the candidate gene needs to be measured; however, it is not feasible to genotype all variants within a gene. The genetic coverage of each of the candidate genes was determined by developing a classification and regression tree model (CART) using data from HapMap and SeattleSNPs. Subsequently, this model was used to determine genetic coverage of the candidate genes in the current study.

1B. Ancestry estimation (Chapter 3)

Ancestry information for each individual was estimated using eigen analysis of genotype data. Based on these analyses, individuals were assigned to the Caucasian American or African American ancestry groups. Subsequently, a set of eigen vectors was identified within each ancestry group to control for possible substructures.

AIM 2: **Determine the effect of a variety of covariates on areal BMD traits in a population of originally healthy African Americans and Caucasian Americans > 68 yrs old using univariate and multivariate (longitudinal) analyses. (Chapter 4).**

Data were available on areal (DXA) bone mineral density (BMD) for several skeletal sites including hip, spine, and whole body. The relationship between these bone traits and a variety of covariates including demographic (e.g., age, sex), anthropometric (e.g., height, weight), lifestyle (e.g., smoking, drinking, physical activity), and medical (e.g. diabetes, etc) traits was assessed.

AIM 3: **Test for associations between candidate gene polymorphisms and BMD variation using three different association methods and try to replicate any significant results using the MrOS cohort. The three association methods were:**

3A: Single SNP-Single trait association using simple regression analysis (Chapter 5).

3B: Single SNP-Multiple trait association using principal components of heritability (PCH) analysis (Chapter 5).

3C: Multiple SNP-Single Trait association using pathway-based enrichment scores (Chapter 6).

2.0 PREDICTING GENE COVERAGE: HOW MANY SNPS ARE ENOUGH?

2.1 ABSTRACT

Numerous studies are going on to discover causal genes for different complex human diseases. The goal of these studies is to demonstrate linkage and/or association between human polymorphisms and phenotypes. We evaluate the utility of the HAPMAP database in genetic studies by contrasting it with a database containing fully sequenced regions of the genome. By selecting an independent set of SNPs, that is, tag SNPs, we estimate how well tag SNP genotypes predict other SNP genotypes in a region of interest. HAPMAP contains several SNPs in high linkage disequilibrium (LD). So tag SNP selection drops almost half of the SNPs. We based our study on European-ancestry population only, and the average prediction of unmeasured SNPs is quite high (55.8%). Even if there are only a few SNPs available from HAPMAP for the region, the prediction is quite good (55.15%). We also investigate how well covered the genes are for our own association study and found we would need an average of 3 additional SNPs per gene to get coverage as good as HAPMAP.

2.2 INTRODUCTION

Most common human diseases result from the interplay of genetic and environmental causes, therefore considerable research is being done to identify the genetic components of these diseases (Weedon et al. 2007; Styrkarsdottir et al. 2008). Given the completion of the first stages of the human genome sequence

(Collins et al. 2003) and the identification of single nucleotide polymorphisms (SNP) across the human genome (2003), along with the development of high-throughput, low cost genotyping, the identification of possible genetic causes of disease has become easier. Theoretically, researchers could genotype 10 million (or more) SNPs across the human genome and try to find the causal variants that affect complex diseases, however, it currently is very expensive to genotype so many SNPs. Information on all reported SNPs is maintained in dbSNP, the world's largest database for nucleotide polymorphisms. It is a part of the National Center for Biotechnology Information (NCBI) [http://www.ncbi.nlm.nih.gov]. In contrast, HAPMAP comprises a subset of all possible kinds of genetic variants. The phase I data of HAPMAP only had SNPs with MAF > 5%, but in phase 2 HAPMAP data was aimed at increasing the density of SNPs and tried to genotype rare SNPs also (Consortium 2007). The individuals genotyped for the HAPMAP database include 30 trios (mother, father, and offspring) each from Nigeria (YRI) and Utah (CEU), the latter population had ancestry from northern and western Europe. There are 45 unrelated individuals each from Han Chinese (CHB) and Japanese populations (JPN). The Copy Number Variation (CNV) information is also added in the latest build. The utility of HAPMAP is that it will help researchers identify a subset of 20,000 to 1 million tag SNPs as informative as the 10 million known SNPs. This procedure will reduce the cost of a study substantially (www.hapmap.org).

Several studies have evaluated how well genetic polymorphisms in HAPMAP, especially tagging SNPs, represent the total genetic variation in the human genome. Tag SNPs are SNPs that are in high linkage disequilibrium (LD) with a set of SNPs, and can successfully predict the genotype of that set of SNPs, once we have the genotype information of the tag SNPs. In particular, previous studies were done to estimate how well the HAPMAP tag SNPs predict the unmeasured SNPs within HAPMAP. One of these studies is the National Institute of Environmental Health Sciences Environmental Genome Project (NIEHS) (Tantoso et al. 2006). The genes for NIEHS study were involved in DNA repair, cell cycle regulation, apoptosis (Livingston et al. 2004). The tag SNPs from HAPMAP are transferable to NIEHS SNPs but the coverage for untyped SNPs is around 50% for Asian and European ancestry and is poor for African ancestry (30%). In certain regions of the genome there may be loss of information because of LD

differences between the two populations. So we will lose some information when HAPMAP tag SNPs are transferred to a different population (Mueller et al. 2005). When tested in an Estonian sample, the performance of tag SNPs from CEU sample of HAPMAP is very good for SNPs from two 500kb Encode regions in chromosome 2 (ENr112, ENr131). The tag SNPs captured 90-95% of the variation in the Estonian sample (Montpetit et al. 2006).

In addition to using tag SNPs for genome-wide analyses to detect chromosomal regions that may harbor a causal genetic variant, researchers perform follow-up studies to determine the relationship between genetic variation within specific candidate genes or, genomic regions and the trait or disease of interest. The effectiveness of any candidate gene study depends, in part, on how well tag SNPs identified from HAPMAP represent (or cover) the total genetic variation in these genes. In other words, are tag SNPs from HAPMAP highly correlated with all SNP variation within a gene? Furthermore, is all genetic variation within any specific gene well represented by a set of HAPMAP SNPs? In the current study we evaluated how well the HAPMAP SNP data represent all known SNP variation that is, how well do HAPMAP SNPs cover a gene? We assessed coverage in genes from individuals with European ancestry. We also built a model to help us answer the question how well any specific set of genotyped SNPs represents all SNP variation within a set of genes. To answer the first question, we selected tag SNPs for a set of genes from the HAPMAP database and determined how well these tag SNPs predicted genetic variation in all known SNPs within the genes. Information on all known genetic variation was obtained from the SeattleSNPs database. SeattleSNP is a resequencing project in which complete sequencing data are generated on a sample 24 African American and 23 European American individuals for a specific set of genes. The SeattleSNP database contains data on base-pair sequences (for > 250 genes, most of which are candidate genes for cardiovascular disease). However, not all intronic regions were sequenced in all genes. Finally, we assessed how well a set of tag SNPs from one of our own association studies represents all known genetic variation within the set of candidate genes we assayed.

2.3 MATERIAL AND METHODS

We performed our analyses using SNP genotype data on individuals with European ancestry from two sources: the HAPMAP repository (for our study we used NCBI build 35) and the SeattleSNPs database (Sept, 2006). Data in HAPMAP contains genotypes of 30 parent-child trios of European descent, that is 60 unrelated individuals. Data on 23 individuals of European descent were were available in the SeattleSNPs database. All our analyses were genotype not haplotype based.

Our first question was to determine how well SNPs in the HAPMAP represented all SNP variation within a gene. A flowchart describing our overall approach for this question is given in Figure 2.1. Because HAPMAP does not contain information on all SNPs, we used the SeattleSNP database as our starting dataset because all detected SNP variants are reported. However, because the numbers of individuals genotyped by HAPMAP or sequenced by SeattleSNP are not large, it is possible that SNPs may be present in one database but not the other. [Aside: the probability of detecting SNPs with different minor allele frequencies of 0.05 and 0.01 in the SeattleSNP databases is 0.9 and 0.37 respectively, the same for HAPMAP database is 0.99 and 0.7 respectively]. However, SeattleSNP will, in general, contain more SNPs.

There were 251 genes (and a total of 19448 SNPs) in the SeattleSNPs database. We used the dbSNP assigned rs-numbers to scan the HAPMAP database for these 251 genes. Of the total 251 genes in SeattleSNP database, 186 genes had at least one SNP in common in both two databases. So we developed two databases, one with the common SNPs between the SeattleSNP and HAPMAP database (we call it adjusted HAPMAP database); the other with the SNPs which were present in SeattleSNP but not in HAPMAP (we call it adjusted SeattleSNP data base). Because some introns of some genes in SeattleSNPs were not sequenced, there could be a SNP in HapMap with MAF > 5% that was not in SeattleSNPs. These SNPs were also removed from our database because we wanted to see how a subset from HAPMAP database predicts a superset from the SeattleSNP database. So the predictor set of SNPs (from HAPMAP) should not have SNPs that are not present in the superset of SNPs (from SeattleSNP). We

deleted any SNP that has a MAF < 5% from the HAPMAP genotype file, but all the SNPs were kept in the adjusted SeattleSNP database, because we wanted to see how well the adjusted HAPMAP SNPs can predict rare SNPs (MAF < 5%) also.

Now that we had a "representative" set of SNPs in both databases, we used Hclust (Rinaldo et al. 2005) to choose tag SNPs from the SNPs in the adjusted HAPMAP database. The Pearson's correlation coefficient r_p^2 is used because the theoretical properties are well studied and it also measures how well a SNP can be used as a proxy for any other SNP (2005). That means for each gene we took the set of SNPs that are common in adjusted SeattleSNP and adjusted HAPMAP and ran Hclust on them using the genotype data from HAPMAP. The set of tag SNPs forms the predictor set for our CART model. Hclust is a hierarchical clustering algorithm that identifies tag SNP based upon genotype data. The algorithm computes a similarity matrix from the square of Pearson's correlation (r_p^2) between allele counts at pairs of loci; uses hierarchical clustering to group correlated SNPs; then chooses a SNP to represent each cluster. Each unmeasured SNP then has $r_p^2 \geq \eta$ for at least one tag SNP. η for selecting tag SNPs was taken as 0.5, that means any set of SNPs that have correlation greater than sqrt(0.5) will fall in the same cluster, and one of them will be selected as tag SNP. Eight of the genes had only one single SNP in common between the two databases and so these eight genes were removed because they would be uninformative. Also, after running Hclust we dropped 5 more genes because the SNPs from those genes were yielding singular matrices, so Hclust was not able to compute tag SNPs. So the final analysis was based on 173 genes. The next sets of analyses were all based on genotype data from our adjusted SeattleSNP set of genes. For each gene in the adjusted SeattleSNP data set we developed two files, one with tag SNP genotypes, selected using Hclust as described above, and one with genotypes of those SNPs that are present in adjusted SeattleSNP but not in adjusted HAPMAP. Next, we performed regression analyses, in which, for a specific gene every SNP that was not in adjusted HAPMAP, was used as the dependent variable, and all of the tag SNP from the same gene were used as independent variables. For each of the SNPs that were not present in adjusted HAPMAP, we estimated the β explained by tag SNPs.

The mean of all of the R^2 obtained from analyses of each of the SNPs that were not present in adjusted HAPMAP was assumed to be the representative R^2 for that gene.

To obtain an easily interpretable relationship between the mean R^2 explained by tag SNPs and various characteristics of genes (such as SNP density, gene size, etc.), we employed Classification and Regression Tree (CART) (Brieman et al. 1984) as implemented in the "rpart" function available in the statistical package R (R Development Core Team 2008). We used the anova option in rpart, which fits a regression tree in which the criteria for pruning is the mean square error. So at each specific node in the tree, there will be a binary split in the data based on minimizing the mean square error. Before performing the CART modeling, we assessed several variance-stabilizing transformations of mean R^2, the dependent variable, and determined no transformations were necessary (Figure 2.2). The independent variables in CART model were (1) the number of tag SNPs (ntag) which is the number of tag SNPs we got from the adjusted HAPMAP database, (2) the density of tag SNPs per 10 kbp (dtag), which is derived from the number of tags (ntag) and the gene length in kbp and (3) the density of all SNPs per 10kbp, dsnp, which is derived from the number of SNPs that are present in SeattleSNP (nsnp) divided by the length of the gene in 10 kbp. The coverage of a gene by tag SNPs depends on the density of SNPs in that gene, density of tag SNPS in that gene and also the total number of tag SNPs in that gene. If a gene has a very poor density of tag SNPs then the coverage won't be good.

Our second question was how well does a specific set of tag SNP represents the SNP variation within a specific set of genes. Using the CART model and a list of SNPs, we can estimate whether a specific candidate gene (or group of candidate genes) is well covered or poorly covered, even if all available tag SNPs in HAPMAP have been chosen. However, in order to apply the CART model, one needs information on the density of all SNPs within a gene, which is not yet available for all genes. Therefore, we used regression models to develop an estimate of SNP density (dsnp) using data from on a set of candidate genes for the MrOs project. In other words, we tried to predict dsnp based on dtag, ntag, number of SNPs in Hapmap (nhap) and interaction between the three variables. A flow-chart describing

this approach is presented in Figure 2.3. From our MrOS candidate gene study, we had genotype data on 28 genes that were also present in the Seattle SNP database. We divided this set of 28 genes into 4 groups according to the size of the genes, as we wanted to maximize the number of genes in testing group without sacrificing the size of the training group. We used a sampling scheme without replacement (SRSWOR), which means we took 4 genes as testing set without any overlap in any of the training sets. We took each group as a testing group and the rest as training groups and then fitted stepwise regression models for each set. So we had 4 genes as testing set and 24 genes for training set for each model. As described above, the model we estimated for a specific set of 4 genes was,

$$\text{dsnp} = \beta_0 + \beta_1\text{*dtag+ } \beta_2\text{*ntag+ } \beta_3\text{*nhap+ } \beta_4\text{*dtag:ntag+ } \beta_5\text{*dtag:nhap+ } \beta_6\text{*ntag:nhap,}$$

where dtag:nhap means the interaction between dtag and nhap. So we used a 7 fold cross validation technique for building our model. A 5 fold to 10 fold cross validation is generally sufficient for model building (Hastie et al. 2001). This above model was fitted using each training set, and the average squared error was calculated for each test set. The model that had the minimum squared error was chosen as the best model. We then fitted the chosen model on all the 28 genes and obtained the final coefficients for the independent variables for predicting dsnp. Using this final model, we estimated dsnp for 257 candidate genes in our MrOS study. Finally, we applied our CART model to these data to obtain an estimate of mean SNP coverage for our specific set of candidate genes.

2.4 RESULTS

One of the aims of the HAPMAP project is to generate a sufficient number of SNPs within each gene to facilitate local and genome wide association studies. This goal is achieved if HAPMAP includes a sufficient panel of SNPs to successfully predict the genotypes of many of the unmeasured SNPs in the

genome. To assess how successfully the HAPMAP meets this goal, we developed a metric that requires knowledge of the full set of SNPs in a variety of regions in the genome. The SeattleSNP database (Carlson et al. 2003), comprising sequence data from numerous genes, currently provides this knowledge for 315 genes. When we started our analysis (Sept, 2006), data on 251 genes was contained in SeattleSNP database.

From the SeattleSNPs database, 173 genes had data that could be compared with HapMap data (see Methods). The number of SNPs per gene in HAPMAP varied from 2-94, with an average of 17 SNPs per gene and in SeattleSNP the number varied from 18-312 (Table 2.2). We screened the SNP genotype data in HAPMAP database for these 173 genes to select a subset that conveys non-redundant information, i.e., tag SNPs, by using a tag SNP selection procedure Hclust (Rinaldo et al. 2005). The number of tag SNPs per gene varied from 1 to 33, with an average of 6 tag SNPs per gene.

As expected the number of SNPs per gene in HAPMAP is relatively small compared to the SeattleSNP database and a substantial fraction of HAPMAP SNPs are not selected as tag SNPs. This fraction depends on the minimum level of correlation η pre-specified in the tag SNP selection procedure. For correlation level of, η = 0.5, 65.4% of the SNPs are dropped because they are highly correlated with tag SNPs. As expected this fraction increases with the number of SNPs originally included in HAPMAP (Figure 2.4). Thus, for an association analysis targeting a specified region, it will usually be unnecessary to genotype HAPMAP's full set of SNPs in that region. As expected many of the SNPs are nearly redundant with the other SNPs in the database.

The ability to predict unmeasured SNP genotypes using the full set of tag SNP genotypes can be assessed by the coefficient of determination (R^2) of the linear model predicting the allele count of an unmeasured SNP based on the tag SNP allele counts. This R^2 is a function of the number of tag SNPs in each gene (Figure 2.5). Genes with 1-2 tag SNP have a lower mean R^2 when compared with genes that have 3 or more tag SNPs. Nevertheless, the coverage is quite good on average for many genes, especially those with greater than 5 tag SNPs. For genes with only one or two tag SNPs, the mean R^2 is 55.15%. For

genes with >4 tag SNPs the mean R^2 was 56.35% and for genes with 3-4 tag SNPs the mean R^2 was 55.2%.

Although informative, the results presented above do not inform the investigator when he has sufficient SNPs to adequately cover the total SNP-variation within a gene versus when he would need to supplement HapMap SNPs with additional SNPs. [These additional SNPs could be obtained from other databases or from re-sequencing areas of the candidate genes of interest.] Therefore, we built a classification and regression tree-based model to predict average R^2 based upon three covariates: the number of tag SNPs (nTag), the number of SNPs per 10 kbp (dSNP), and the number of tag SNPs per 10 kbp (dTag). The model delineates covariates that yield partitions with a similar response (Figure 2.6). For example, if a gene has density of tag SNPs between 5.86 and 6.71 with the total number of SNPs less than 117.6, then R^2 is likely to be excellent (about 0.73). But if there are less than 4 tag SNPs in a relatively large gene (the density of tag SNP (dtag < 1.53), then R^2 is likely to be about 0.46. As expected, for genes with low linkage disequilibrium between SNPs, and high total SNP density, prediction was low (for genes with dtag ≥ 4.3 and dsnp ≥ 117.6 the prediction is 49%). After dividing the total prediction space into 11 cells, the model predicts 55.8% of the variability in the average R^2 (Figure 2.6).

Finally, we wished to apply the CART model to a set of data from our own candidate gene study (MrOS). However, because all of our candidate genes are not present in the SeattleSNP database, or any other database, we did not have an estimate of the total SNP density (dSNP) for these genes and thus we could not use the CART model developed above. As described in the methods, we used a 7-fold cross-validation regression model approach to predict dSNP using data on 28 genes that were in common between our data and SeattleSNPs. For all seven models, four genes were used as the training set, and the remaining 24 genes were used to test the regression model. Results from all 7 training models from the cross-validation approach and the corresponding mean square errors from the corresponding testing set are given in Table 2.2. The training model with the lowest mean square error is highlighted in red, after

obtaining the regression model with the smallest square error, we then used the data on all the 28 genes to estimate the regression coefficients of the final model. The final prediction model is given as:

dSNP = 65.23-0.34*nhap-13.73*ntag+9.98*dtag+0.24*nhap:ntag+0.29*nhap:dtag,

where nhap:ntag is the interaction between nhap and ntag, nhap:dtag is the interaction between nhap and dtag. Table 2.3 contains the regression coefficients (and standard errors) for the final model.

We applied this final prediction model to estimate dsnp for all 257 of our candidate genes. We then used our CART model to obtain an estimate of the average R^2 for the genes in our study. Our estimate of R^2 for the MrOS study is 52.99%, which is less than the 55.8% obtained by Hapmap overall. The range of average R^2 varies from 43% to 73%. Based on our CART model results for MrOS we calculated how many SNPs are genotyped for each gene in MrOS and how many SNPs are selected as tag SNPs for each gene from HAPMAP data. For an example for gene A, suppose we have m_A SNPs in MrOS and for the same gene A we have h_A tag SNPs from HAPMAP. Then the number we need is h_A-m_A for gene A to get the comparable coverage as in HAPMAP. This quantity is calculated for all the genes in MrOS dataset and the overall mean is obtained. We estimated that we would need to assay an additional 2.84 SNPs per candidate gene on average to get equivalent coverage as is currently available in HAPMAP.

2.5 DISCUSSION

HapMap contains numerous SNPs in high LD with other SNPs. For a region of interest, about half the SNPs in HAPMAP are dropped after tag SNP selection (Table 2.2). How well do tag SNPs cover the region of interest? For studies involving individuals of European ancestry, average prediction of

unmeasured SNPs is quite good when the density of tag SNPs in the region is high (Figures 2.5 and 2.6), and surprisingly good even when few SNPs are available from HAPMAP for the region. Although several groups have reported on how well HAPMAP SNPs cover unmeasured genotypes (Tantoso et al. 2006), there are differences with our approach. First, we chose tag SNPs with a $r^2 = 0.5$ versus $r^2 = 0.8$. This choice reduced the number of tag SNPs in our analyses, and thus we were trying to predict untyped SNPs using a minimal r^2 among our tag SNPs. In addition, we also used tag SNPs to try to predict the rare untyped SNPs, that is, MAF < 5%, whereas other investigators have tried to predict all SNPs with MAF > 5%. These two reasons likely influence our obtaining an overall average $R^2 = 55.8\%$. Given that we chose lower r^2 value and tried to predict the rare SNPs, we conclude that the coverage is pretty good.

One of the most important utilizations of our model is in predicting the performance of tag SNPs chosen for our candidate gene study. We can apply the CART model to our list of genes to see how good the coverage is for our genes. Based on specific characteristics of our genes (gene length, number of tag SNPs from HAPMAP, number of tag SNPs present in our study and the total number of SNPs for that particular gene) we can predict how well covered our genes are, as compared to SNPs within the current HAPMAP database. If investigator wants, he/she can apply the CART model to the set of genes, and then determine whether additional SNPs are required to obtain satisfactory coverage. As stated in the results section our MrOS study needs another ~3 SNPs on average per gene to get similar coverage as is currently available in HAPMAP.

We need to examine the coverage of HAPMAP for other populations as well. From previous results, however, it is reasonable to expect that HAPMAPs utility would be roughly the same for populations of Asian or European descent because the magnitude of LD tends to be similar in these populations/samples (Gabriel et al. 2002). Populations of African descent, on the other hand, tend to have substantially smaller LD than other populations, so coverage in HAPMAP will be more challenging (Gabriel et al. 2002). We tried to build a CART model for the African sample from HAPMAP. But the SeattleSNP database has data on the African-American population. The LD structures of these two

populations are not same (Sawyer et al. 2005). So tag SNPs are unlikely to be readily transferable between these two populations. But developing a CART model for the African population could be very useful, because in Africans there is not as much long range LD. In other words the LD blocks in Africans are small, and thus the performance of tag SNPs may not be very good in this population.

2.6 TABLES AND FIGURES

Table 2.1:The number SNPs/gene 173 genes in common in the HAPMAP and SeattleSNP databases

SNPs/GENE	HAPMAP SNPs	Seattle SNP	TAG SNPs
Average	17	112	6
Max	94	312	33
Min	2	18	1
Total	2886	19448	999

Table 2.2:Description of the seven training models used in the cross-validation approach to predict dSNP and the mean square error obtained from the seven testing sets

Genes in Testing Set	Variables in the Model	Mean Square Error
MMP9,VEGF,TRPV5,PPARA	nhap,ntag,dtag,nhap:ntag,nhap:dtag	185.33
IGF2,IL1RN,IL1R2,CHUK	nhap,ntag,dtag,nhap:ntag	831.18
IL6,TRAF6,IL1R1,TRAF2	nhap,ntag,dtag,dtag:ntag	3309.23
NFKBIB,TNFRSF1A,DCN,TNFRSF1B	nhap,ntag,dtag,nhap:ntag,nhap:dtag	86.31
CSF2,IL1A,RIPK1,IGF1	nhap,ntag,dtag,nhap:dtag	658.99
ALOX15,CSF3R,IFNAR2,IKBKB	nhap,ntag,dtag,nhap:ntag,nhap:dtag	1769.48
IL1B,TNFAIP3,IFNAR1,PPARG	nhap,ntag,dtag,nhap:ntag	1210.20

Table 2.3: Description of the Final Model used to Estimate dSNP in the MrOS Data Set

Variables in the Model	Regression Coefficient	Std. Error
Intercept	65.23	10.41
nhap	-0.34	0.50
ntag	-13.73	3.00
dtag	9.98	2.01
nhap:ntag	0.24	0.10
nhap:dtag	0.29	0.20

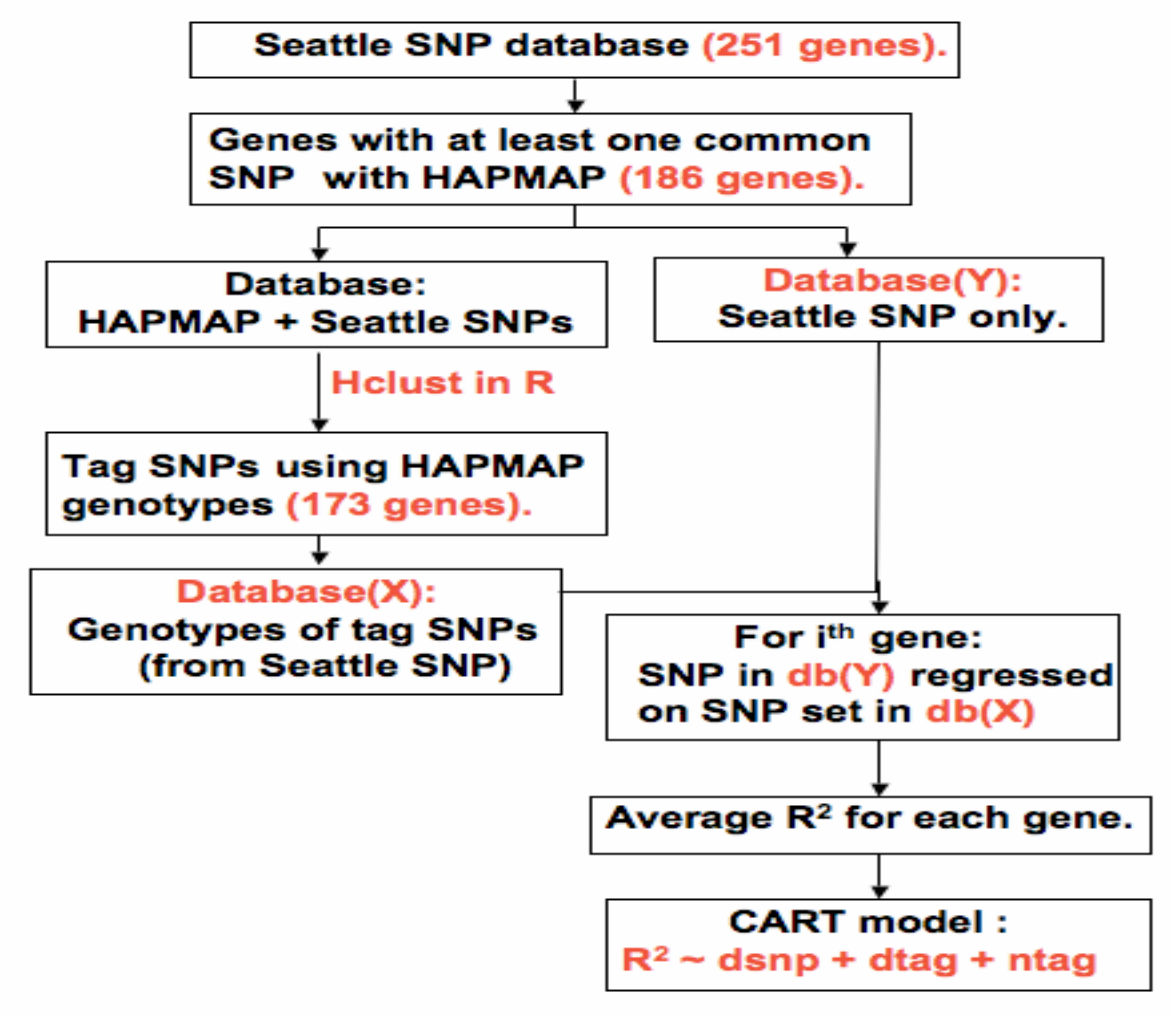


Figure 2.1: Flow Chart of the approach used to develop the CART model

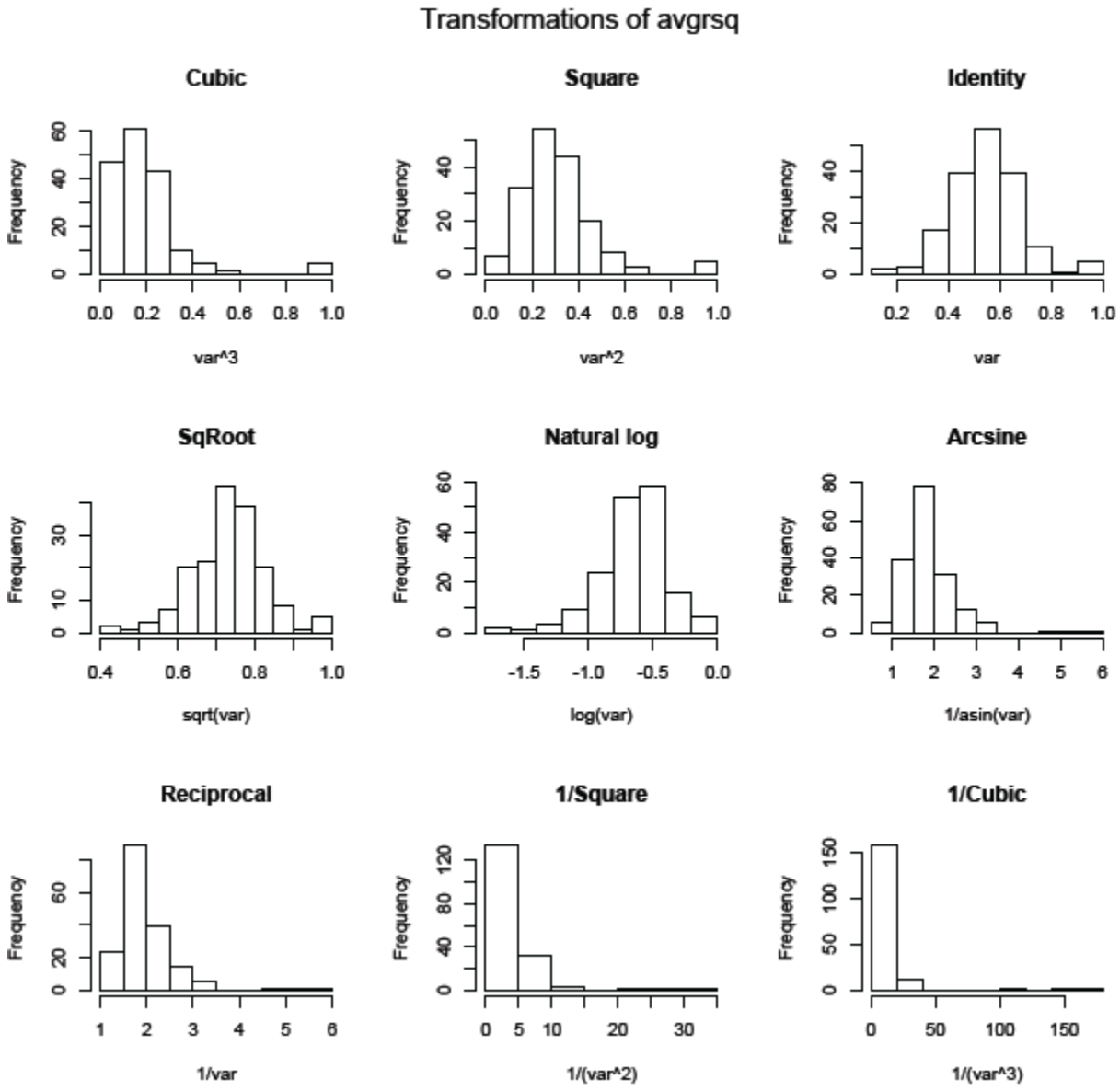


Figure 2.2: Distribution of Transformations of average R^2

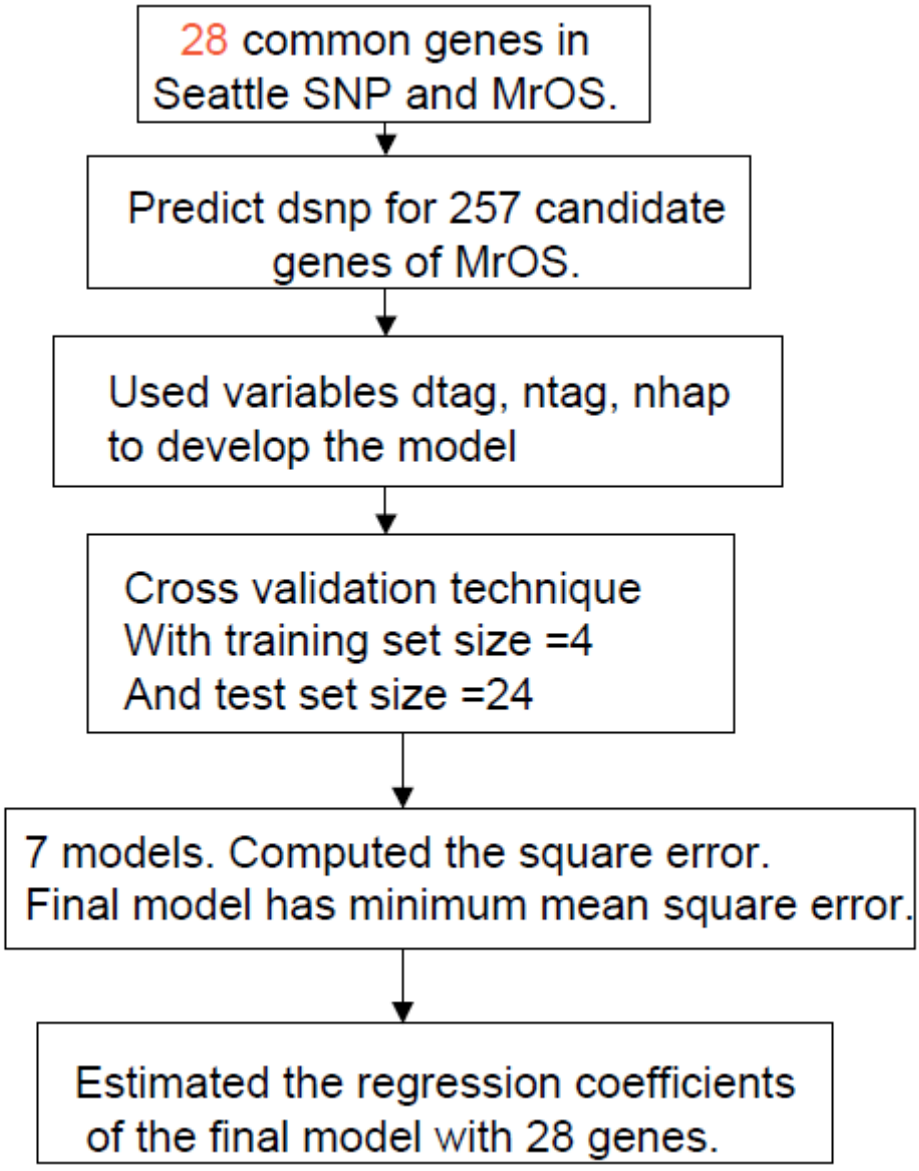


Figure 2.3: Flow Chart of the cross validation approach to obtain estimates of density of total number of SNPs/ 10kbp (dsnp)

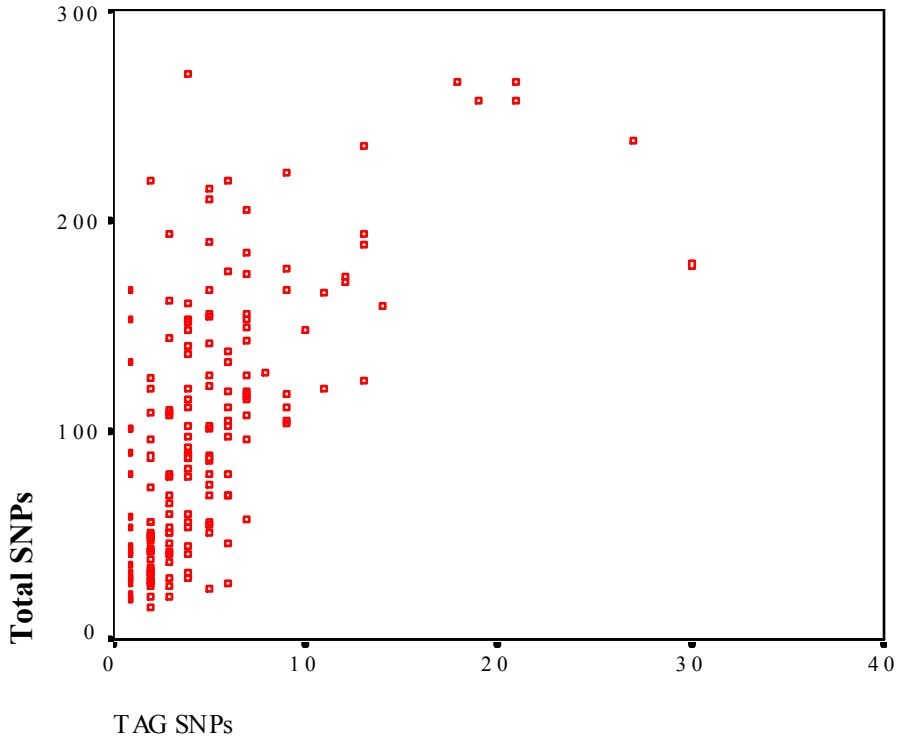


Figure 2.4: A plot of tag SNPs per gene selected from the HapMap database for η = 0.5 by the number of total SNPs present in HAPMAP database

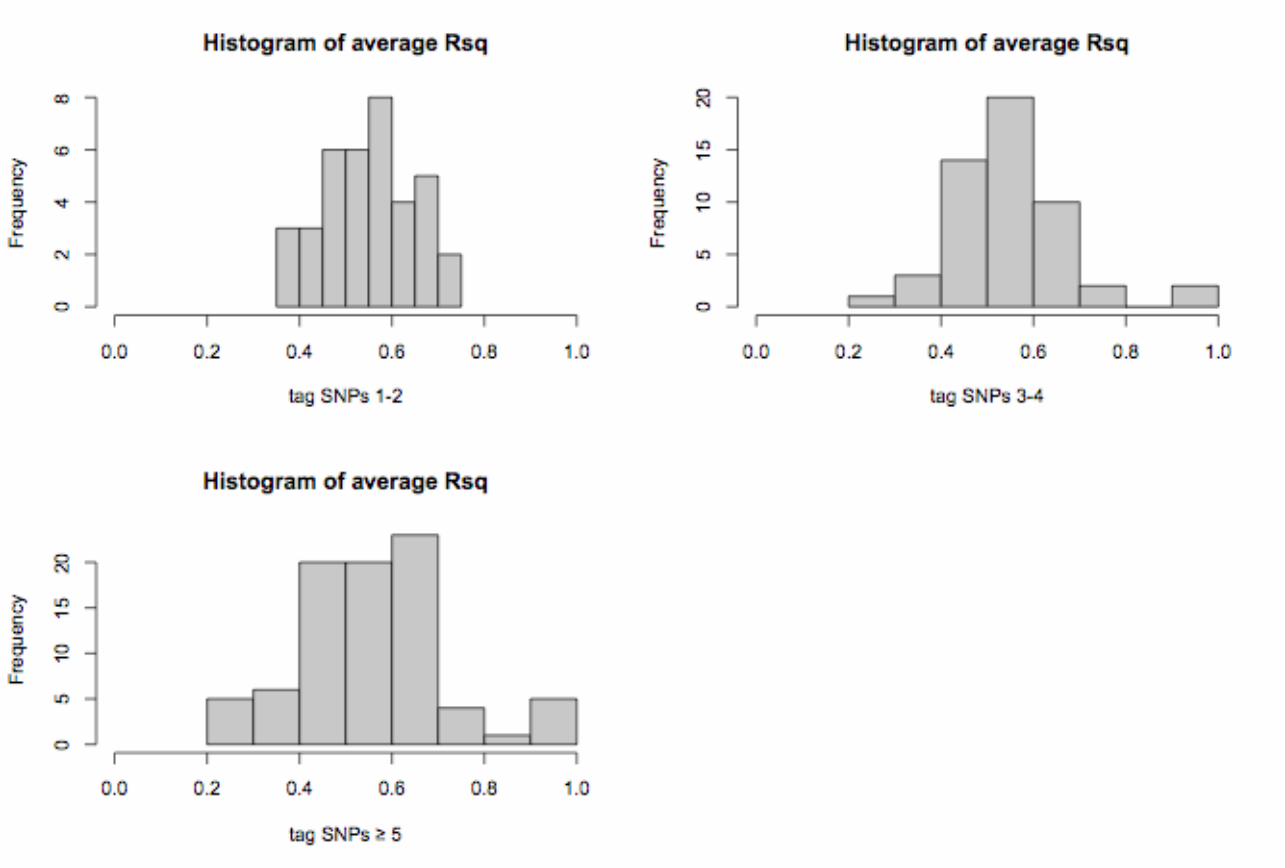


Figure 2.5: Histogram of Average R^2 for predicting unmeasured SNPs within genes by the number of tag SNP per gene: 1-2 tag SNPs, 3-4 tag SNPs, and ≥ 5 tag SNPs

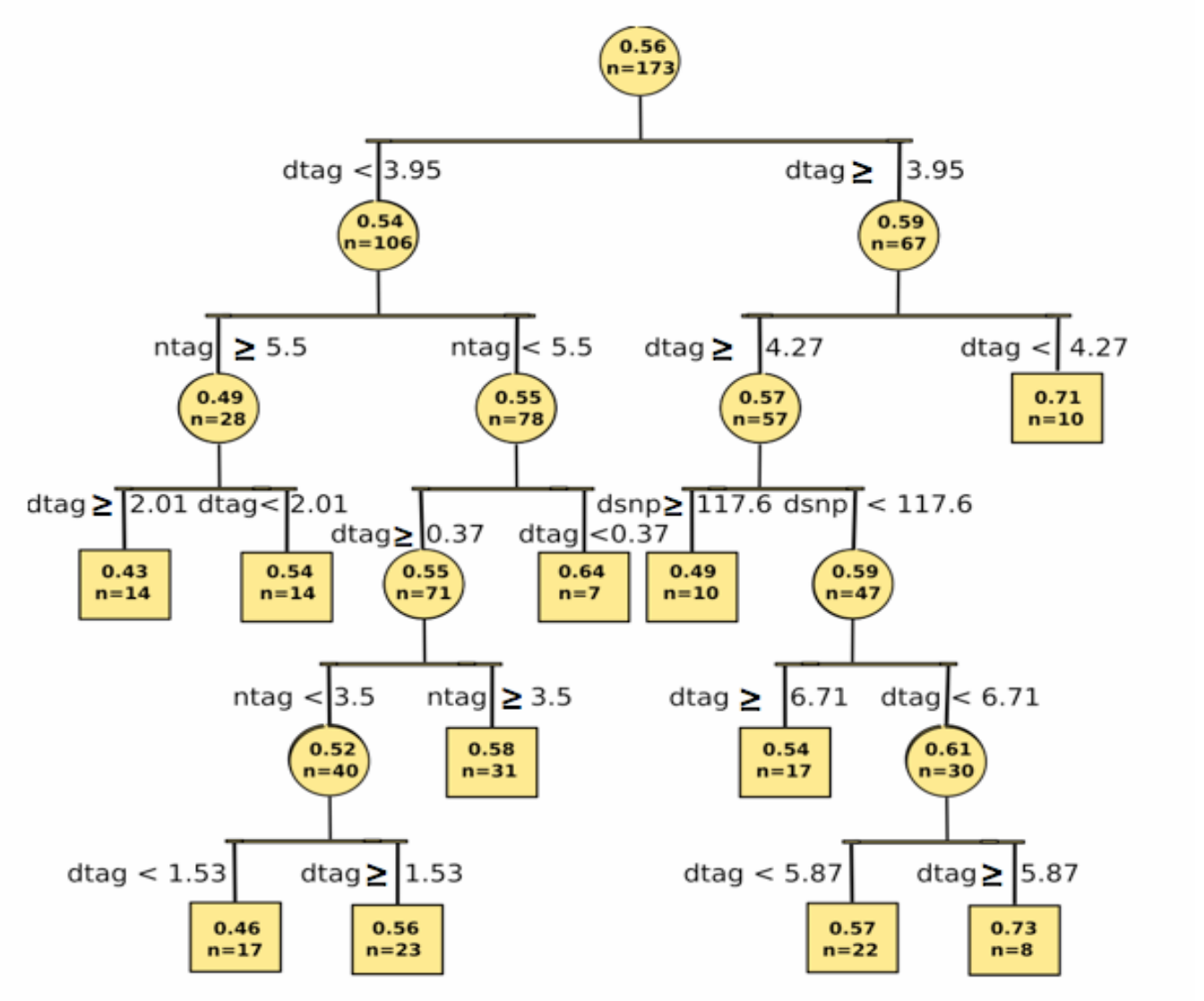


Figure 2.6: CART model for predicting coverage of HAPMAP tag SNPs

3.0 ASSESSING ANCESTRY AND POPULATION SUBSTRUCTURE AMONG THE HEALTH ABC COHORT

3.1 INTRODUCTION

One of the major concerns with association studies is that false positive results may occur due to the presence of unrecognized population substructure (Devlin et al. 1999). Population stratification may occur in admixed populations like Mexican Americans (Salari et al. 2005) as well as in Europeans (Campbell et al. 2005) due to variation in genetic ancestry among individuals both within and between racial/ethnic groups. For example, there is a cline in lactose dehydrogenase across European population (Burger et al. 2007). In 1999, Devlin and colleagues developed the genomic control method to address the issue of population stratification (Devlin et al. 1999). This method adjusts for population substructure at population level by estimating an inflation parameter based on results from a set of null markers. Another method to estimate population substructure uses data on ancestry informative markers (AIM) (Smith et al. 2001). AIMs are a set of markers that have large differences in allele frequencies between different populations. Using genotype data from the AIMS, individuals can be grouped into clusters or subpopulations. One common method used to assess population substructure is the program STRUCTURE (Pritchard et al. 2000) which uses genotype data to assign individuals to discrete population clusters. However, assignment of individuals to different clusters is sensitive to the number of clusters as well as the metric used in defining the clusters (Price et al. 2006). We chose to address the problem of possible substructure using a principal components method (Patterson et al. 2006), which uses

the correlation structure between individuals at several uncorrelated loci to assess the structure present between or within populations.

We investigated population substructure for two reasons. First, because we know that BMD differs among individuals with different ancestry (see Introduction, section 1.3), we decided to classify each individual in the HealthABC cohort into the most probable ancestry group (European American or African-American) based on their genetic information, rather than self-reported ethnicity. Next we wanted to control for possible substructure within each genetic ancestry group, so we re-analyzed the genotypic data within each group to assess how many eigenvectors should be included as covariates in our subsequent candidate gene analyses.

3.2 DATA

3.2.1 Identification of Candidate Genes and SNPs

To identify a set of candidate genes to be genotyped, public databases, especially PubMed, were searched to identify physiologically relevant candidate genes for musculoskeletal traits. After identifying a set of candidate genes, SNPs were identified using HAPMAP phase I data. All SNPs that were within 30kb upstream or 10kb downstream of the candidate gene, and had a minor allele frequency of $\geq$5% were identified (www.hapmap.org) (2003). From this list, tag SNPs were chosen using at a pairwise correlation level of r^2= 0.8 (Roeder et al, 2005). Candidate genes that were clustered near each other were considered to be a single unit, and tag SNPs were chosen for the entire cluster. For an example IGFBP2 and IGFBP3 both are located on Chromosome 2 within 7.6 kb of each other, so they were treated as single entity, called the IGFBP cluster. In addition to tag SNPs, potentially functional non-synonymous SNPs were also included, based on information in the databases PupaSNP (pupasite.bioinfo.cipf.es) (Zhaoet al, 2004) and

Promolign (polly.wustl.edu/promolign/main.html) (Conde et al 2004). All selected SNPs were then genotyped using the Illumina Golden Gate custom assay.

3.2.2 Data cleaning

After receiving the genotype data from the laboratory, the following quality control procedures were implemented to ensure the accuracy of our genotype data. We started with 3001 individuals and 1536 SNPs in 138 genes from the Illumina Golden Gate assay (Table 3.1). Of these 1536 SNPs 71 had no data, so we did our data cleaning on 1465 SNPs. We also had information on 126 SNPs, that were genotyped in Dr. Ferrell's lab, for 2815 individuals. We separately cleaned the two data sets because the sample sizes differed. For the Illumina data set, individuals missing > 7% of the genotype data were omitted from subsequent analysis due to poor DNA samples (22 individuals were removed). Next, all SNPs with a call rate of < 90% were omitted (19 SNPs were dropped) due to poor genotyping fidelity. Monomorphic (i.e. non-informative) SNPs were omitted. Next, Hardy-Weinberg equilibrium (HWE) was assessed for each SNP using the exact test separately for each self-reported race, and SNPs that are not in HWE (i.e. p-value < 0.001) were flagged, but not removed at this stage. Each of the association analysis programs that we used (see Chapter 5 and 6) also perform internal checks for the HWE within each sex by ancestry group. For all of these programs, SNPs with a p-value < 0.001 for HWE were discarded from subsequent analysis. Of the 1465 SNPs assayed using the Illumina Golden Gate method, 7 SNPs from 3 genes were located on the X chromosome. We detected male heterozygotes for these SNPs, implying genotyping errors, so they were deleted from any subsequent analysis. As I do not have case control data so I did not apply quality control procedures that are specific to Case Control studies.

For the genotype data from Dr. Ferrell's lab, all the individuals and SNPs passed the quality control filters described above. We combined the two data sets (2979 individuals, 1439 SNPs from Illumina, and 2815 individuals and 126 SNPs from the Dr. Ferrell's lab). For these SNPs we plotted the frequency distribution of minor allele frequency in the two populations (Figure 3.1). As expected, the

MAF distribution differs between the two self-reported ancestry groups, and African Americans have more SNPs with MAF < 0.1.

After performing the quality control procedures above, we deleted all SNPs that had minor allele frequency < 5% within each self-reported ancestry group (78 SNPs were deleted). Therefore, after data-cleaning, we had genotypic data on 2979 individuals (1653 self-reported European Americans, 1180 self-reported African- Americans, and 146 unknown). Genotypes were available on total 1487 total SNPs in 151 total genes.

Originally, the SNPs that were genotyped in Dr. Ferrell's lab were not assigned a RefSNP number (rs number) because most of these SNPs were chosen from the literature. However, each SNP that is submitted in dbSNP is assigned a reference SNP id by dbSNP (Smigielski et al. 2000) and we wished to obtain rs numbers for all of our SNPs for consistency and accuracy. So we had to determine the rs number for each SNP using the information on the forward and reverse primers for all the SNPs. We used UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgBlat) and submitted the forward and reverse primers to obtain the DNA sequence, and the SNP, between those primers. For some of the primer pairs, we obtained multiple SNPs and rs numbers and we consulted with Dr. Ferrell to obtain the correct rs number.

3.3 METHODS

3.3.1 Eigen analysis methods

In general, eigen analysis is a method used to reduce the information contained in a regular matrix into a diagonal matrix. We used principal component analysis (PCA), a form of eigen analysis, to reduce the individual covariance matrix (in which each element contains the

estimated covariances between two individuals) and obtain a set of principal components (eigen vectors) that can adequately explain the genetic variation present in the set of SNPs. One of the requirements of eigen analysis of genotypes is that the genotypes must be uncorrelated, otherwise, individuals will form clusters because they had common genotypes (see exploratory analysis in Results section below). Therefore, we first selected a set of uncorrelated SNPs from the total number of available SNPs. We used HCLUST (Rinaldo et al. 2005) which allows us to select tagging SNPs with a specific correlation relationship, and we selected $r^2 \leq 0.04$.

To assess ancestry using genotypic data on a set of SNPs for each individual in our HABC cohort, we first calculated the covariance structure among the individuals. We recoded the genotype data for each SNP as 0,1,2 according to the number of minor alleles present in the genotype of a specific individual. First we calculated the mean and standard deviation for each SNP. The mean was calculated as sum of SNP counts/number of counts. For example the number of counts is m (if there is no missing data for a particular SNP), and if m_0, m_1, m_2 are the number of 0,1 and 2's respectively in the data for that SNP, then the mean is calculated as: mean = $((m_1*1+m_2*2)/m)$. The standard deviation is estimated as sqrt(2*p*(1-p)) where p is the allele frequency of the minor allele for each SNP, and p is calculated as mean/2.

Then we determined the standardized allele counts by taking [(actual count)-mean]/sd for each SNP. Finally we replaced missing SNP counts with 0 (the expected value, because for a standardized variable the expected value is always zero). We obtained the covariance matrix by calculating the sum of cross- products for the standardized SNP count divided by the number of SNP.

To calculate $(i,j)^{th}$ element of the covariance matrix, that is the covariance between $individual_i$ and $individual_j$, we can simply calculate the sum of the element-wise product of the i^{th} and j^{th} row of the adjusted SNP count matrix

$$C_{i,j} = \sum x_i * x'_j$$

where x_i is the i^{th} row and x_j is the j^{th} row of the standardized SNP count matrix.

After obtaining the covariance matrix, we estimated the eigenvalues and eigenvectors (principal components, PC) of the covariance matrix. The eigenvalues were modified using formulas given in Patterson et al (2006) because the modified eigen values follow a Tracy Wisdom Distribution. For the first 3 PCs, we also calculated a vector as

$eigenvector_i * sqrt(eigenvalue_i)$ where i= 1,2,3.

These modified vectors rescale the variability by multiplying each vector by the square root of the corresponding eigen value, so the vectors will be comparable.

Finally, to discriminate between different ancestry groups or to assess whether there is evidence of subpopulations within an ancestry groups, we made pairwise plots of the first three PCs, both overall and within groups. These plots were used to discriminate between European Americans and African Americans, and also to assess whether there was obvious substructure within the Caucasian American and African-American groups and how many PCs should be included as covariates in the candidate gene analyses.

3.4 RESULTS

3.4.1 Exploring ancestry estimation using the eigen analysis method

To obtain an understanding of what eigen analysis does, and the effect of violating some assumptions, I performed a series of exploratory analyses. First, to determine the effects of using correlated SNPs on ancestry determination, I performed eigen analyses on 119 correlated SNPs and plotted the results (Figure 3.2). As can be seen, we obtain evidence for two ancestry groups (based on PC2), but also artifactual

evidence for three sub-groups within each ancestry groups (based on PC3). Our collaborators related that such substructure is also seen in other studies when one uses correlated SNPs, and this substructure represents genotypes.

3.4.2 Assigning Individuals to Ancestry Groups

To classify individuals by African American versus European American ancestry, we obtained a set of 429 uncorrelated ($r^2 \leq 0.04$) SNPs using data on all individuals and performed eigen analyses. Figure 3.3 presents the plots of (modified) PC1 by PC2 for all individuals. These two eigenvectors contain most of the information regarding ancestry groups. In Figure 3, blue dots represent self-reported African-Americans, green dots represent self-reported Caucasian Americans, and red dots represent unknown ethnicity.

We next calculated the centroids (medians) of the first two eigen vectors separately for the two clusters and calculated the distance of those centroids (medians) from each of the points. Below is the formula for calculating the distances.

$$d_{ij} = eigenvalue1 * (eigenvector1_{ij} - centroid1)^2 + eigenvalue2 * (eigenvector2_{ij} - centroid2)^2,$$

where d_{ij} = the distance of i^{th} individual and the j^{th} population. i=1,2,….2979; j=1,2. So for each individual two distances are calculated: one from the Caucasian American cluster, the other from the African-American cluster. All of the unknown individuals were assigned to either the African American or European American ancestry group by comparing the two distances and assigning them to the group with the minimum distance. Based on these results, and in consultation with Drs. Roeder and Devlin, we decided to classify the HABC individuals into two ancestry groups using transformed eigen vector 1 = 0.04 as a cutoff value. Any individual with a higher value was assigned to the Caucasian American group

and any individual with a lower value was assigned in African-American group. Eight individuals previously classified as African-Americans were classified as Caucasian Americans and five of the Caucasian Americans were classified as African-Americans. All the 146 unknowns were classified into one of the two groups (86 Caucasian American and 60 African-Americans). So all of the HABC individuals were assigned to one of the two ancestry groups: 1742 Caucasian Americans, 1237 African-Americans, for a total of 2979 individuals (see Table 3.2).

3.4.3 Assessing Eigenvectors within Groups

After classifying all HABC individuals into two ancestry groups, we then assessed whether there was any substructure within each group. We performed a similar procedure to that described above. First, we used HCLUST to select uncorrelated (r^2=0.04) SNPs for each of the two groups separately: for the Caucasian American group 399 tag SNPs were selected and for African-Americans 499 tag SNPs were selected. We re-selected a set of uncorrelated SNPs within each ancestry group, rather than using the original set of uncorrelated SNPs, because (1) the number of people changed in the two groups, and (2) the LD pattern will change in the new groups. This new set of tag SNPs might be able to detect more subtle structures within the two populations. The plots of the first three PCs (pairwise) are given for the African-American ancestry group (Figure 3.4) and the Caucasian American ancestry group (Figure 3.5).

After visual inspection of the PC plots, we did not observe any obvious substructure present within the Caucasian American ancestry group. However in the African-American group, we see some substructure along PC2 axis. We wondered whether this substructure might be related to recruitment site (Pittsburgh versus Memphis), however, individuals from both sites were approximately equally represented in both clusters, and there was no difference in mean PC scores between individuals from Memphis and Pittsburgh (results not shown). We did the same analyses by sex, and also found no differences. Thus, this substructure is not due to recruitment site or sex. To mitigate possible effects of subtle substructure, we decided to incorporate the first three eigenvectors as covariates in our subsequent

candidate genes analyses for both the ancestry groups. Based on our analyses of environmental covariates influencing the bone traits (see Chapter 4), we will also remove variation for other significant environmental covariates. Thus, all of our candidate gene association analyses were done using residual trait phenotypes.

3.5 DISCUSSION

Population substructure is one of the major concerns in the performance of genetic association studies in populations of unrelated individuals. Investigators have proposed a variety of methods to mitigate the effects of population substructure, including genomic control, AIMs, and eigen analysis. Given our recent advances in high-throughput genotyping capabilities, most investigators propose to control for or assess population substructure using eigen analysis methods on data on large sets of SNPs. Because we had genotypic data on multiple SNPs in 138 candidate genes, rather than data on 100,000 or more AIM SNPs as has been used in several studies (Hinds et al. 2005; Shriver et al. 2005), we also wanted to determine whether we had sufficient number of uncorrelated SNPs to detect population substructure. The number of uncorrelated SNPs necessary to detect population substructure is not well-known with some investigators recommending a few hundred and others recommending more (Collins-Schramm et al. 2004; Hinds et al. 2005).

We first assessed how well our set of ~429 uncorrelated SNPs were able to classify individuals of African American versus European American ancestry. As can be seen in Figure 4, most individuals were easily classified into the two ancestry groups. Thus, as might be expected, data on 429 uncorrelated SNPs is sufficient to distinguish between major continental ancestral groups, such as Europeans versus Africans. Then we assessed substructure within Europeans (399 tag SNPs) and African-Americans (499 tag SNPs), and observed no apparent population substructure present within these populations. Although

we see some possible substructure in African American ancestry group, a larger number of uncorrelated SNPs may be necessary to detect more subtle structure.

Although ancestry estimation using AIMs can be much more powerful, it depends on the availability of allele and genotype frequency data on a specific set of markers for presumed ancestry populations (Barnholtz-Sloan et al. 2008). We did not have the opportunity to work with the population-specific AIMs, so we used the principal component method to analyze the possible substructure present in our population. This method corrects for population structure using continuous axes of variation, which gives useful information regarding within-continent variation (Price et al. 2006). By using this method we tried to have continuous axes of variation rather than assigning each individual into clusters, as is done by the program STRUCTURE (Patterson et al. 2006).

3.6 TABLES AND FIGURES

Table 3.1: Details of genes studied in HABC cohort

gene	Full Name	chr	# of SNPs
ACVR2B	activin A receptor, type IIB	3	4
ADIPOR1	adiponectinreceptor1 (MUSCLE)	1	7
AR	Androgen receptor	X	4
ARF3	ADP-ribosylation factor 3	12	1
BDNF	Brain derived neurotrophic factor	11	10
BMPR2	bone morphogenetic protein receptor, type II	2	8
CCDC65	coiled-coil domain containing 65	12	1
CNTF	Ciliary neurotrophic factor	11	3
CNTFR	CNTF receptor	9	12
COMT	Catechol-O-methyl transferase	22	16
CRH	CORTICOTROPIN-RELEASING HORMONE	8	4
CRHBP	CORTICOTROPIN RELEASING HORMONE-BINDING PROTEIN	5	9
CRHR1	CORTICOTROPIN-RELEASING HORMONE RECEPTOR 1	17	8
CRHR2	CORTICOTROPIN-RELEASING HORMONE RECEPTOR 2	7	14
CSF1R	COLONY-STIMULATING FACTOR 1 RECEPTOR	5	17
CSF2	COLONY-STIMULATING FACTOR 2	5	12
CSF2RB	colony stimulating factor 2	22	12
CSF3R	colony stimulating factor 3 receptor	1	15
CSK	c-src tyrosine kinase	15	1
CYP11A1	Cholesterol side-chain cleavage enzyme	15	7
CYP11B1	cytochrome P450, family 11, subfamily B, polypeptide	8	12
CYP17A1	CYTOCHROME P450, FAMILY 17, SUBFAMILY A, POLYPEPTIDE 1	10	10
CYP19A1	CYTOCHROME P450, FAMILY 19, SUBFAMILY A, POLYPEPTIDE 1	15	19
CYP1A1	cytochrome P450, family 1, subfamily A, polypeptide 1	15	3
CYP1A2	cytochrome P450, family 1, subfamily A, polypeptide 2	15	2
CYP1B1	cytochrome P450, family 2, subfamily B, polypeptide 1	2	12
CYP21A2	Steroid 21-hydroxalase	6	2
CYP24A1	cytochrome P450, family 24, subfamily A, polypeptide 1	20	15
CYP27B1	cytochrome P450, family 27, subfamily B, polypeptide 1	12	5
CYP3A4	cytochrome P450, family 27, subfamily B, polypeptide 1	7	4
ESR1	Estrogen receptor α	6	34
ESR2	Estrogen receptor β	14	11
FBXO32	F-box protein 32	8	18
FRZB	frizzled-related protein	2	12
FST	FOLLISTATIN	5	9
GC	group-specific component (vitamin D binding protein)	4	10
G-CSF	colony stimulating factor 3 (granulocyte)	17	8
GDF8	Myostatin ;growth differentiation factor 8	2	5
GH1	Growth Hormone	17	5
GHR	Growth Hormone Receptor	5	17
GHRH	Growth Hormone Releasing Hormone	20	4
GHRHR	GHRH receptor	7	10
GHRL	ghrelin precursor	3	2
GNRH1	GONADOTROPIN-RELEASING HORMONE 1	8	5
GNRH2	GONADOTROPIN-RELEASING HORMONE 2	20	14
GNRHR	GONADOTROPIN-RELEASING HORMONE RECEPTOR	4	9

Table 3.1 continued

gene	Full Name	chr	# of SNPs
HOXA1	HOMEOBOX A1	7	3
HOXA10	HOMEOBOX A10	7	1
HOXA11	HOMEOBOX A11	7	4
HOXA13	HOMEOBOX A13	7	7
HOXA2	HOMEOBOX A2	7	3
HOXA3	HOMEOBOX A3	7	1
HOXA4	HOMEOBOX A4	7	1
HOXA5	HOMEOBOX A5	7	1
HOXA6	HOMEOBOX A6	7	2
HOXA7	HOMEOBOX A7	7	1
HOXA9	HOMEOBOX A9	7	3
HSD11B1	11-@BETA-HYDROXYSTEROID DEHYDROGENASE, TYPE I	1	11
HSD17B1	17-@BETA-HYDROXYSTEROID DEHYDROGENASE I	17	5
HSD17B2	17-@BETA-HYDROXYSTEROID DEHYDROGENASE 2	16	11
HSD17B3	17-@BETA-HYDROXYSTEROID DEHYDROGENASE 3	9	15
HSD17B4	17-@BETA-HYDROXYSTEROID DEHYDROGENASE 4	5	13
HSD3B1	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase	1	3
HSD3B2	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase	1	5
IGF1	Insulin-like growth factor-1	12	14
IGF1R	IGF receptor 1	15	39
IGF2	Insulin-like growth factor-2	11	5
IGF2R	IGF receptor 2	6	37
IGFALS	insulin-like growth factor binding protein, acid labile subunit	16	3
IGFBP1	insulin-like growth factor binding protein 1	7	5
IGFBP2	IGF binding protein 2	2	10
IGFBP3	insulin-like growth factor binding protein 3	7	4
IGFBP4	insulin-like growth factor binding protein 4	17	8
IGFBP5	insulin-like growth factor binding protein 5	2	10
IGFBP6	insulin-like growth factor binding protein 6	12	5
IL1A	interleukin 1, alpha	2	9
IL1B	interleukin 1, beta	2	10
IL1R1	IL1 receptor type I	2	19
IL1R2	IL1 receptor type II	2	12
IL1RN	IL1 receptor agonist	2	12
IL6	Interleukin 6	7	11
IL6R	IL6 receptor gp80 and gp130	1	10
IL6ST	interleukin 6 signal transducer	5	6
IRAK1	interleukin-1 receptor-associated kinase 1	X	1
IRAK3	interleukin-1 receptor-associated kinase 3	12	11
LHB	luteinizing hormone beta polypeptide	19	5
LHCGR	luteinizing hormone/choriogonadotropin receptor	2	29
LIF	leukemia inhibitory factor (cholinergic differentiation factor	22	14
LIFR	leukemia inhibitory factor receptor	5	15
LRP5	low density lipoprotein receptor-related protein 5	11	11
LRP6	low density lipoprotein receptor-related protein 6	12	18
LTBP1	latent transforming growth factor beta binding protein 1	2	46
LTBP2	latent transforming growth factor beta binding protein 2	14	21

Table 3.1 continued

gene	Full Name	chr	# of SNPs
LTBP3	latent transforming growth factor beta binding protein 3	11	5
MC2R	melanocortin 2 receptor (adrenocorticotropic hormone)	18	4
MEF2A	MADS box transcription enhancer factor 2, polypeptide A	15	9
MEF2B	MADS box transcription enhancer factor 2, polypeptide B	19	7
MEF2C	MADS box transcription enhancer factor 2, polypeptide C	5	12
MEF2D	MADS box transcription enhancer factor 2, polypeptide D	1	14
MYF5	myogenic factor 5	12	5
MYF6	myogenic factor 6 (herculin)	12	6
MYOD1	myogenic differentiation 1	11	5
MYOG	myogenic factor 4	1	2
NCOA1	nuclear receptor coactivator 1	2	11
NCOA2	nuclear receptor coactivator 2	8	20
NCOA3	nuclear receptor coactivator 3	20	16
NFKB1	Nuclear factor-k-B	4	14
NR3C1	Glucocorticoid receptor (GR)	5	10
NTF3	neurotrophin 3	12	11
NTRK1	neurotrophic tyrosine kinase, receptor type1	1	12
NTRK2	neurotrophic tyrosine kinase, receptor type2	9	16
NTRK3	neurotrophic tyrosine kinase, receptor type3	15	44
POMC	proopiomelanocortin	2	6
PRKAG1	protein kinase, AMP-activated, gamma 1 non-catalytic subunit	12	1
SHBG	sex hormone-binding globulin	17	7
SMAD1	SMAD, mothers against DPP homolog 1 (Drosophila)	4	10
SMAD2	SMAD, mothers against DPP homolog 2 (Drosophila)	18	9
SMAD3	SMAD, mothers against DPP homolog 3 (Drosophila)	15	25
SMAD4	SMAD, mothers against DPP homolog 4 (Drosophila)	18	3
STAR	steroidogenic acute regulator	8	3
STS	Steroid sulfatase	X	2
TGFB1	Transforming growth factor-β	19	6
TGFB2	Transforming growth factor-β	1	18
TGFB3	Transforming growth factor-β	14	9
TGFBR1	TGF-receptors	9	7
TGFBR2	TGF-receptors	3	23
TGFBR3	TGF-receptors	1	40
THRA	thyroid hormone receptor, alpha	17	7
TIEG	Kruppel-like factor 10	8	10
TNFA	Tumor necrosis factor-α	6	7
TNFRSF11A	TNFSF11 receptor	18	13
TNFRSF11B	Osteoprotegerin	8	12
TNFRSF1A	tumor necrosis factor receptor superfamily, member 1A	12	9
TNFRSF1B	tumor necrosis factor receptor superfamily, member 1B	1	22
TNFSF10	tumor necrosis factor (ligand) superfamily, member 10	3	12
TNFSF11	Osteoprotegerin ligand	13	7
VDR	Vitamin D receptor	12	20
WNT10B	wingless-type MMTV integration site family, member 10B	12	6
WNT1B	wingless-type MMTV integration site family, member 1B	12	1
WNT3A	wingless-type MMTV integration site family, member 3A	1	6
WNT5A	wingless-type MMTV integration site family, member 5A	3	10

Table 3.2: Table of self reported and genetically assigned ancestry groups

	Ancestry Group		
	African- American	**Caucasian American**	**Unknown**
Self reported	1180	1653	146
Genetic (based on 429 tag SNPs)	1237	1742	0

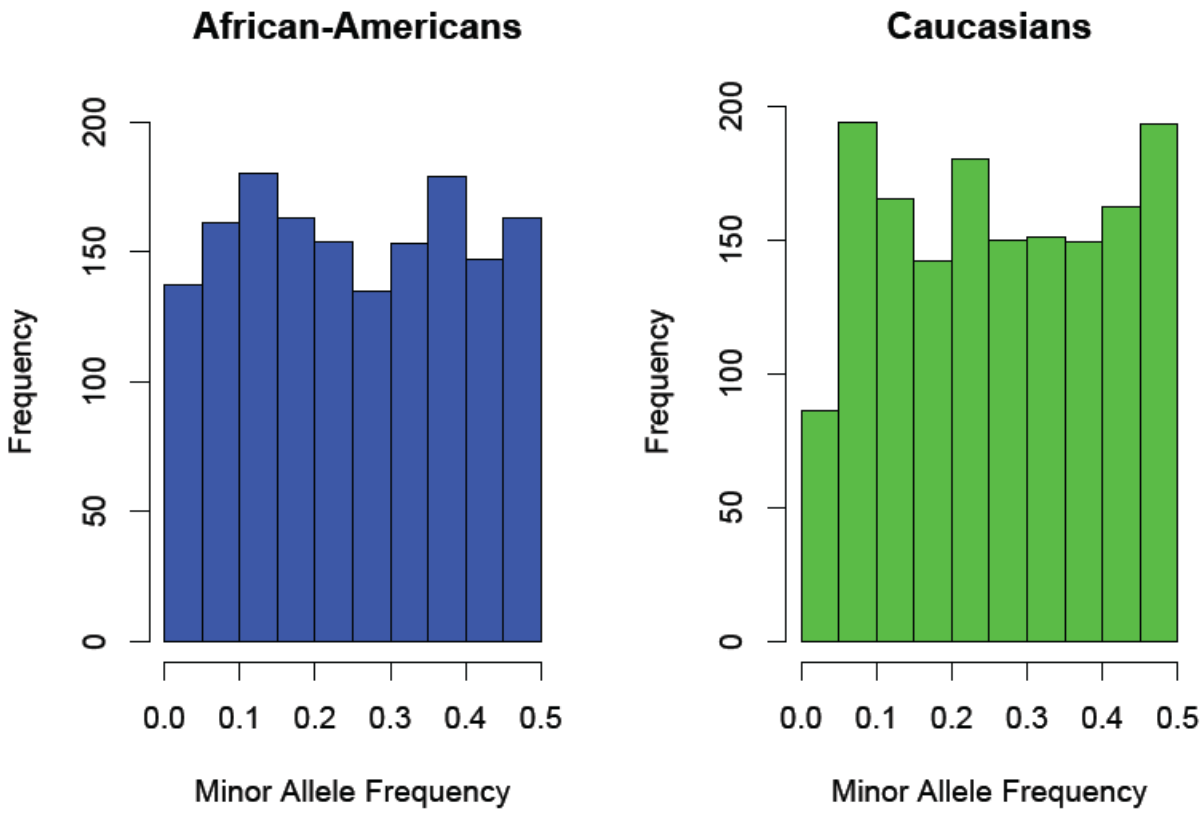


Figure 3.1: Histogram for minor allele frequencies in the two self-reported race groups

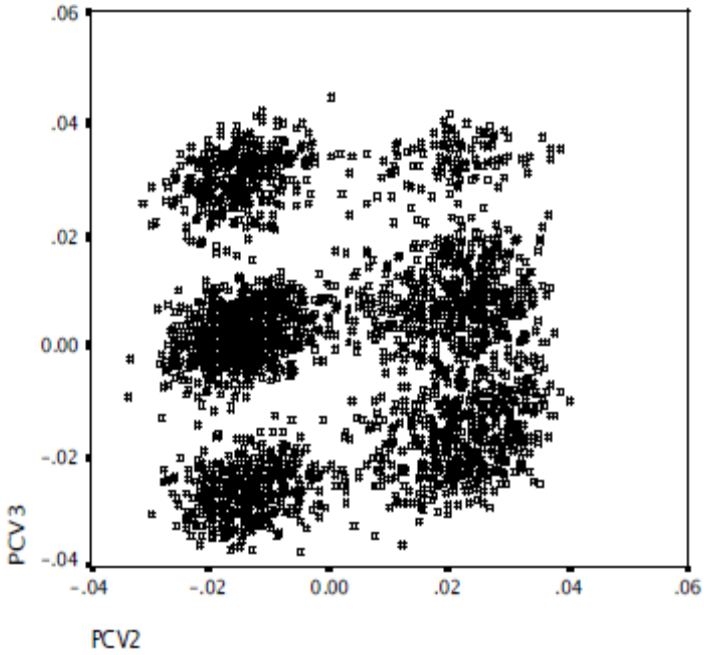


Figure 3.2: Plot of second and third eigen vectors when SNPs are correlated

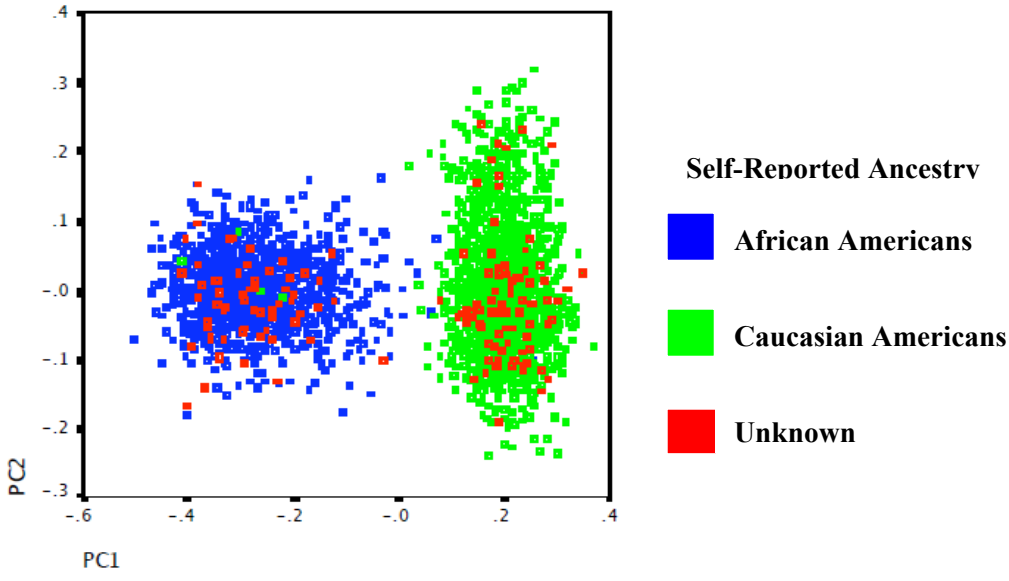


Figure 3.3: Estimated genetic ancestry (represented by the first two eigen analysis components) versus self-reported ancestry

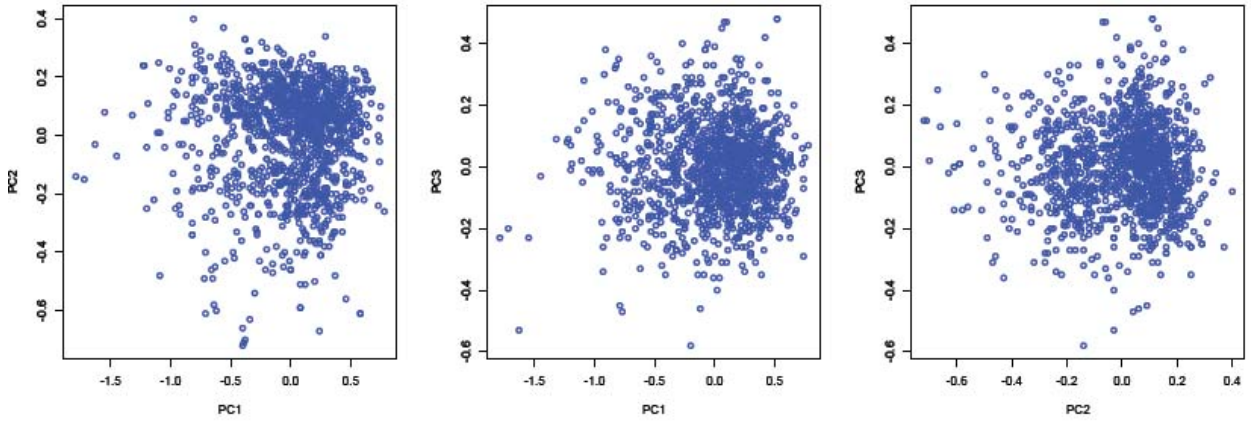


Figure 3.4: Pairwise plots of first three principal components for individuals in the African American ancestry group

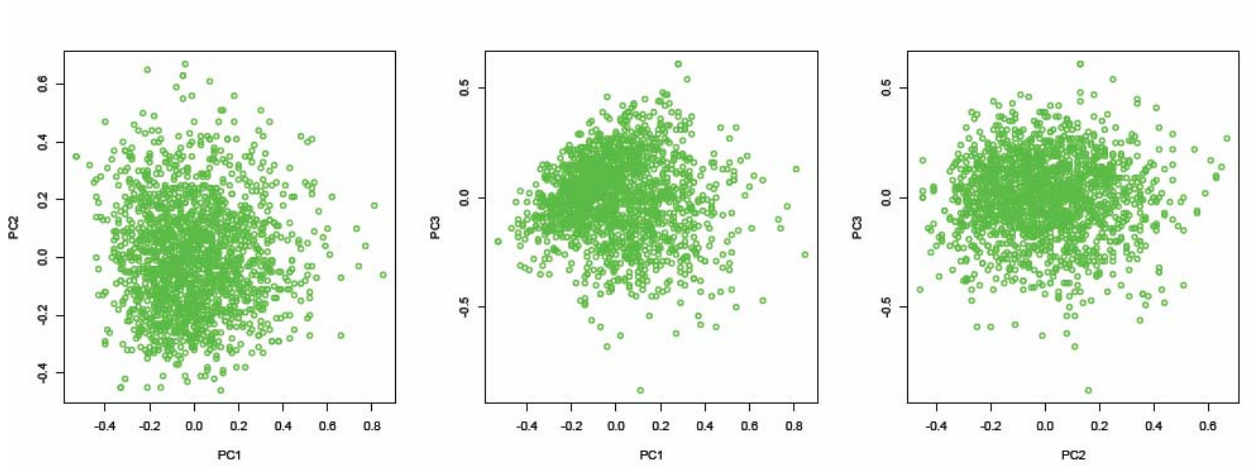


Figure 3.5: Pairwise plots of first three principal components for individuals in the Caucasian American ancestry group

4.0 ANALYZING THE EFFECTS OF COVARIATES ON BMD ACROSS DIFFERENT ETHNIC AS WELL AS GENDER GROUPS

4.1 INTRODUCTION

As described in the Introduction to this dissertation, osteoporosis is a major public health concern and is characterized by decline in bone mass, which can result in fragility of bone. More than 10 million Americans may develop osteoporosis each year. Low level of BMD is a key factor in development of osteoporosis, so studying the genetic as well as environmental factors affecting BMD is very important.

Several investigators have reported that the risk of osteoporosis varies by ethnicity and sex. Mean BMD at total hip is higher among African-Americans than Caucasian Americans (Araujo et al. 2007). Women tend to loose BMD at a more rapid rate and Caucasian Americans also loose BMD more rapidly than African Americans (Cauley et al. 2005; Evans et al. 2005; Wilkins et al. 2005). There are several lifestyle and anthropometric factors that influence BMD level and osteoporosis. A previous longitudinal study on 116229 nurses showed significant increase in risk of hip fracture (Cornuz et al. 1999). Current smoking also increases bone resorption among males (Szulc et al. 2002). Vitamin D and calcium intake significantly increases the level of BMD at whole body and vertebral sites (Grados et al. 2003). The risk of hip fracture is reduced with intake of calcium (Feskanich et al. 2003). Low weight is a significant risk factor for low levels of BMD and fracture at spine and hip sites. Short height is associated with low level of BMD at femoral neck and lumbar spine (Ho et al. 2005). Furthermore, the effects of covariates may also differ by age and sex. For example, in 45-92 year old individuals, serum leptin levels are associated with BMD in women but not in men (Weiss et al. 2006). Likewise, postmenopausal women have a

positive association of cyclooxygenase-2 use and BMD but the association is reversed in older men (Richards et al. 2006). As described in the introduction to this dissertation there are two components of bone: trabecular and cortical. Genetic and environmental components affect the two types of bone differently (Duan et al. 1999). The proportion of bone components varies across different skeletal sites (Silverberg et al. 1995), so studying the different bone sites is important.

The goal of the current study is to analyze the effect of different risk factors on bone traits at three different time points separately for originally healthy U.S. African Americans and Caucasian Americans age greater than 68 years of age. We also investigated the effects of the covariates on the BMD traits over time. We know that the epidemiology of BMD differs between different ethnic groups, but how different covariates is affecting the BMD of different sex and ethnic groups is not very well characterized. We would like to answer questions such as how different covariates influence bone traits in different sex and ethnic groups (Araujo et al. 2007). We will also try to analyze how the effect of the covariates is changing over time. A comparative study of different ethnic and sex groups is important to further investigate the genetic causes behind high or low BMD values.

4.2 DATA AND METHODS

The Health Aging and Body Composition Study (Health ABC) is an observational cohort study. The data has been collected over six different time points (1 year apart) on African American and Caucasian American men and women in the U.S. At baseline, data are available on 2830 individuals (1653 Caucasian Americans [504 men and 673 women] and 1177 African-Americans [872 men and 781 women]) between the ages of 68 to 80 years old. The data were collected at two sites, Pittsburgh and Memphis.

Data on different anthropometric measurements, lifestyle, medical characteristics and different fat, muscle and bone related traits were collected. Exercise level (kcal/kg/week) was estimated based on a

questionnaire administered by the interviewer. Height and weight were estimated using a stadiometer (Harpenend, Wales, UK) and balance beam scale, respectively, without wearing any shoes. The difference between sitting height and a 30 cm seat was denoted as trunk length and the difference between total height and trunk length was denoted by leg length. Smoking and drinking history were assessed by self report. Race and ethnicity were also assessed by self-report, but we calculated genetic ancestry from genotype data and used genetic ancestry to classify individuals in our study into two ancestry groups (see chapter 3). Physical activity level was estimated based on questionnaire administered by an interviewer (Visser et al. 2005). Medications were assessed by transcription of all prescription. All participants provided written informed consent for examination. Institutional Review Boards at the University of Pittsburgh and University of Tennessee Health Science Center approved all protocols.

Prior to doing analysis we plotted several transformations of BMD traits, but for each trait, the original measurement appeared to be approximately normally distributed. We started with 3075 individuals with phenotypic data, but after removing outliers (any individual with a value ±4 sd from the mean), merging with the genotype file and reclassifying individuals according to genetic ancestry (as described in chapter 3) we were left with 2830 individuals. All subsequent analyses were performed on this set of 2830 individuals. We analyzed the effects of covariates on different bone traits by using linear regression methods after dividing the subjects into four groups according to their genetic ancestry (as defined in chapter 3) and sex because of known differences or possible differences in the effects of covariates on different ethnic groups or between men and women (see Introduction above).

We fit the multiple regression model as follows:

$$y_i = \sum_{j=1}^{p} X_{ij}\beta_j + \varepsilon_i ,$$

where y is the trait of interest (eg. Hip BMD), X is the matrix of constants and covariates, β is the parameter to be estimated. There are a total of p parameters, one intercept and the remaining p-1 parameters correspond to covariates, ε_i is the observational error. We did not use stepwise regression because order of parameters and the method used can affect the selected model [Derksen et al. 1992].

First we included a larger list of covariates in the model, and then left out those covariates in the final model building which had p-value > 0.1.

We also wanted to see how proportion of the total phenotypic variation of the BMD traits is explained by different categories of covariates. We grouped our covariates into five categories (ancestry, anthropometrics, demographics, lifestyle and medical history). To calculate the effect of a particular category we fitted the above regression model taking all the covariates together and then dropped covariates of different groups one by one. We dropped the ancestry group at first and then subtracted the R^2 from the resulting submodel from that of the full model. Then we dropped the demographics covariates and calculated the resulting R^2 and subtracted it from that of the model we got by deleting the ancestry covariates, and so on.

In addition to data obtained at baseline, follow up data on areal BMD were available at two years and four years after the baseline data were collected. Therefore, we had the opportunity to determine whether effect of different covariates changes over time, as well as whether they had effects in a cross-sectional analysis. These longitudinal analyses were also performed within the four sex and ethnic groups separately. Because we did not have measures of all the covariates at different time points, we assumed that there was no interaction between the covariates and time in our models. We analyzed the effect of the covariates by using observations over three time points. Generalized Estimating Equation (GEE) methods were used to analyze the longitudinal effects of the covariates. We included time as one of the covariates in addition to any significant covariates.

Because the observations over three time points are correlated, we needed to specify a correlation structure. GEE is very robust against wrongly specifying the correlations structure because estimates of the regression parameters remain consistent. Thus there is little gain in efficiency by specifying the correct correlation (Zeger et al. 1986). We specified the correlation structure as unstructured because then all possible correlations between all the different responses for a specific subject are considered. These unstructured correlations are then used in estimating the model parameters (Fitzmaurice et al. 1993).

Because our response variables (the areal BMD traits) were normally distributed, we specified the link function as identity.

The model is given as

$$\mu_i = E(y_i),\ i = 1 \ldots\ldots n$$

$$h(\mu_i) = X_i\beta$$

where $Y_i = [y_{i1}.....y_{iT}]$, Y_i is the set of measurements for the i^{th} individual. T is the number of repeated observation for each individual, h(.) is the link function, in our case it is identity. The correlation matrix between Y_{ij}'s is given by R_i , we specified it as unstructured.

4.3 RESULTS

Anthropometric, demographic, lifestyle, and medical history characteristics obtained at baseline for the Caucasian American and African American men and women are given in Table 4.1. The number of individuals from the two ancestry groups recruited from Pittsburgh and Memphis were almost equal in number, that is, 846 versus 807 Caucasian Americans (1653 total) and 557 versus 620 (1177 total) African Americans, respectively. The average ages for all the groups were similar. Men from both the ancestry groups are taller than their women counterparts. Caucasian American women have the lowest weight of the four groups. Among Caucasian Americans fewer women (77.2%) than men (80.1%) consume alcohol and the difference is larger in African-Americans: men (83.9%) and women (64.8%). As expected, the prevalence of smoking history is higher among men (69.8 and 71.2%) compared to women (44.1 and 41.7%) in both ancestry groups. Caucasian American men and women have a higher level of exercise than their African-American counterparts. African-Americans have a higher prevalence of diabetes and high blood pressure than Caucasian American, but Caucasian Americans have a higher

prevalence of cancer. Overall, men have a higher rate of CHD than women in their respective ancestry groups. These results are similar to results from the Framingham osteoporosis study data (Tucker et al. 2006). The medical condition covariates were all self reported.

The raw means and standard deviation of BMD measured at 15 skeletal sites measured at baseline is presented in Table 4.2. We analyzed the effects of covariates on areal BMD at multiple skeletal sites that most investigators do not report because the power of one of our SNP association methods (the principal component of heritability) should be increased using multiple measures, as long as the measures are not highly correlated (Klei et al. 2008). However, for ease of understanding and comparison with results from other studies, for the remainder of this Chapter, only results of analyses performed on areal BMD of the whole body, lumbar spine, total hip, femoral neck, mean arm and mean leg will be presented. Results of analyses of other skeletal sites are presented in tables in the Appendix.

As can be seen from Table 4.2, mean BMD at all skeletal sites are higher in males than females and also higher in African-Americans compared to Caucasian Americans. These results are similar to those obtained from NHANES (Looker et al. 1998). I also examined relationship of mean BMD at different time points. As can be seen from Figure 4.1, for most of the skeletal traits across all sex by ancestry groups, mean BMD value gradually decreases with time, except for lumbar spine BMD.

We then used linear regression analyses of the three different time points BMD data to estimate the effects of different covariates. As described in Chapter 3, we included the first three principal components from our ancestry analysis as covariates to see if there is an effect of population substructure on the BMD traits. However, we included them in the final model, even if they were not significant, to control for effects of population substructure.

4.3.1 Results from analyses of covariates at baseline

The results of regression analyses of effects of covariates on BMD at six skeletal sites measured at baseline (or time point 1) are presented in Table 4.3. In general, results for the other two time points (time

points 3 and 5) were similar to those obtained at baseline, and these results are presented in the Appendix. In addition, results for the other skeletal sites at all three time points are also presented in the Appendix. As stated above, these analyses were performed within each sex by ancestry group separately.

a) Caucasian American Males:

As can be seen in Table 4.3A population substructure does not have any significant effect on any of the BMD traits at $p \leq 0.05$, although all PCs are significant at $p \leq 0.10$. Age has significant effect ($p < 0.01$) on mean arm BMD. Site does have a significant effect on all the traits ($p<0.05$ for all), except femoral neck BMD and arm BMD ($p \geq 0.05$). As has been observed in many studies (Looker et al. 1998), BMD at all skeletal sites increased with increasing weight. In contrast, as waist circumference increases, BMD significantly deceases at all the sites except lumbar spine. Height has significant positive effect on whole body, mean arm and mean leg BMD. Among the lifestyle factors, only smoking had significant negative effect ($p < 0.01$) at all sites except lumbar spine. Among the prevalent diseases, osteoporosis status has significant negative effects at all skeletal sites ($p < 0.001$). Calcium supplement have a significant positive effect on mean leg BMD ($p < 0.05$).

b) Caucasian American Females:

Among Caucasian American women (Table 4.3B), population substructure also does not have any significant effect on any of the BMD traits ($p > 0.05$), but is significant at a more liberal $p < 0.10$. Age and site both have significant effects on lumbar spine BMD, site also has significant effect on mean leg BMD (p value < 0.01). Among the anthropometric factors, weight significantly affects all the BMD traits, as is observed in all the other groups. Height has significant positive effect at all the BMD sites except for femoral neck and hip BMD. Among the lifestyle factors, none have any significant effect on any of the traits, except for prevalence of smoking history on hip BMD ($p < 0.05$). Among the prevalent diseases, osteoporosis status has significant negative effects at all skeletal sites ($p < 0.001$). Likewise, prevalent

osteoarthritis at knee and oral estrogen use both have significant positive effect on BMD at all skeletal sites.

c) African-American Males:

Among African American men (Table 4.3C), neither population substructure nor age, have any significant effects ($p \geq 0.05$) on any of the areal BMD traits. Site does have a significant effect on all the BMD traits except femoral neck BMD ($p \geq 0.05$). The regression coefficient for all the traits is positive, so it means that on average individuals from Pittsburgh have higher BMD values than their Memphis counterparts. Increasing weight is significantly associated with increasing BMD at all sites. In contrast, as waist circumference increases, BMD significantly deceases at the whole body ($p < 0.001$), lumbar spine ($p < 0.001$), mean arm ($p < 0.01$) and mean leg ($p< 0.01$), skeletal sites. Height has significant negative effect ($p < 0.01$) on hip BMD, but at all other skeletal sites height was not significant. Among the lifestyle factors, only exercise had significant effect, increased exercise activity was correlated with increased BMD ($p < 0.05$ at whole body and lumber spine). Among the prevalent diseases, osteoporosis status is associated with a significant decrease in BMD at all skeletal sites ($p < 0.001$).

d) African-American Females:

Unlike the other three groups, the first principal component of population substructure is significantly correlated ($p \leq 0.05$) with all of the BMD traits, except mean leg. Similar to African- American males age does not have any significant effect ($p \geq 0.05$) on any of the traits, but site does have a significant positive effect on all BMD traits except femoral neck and mean arm BMD. When considering the anthropometric factors, weight significantly affects all the BMD traits. Waist circumference does not have any significant effect on any of the skeletal sites, but increasing height is correlated with increasing mean arm ($p < 0.001$) and lumbar spine ($p < 0.05$) BMD. Among the lifestyle factors, exercise has significant effect on mean arm BMD ($p < 0.01$) and mean leg BMD ($p < 0.05$); increased exercise activity was correlated with

increased BMD. As expected, osteoporosis status is associated with decreased BMD at all skeletal sites ($p < 0.001$). Prevalent osteoarthritis at the hip is significantly associated with increased BMD at all sites. In contrast, prevalent diabetes is associated with decreased BMD at the hip only.

4.3.2 Proportion of total phenotypic variation attributable to different covariate categories

The proportion of total variation in areal BMD explained by different sets of covariates is given in Table 4A, B, C and D. In general, measured covariates accounted for less of the total BMD variation in Caucasian American men (10-26%) than any of the other three groups (23-53%). Measured covariates accounted for the most variation among Caucasian American women (27-44%). This difference in total percent variation is due the proportion of variance in areal BMD accounted for by medical conditions which ranges from 22-47% in Caucasian American women to 13-22% in African American women to 12-19% in African American men to 4-8% in Caucasian American men. In fact, medical conditions accounted for the largest proportion of variance among all five categories of covariates among Caucasian American women.

The category that accounted for the next largest proportion of variation across all groups (except Caucasian American women) was the set of anthropometric covariates (height, weight, and waist circumference). This category accounted for the smallest proportion of variance in Caucasian American women (3-6%), and moderate among the other three groups: 5-16% in Caucasian American men, 7-13% in African American men, and 7-17% among African American women. Anthropometric covariates accounted for the largest proportion of variance in areal BMD among Caucasian American men.

The remaining categories of covariates each accounted for < 2% of the total variation in areal BMD at all sites in all ancestry by sex groups.

4.3.3 Results from longitudinal analysis of BMD

Because we have data on individuals at three time points at two-year intervals, we had the opportunity to investigate whether time has any effect on mean BMD and also to test for effects of covariates across time. Results from the longitudinal analyses of the six BMD sites for each sex by ancestry group are presented in Table 4.5a, b, c, and d.

As expected, time is a significant covariate for all the groups across all the skeletal sites, except femoral neck, hip and mean arm sites for Caucasian American females and at hip BMD for African-American females. Also, the effects of demographic, anthropometric, lifestyle, and medical history covariates were similar to those reported in the analysis of baseline data (Table 4.3); except that the effect of weight is no longer significant for many of the traits. So overall we conclude that although time plays a very important role in variation in areal BMD in all groups, the effect of other covariates remains similar.

4.4 DISCUSSION

Data on BMD at different skeletal sites on the Health ABC cohort provides an opportunity to analyze both cross-sectional and longitudinal effects of covariates in healthy aged populations of African-Americans and Caucasian Americans. Because the same covariates and BMD traits were measured on all individuals, we can assess the similarities and differences these covariates may have in males and females and in two different ancestry groups. We can also assess whether the effects of these covariates differ at different skeletal sites and whether the effects of the covariates measured at baseline has varying effect on traits across time, although we could not test for interactions between the covariates and time because we did not have independent measures of the covariates at all time points.

In general, as has been reported by several groups (Melton et al. 2000; Duan et al. 2001; Tuck et al. 2005), including previous analyses of the baseline HABC data, the mean level of BMD is lower in

women than in men at all skeletal sites and at all three time points. Furthermore, as expected from reports of other investigators (Melton et al. 2001; Pothiwala et al. 2006), Caucasian Americans have lower BMD levels than African-Americans, and these differences between sexes and between ancestry groups remained even after incorporating effects of other significant covariates such as weight, height, recruitment site, lifestyle, and medical history.

We also found that smoking, drinking and low level of physical activity are associated with decreased BMD at most of the skeletal sites in both ancestry groups as well as both the sexes. The results tally with previous results regarding the effect of covariates on BMD traits found in other studies of the HABC population (Strotmeyer et al. 2004) as well as studies of older individuals in other cohorts (Brown et al. 2004; Ng et al. 2006). We did not detect a significant effect of age on any of the cohorts, most likely because the population is older (average age = 73 years), so BMD does not change significantly over different ages.

Consistent with other reports, our results also indicate that bone metabolism differs between different ancestry groups, as well as between men and women. How bone metabolism differs among the groups and whether different genes act differently in the groups is investigated in Chapters 5 and 6 in which we report the results of candidate gene association analysis performed on each of the four ancestry by sex groups separately.

One of the primary objectives of these covariate analyses was to identify factors that may differentially influence BMD in a healthy, aged population of men and women comprising two ancestry groups, Caucasian Americana and African American. These significant covariates were then used to generate residuals that were analyzed for associations with specific genotypes at candidate loci for musculoskeletal traits (Chapters 5 and 6). In other words, we wanted to remove the effects of known covariates to increase the power to detect genetic signals due to variation at the candidate loci. We realize that the choice of which covariates to include in the model to generate residuals could also influence the probability of detecting an effect due to genetic variation at the candidate genes. We also included

osteoporosis by T-score as one of the covariates, because we wanted to see the effect of genetic variants after removing the effect of this classification variable.

4.5 TABLES AND FIGURES

Table 4.1: Characteristics of the Covariates in HABC men and women by ethnic group measured at baseline

GROUP	African American		Caucasian American	
SEX	Male (n=504)	Female (n=673)	Male (n=872)	Female (n=781)
Covariates				
Demographics				
Age,yrs	73.5±2.8	73.4±3	73.9±2.9	73.6±2.8
Site(% Memphis)	49.4	45.8	50.3	52.1
Anthropometrics				
Height,mm	1730.7±70.6	1596.5±64.4	1735.1±63.9	1594.4±59.1
Weight,kg	81.8±14.8	75.8±15.9	81.3±12.4	66.3±12.2
Waist Circumference,cm	99.6±12.7	101.2±14.9	101.7±11.7	95.9±12.5
Life style				
Drinking History(%)	80.1	77.2	83.9	64.8
Smoking History(%)	69.8	44.1	71.2	41.7
Exercise,kcal/kg/wk	79.7±75.7	80.2±74.1	82.9±64.6	86.2±62.9
Medical Conditions				
Prevalent CHD(%)	22.9	19.5	28.3	11.9
Prevalent CVD(%)	7.9	8.8	7.6	7.2
Prevalent Diabetes (%)	21.8	21	13.3	7.7
Prevalent Osteoarthritis, knee(%)	4.4	12.5	5.9	13
Prevalent Osteoarthritis, hip(%)	2.2	4.3	3	7.6
Prevalent HBP(%)	55.7	66.2	42	44.2
Prevalent Cancer(%)	14.3	8.8	27.1	21.6
Prevalent Clinical Osteoporosis(%)	2	8.7	2.2	25.9
Osteopporosis by T-score(%)	4.4	6.4	1.3	15.9
Any Osteoporosis Drugs(%)	1	1.8	0.8	12.2
Calcium Supplement(%)	3.2	12.9	9.6	41.9
Vitamin D Supplement(%)	1.6	5.1	4.7	19.5
Oral Estrogen(%)	0.2	12	0	31

Table 4.2: Table of mean and standard deviation of raw BMD traits at baseline for the four sex by ancestry groups in HABC data

Group	African American		Caucasian American	
Sex	Males	Females	Males	Females
BMD at baseline				
	n=504	n=673	n=872	n=781
Whole Body	1.21±0.13	1.04±0.11	1.15±0.11	0.98±0.10
Head	2.34±0.40	2.31±0.45	2.14±0.34	2.13±0.39
Left Arm	0.88±0.08	0.68±0.07	0.82±0.07	0.63±0.06
Mean Arm	0.88±0.08	0.69±0.07	0.83±0.07	0.64±0.06
Left Leg	1.28±0.13	1.04±0.13	1.25±0.13	0.98±0.11
Mean Leg	1.27±0.13	1.04±0.13	1.25±0.13	0.98±0.11
Left Rib	0.73±0.09	0.62±0.07	0.68±0.08	0.57±0.08
Mean Rib	0.74±0.09	0.62±0.07	0.68±0.07	0.57±0.07
Lumbar Spine	1.15±0.23	1.00±0.19	1.08±0.21	0.91±0.17
Thoracic Spine	1.01±0.17	0.85±0.14	0.94±0.14	0.78±0.12
Pelvic	1.30±0.22	1.15±0.17	1.20±0.17	1.07±0.15
Femoral Neck	0.85±0.14	0.75±0.13	0.76±0.13	0.65±0.11
Intertrochanteric	1.19±0.18	1.01±0.18	1.10±0.17	0.91±0.16
Ward's Triangle	0.62±0.17	0.55±0.16	0.53±0.14	0.47±0.13
Hip	1.02±0.16	0.86±0.15	0.94±0.14	0.77±0.13

Table 4.3: Association of Covariates of BMD traits at baseline for four ancestry by sex groups

****** = p-value < 0.001; ***= 0.001 < p-value < 0.01; **= 0.01< p-value <0.05.**

4.3A Caucasian American Male

GROUP	Caucasian Male					
BMD Traits	Femoral Neck	Hip	Whole Body	Lumbar Spine	Mean Arm	Mean leg
Covariates						
Population Substructure						
PC1	4.24E-03	-4.77E-02	-1.39E-01	2.60E-01	2.93E-04	-1.21E-01
PC2	1.51E-01	1.32E-01	1.38E-02	-9.97E-02	1.15E-01	1.17E-01
PC3	-1.51E-01	1.52E-02	-1.54E-02	5.66E-02	-3.64E-02	2.04E-02
Demographics						
Age	-1.34E-03	-1.60E-03	-2.52E-03	2.55E-03	-2.22E-03 ***	-3.49E-03 **
Site	-1.86E-03	3.59E-02 ****	2.28E-02 ***	3.94E-02 **	2.44E-03	5.41E-02 ****
Anthropometrics						
Height	1.09E-05	-9.27E-05	1.87E-04 ***	2.58E-04	1.67E-04 ****	2.11E-04 ***
Weight	4.38E-03 ****	5.95E-03 ****	2.86E-03 ****	4.31E-03 ****	1.76E-03 ****	3.62E-03 ****
Waist Circumference	-1.08E-03 **	-1.71E-03 ***	-1.87E-03 ****	-1.46E-03	-1.10E-03 ****	-1.98E-03 ****
Life style						
Education	3.83E-03	9.23E-04	3.97E-03	4.41E-03	-8.26E-03 **	8.60E-04
Smoking History	-3.87E-04 ***	-4.84E-04 ***	-3.70E-04 ***	-3.87E-04	-2.30E-04 ***	-4.95E-04 ****
Exercise	5.88E-05	3.65E-05	-3.42E-05	-3.09E-05	1.61E-05	-5.17E-05
Medical Conditions						
Prevalent CHD	1.61E-02 **	9.71E-03	1.51E-02 **	1.16E-02	5.40E-03	1.87E-02 **
Prevalent Diabetes	1.66E-02	2.10E-02	5.19E-03	3.56E-02	2.59E-03	2.02E-02
Prevalent Osteoarthritis, knee	1.69E-02	1.16E-02	6.70E-03	1.15E-02	5.73E-03	5.91E-03
Prevalent Osteoarthritis, hip	-1.81E-02	-1.56E-02	-4.05E-03	1.88E-02	-3.88E-04	-1.02E-02
Prevalent HBP	-1.16E-03	3.49E-03	2.91E-03	2.60E-02 **	2.13E-03	-3.94E-03
Osteopporosis by T-score	-1.45E-01 ****	-2.34E-01 ****	-1.61E-01 ****	-1.66E-01 **	-1.18E-01 ****	-1.90E-01 ****
Calcium Supplement	1.26E-02	1.95E-02	3.58E-02 **	4.13E-02	1.32E-02	3.93E-02 **

4.3B Caucasian American Female

GROUP	Caucasian Female					
BMD Traits	Femoral Neck	Hip	Whole Body	Lumbar Spine	Mean Arm	Mean leg
Covariates						
Population Substructure						
PC1	1.67E-03	1.07E-01	-8.49E-02	3.58E-01	-1.59E-02	-9.05E-02
PC2	6.02E-02	-2.71E-02	-6.81E-02	-1.37E-02	-1.17E-01	-8.08E-02
PC3	-2.05E-02	-8.81E-02	-1.09E-01	-6.43E-02	-7.35E-02	-9.94E-02
Demographics						
Age	-2.75E-04	-4.67E-04	-4.57E-04	6.00E-03 ***	6.16E-05	-1.16E-03
Site	-1.36E-02	1.00E-02	4.90E-03	2.89E-02 **	-4.22E-03	2.46E-02 ***
Anthropometrics						
Height	5.36E-05	1.44E-05	2.14E-04 ****	2.40E-04 **	1.66E-04 ****	1.91E-04 ***
Weight	3.01E-03 ****	3.67E-03 ****	1.51E-03 ***	2.82E-03 ***	9.44E-04 ***	2.21E-03 ****
Life style						
Drinking History	6.81E-05	7.14E-05	1.45E-05	-4.58E-05	5.26E-05	3.47E-05
Smoking History	7.44E-03	1.04E-02 **	3.17E-03	-2.80E-03	4.68E-03	7.60E-03
Exercise	2.17E-03	-4.94E-05	3.65E-03	1.53E-02	3.47E-03	4.13E-03
Medical Conditions						
Prevalent Diabetes	1.70E-02	3.38E-02 **	2.76E-02 **	4.27E-02	1.23E-02	2.31E-02
Prevalent Osteoarthritis, knee	1.59E-02 ***	1.35E-02 **	1.00E-02	3.41E-02 ***	1.27E-02 ****	1.48E-02 **
Prevalent Osteoarthritis, hip	-1.49E-02 **	-9.62E-03	-7.35E-03	1.12E-02	-9.83E-03 **	-1.33E-02 **
Prevalent Cancer	1.86E-02 **	1.51E-02	1.80E-02 **	1.16E-02	1.10E-03	1.27E-02
Prevalent Clinical Osteoporosis	-1.15E-02	-9.52E-03	-8.46E-03	-6.09E-03	-2.22E-03	-7.49E-03
Osteopporosis by T-score	-1.30E-01 ****	-1.80E-01 ****	-1.03E-01 ****	-1.43E-01 ****	-5.78E-02 ****	-1.12E-01 ****
Calcium Supplement	-1.53E-02	-1.49E-02	-1.17E-03	-1.06E-02	-1.06E-02 **	-4.36E-03
Vitamin D Supplement	2.09E-02 **	1.35E-02	1.35E-03	1.63E-02	3.00E-03	6.86E-04
Oral Estrogen	2.64E-02 ****	3.40E-02 ****	6.52E-02 ****	7.42E-02 ****	3.51E-02 ****	5.12E-02 ****

4.3C African-American Male

GROUP	African-American Male					
BMD Traits	Femoral Neck	Hip	Whole Body	Lumbar Spine	Mean Arm	Mean leg
Covariates						
Population Substructure						
PC1	1.58E-01	1.67E-01	-5.63E-02	-3.50E-03	4.99E-02	-7.25E-02
PC2	4.04E-01	2.20E-01	2.97E-01	5.89E-01	1.20E-01	3.06E-01
PC3	-2.88E-01	-1.54E-01	1.28E-01	3.75E-01	-2.90E-03	7.06E-02
Demographics						
Age	1.00E-04	1.70E-03	2.90E-03	7.20E-03	-8.00E-04	3.10E-03
Site	6.60E-03	3.59E-02 ***	3.71E-02 ****	5.24E-02 **	1.97E-02 ***	5.99E-02 ****
Anthropometrics						
Height	-2.00E-04	-3.00E-04 ***	0.00E+00	-1.00E-04	0.00E+00	0.00E+00
Weight	4.50E-03 ****	5.80E-03 ****	4.20E-03 ****	7.80E-03 ****	2.90E-03 ****	4.10E-03 ****
Waist Circumference	-1.40E-03	-1.50E-03	-3.30E-03 ****	-5.70E-03 ****	-1.70E-03 ***	-2.60E-03 ***
Life style						
Drinking History	-2.30E-03	-3.80E-03	2.10E-03	8.40E-03	2.40E-03	-2.00E-04
Smoking History	-2.50E-03	-3.10E-03	-1.05E-02	-2.00E-02	-2.70E-03	-8.60E-03
Exercise	3.00E-04	1.00E-04	6.00E-04 **	1.30E-03 **	3.00E-04	5.00E-04
Medical Conditions						
Prevalent CHD	1.08E-03	1.51E-02	1.75E-03	-1.93E-02	3.46E-03	1.25E-02
Prevalent CVD	-2.42E-04	-2.81E-03	1.87E-02	2.30E-03	1.40E-02	3.39E-02
Prevalent Diabetes	2.21E-02	2.44E-02	1.05E-02	4.07E-02	-4.14E-03	1.23E-02
Prevalent Osteoarthritis, knee	6.50E-03	2.88E-03	5.06E-03	3.29E-02	4.25E-03	-1.25E-03
Prevalent Osteoarthritis, hip	-3.66E-02 **	-3.62E-02	-2.31E-02	-2.90E-02	-1.72E-02	-2.19E-02
Prevalent HBP	1.08E-03	-5.89E-03	-7.27E-03	-5.63E-03	-4.94E-03	-5.24E-03
Prevalent Cancer	-1.82E-02	-2.64E-02	-2.86E-02	-4.93E-02	-2.62E-02 ***	-3.32E-02 **
Prevalent Clinical Osteoporosis	6.75E-02 **	3.88E-02	4.50E-02	1.17E-01 **	9.58E-03	4.37E-02
Osteopporosis by T-score	-1.93E-01 ****	-2.51E-01 ****	-1.49E-01 ****	-2.45E-01 ****	-8.63E-02 ****	-1.53E-01 ****
Any Osteoporosis Drugs	-1.86E-01 **	-1.53E-01	-1.31E-01	-3.05E-01 **	-4.84E-02	-1.01E-01
Calcium Supplement	-1.74E-02	2.61E-03	-2.24E-02	-5.72E-02	-1.65E-02	-1.97E-02
Vitamin D Supplement	-2.07E-02	-3.11E-02	5.72E-03	-1.61E-02	-1.85E-03	-9.94E-03
Oral Estrogen	-1.70E-01	-1.39E-01	-2.83E-01 **	-2.63E-01	-1.59E-01 **	-2.27E-01

4.3D African-American Female

GROUP	African-American Female					
BMD Traits	Femoral Neck	Hip	Whole Body	Lumbar Spine	Mean Arm	Mean leg
Covariates						
Population Substructure						
PC1	6.26E-01 ****	6.08E-01 ****	3.88E-01 **	5.67E-01 **	2.46E-01 ***	2.98E-01
PC2	2.59E-02	4.07E-02	1.33E-01	1.48E-02	1.77E-01 **	1.57E-01
PC3	3.00E-02	-1.15E-01	-9.56E-02	-2.39E-01	-1.41E-02	-1.14E-01
Demographics						
Age	-1.30E-03	-2.80E-03	-7.00E-04	3.90E-03	-1.90E-03 **	-2.60E-03
Site	-1.16E-02	2.48E-02 **	2.20E-02 **	6.01E-02 ****	9.90E-03	4.21E-02 ****
Anthropometrics						
Height	-1.00E-04	-1.00E-04	1.00E-04	3.00E-04 **	2.00E-04 ****	1.00E-04
Weight	3.30E-03 ****	4.20E-03 ****	2.40E-03 ****	4.00E-03 ****	1.80E-03 ****	2.70E-03 ****
Waist Circumference	3.00E-04	-1.00E-04	-7.00E-04	-3.00E-04	-6.00E-04	-2.00E-04
Life style						
Drinking History	1.43E-02 **	1.34E-02	1.12E-02	1.59E-02	6.80E-03	1.43E-02 **
Smoking History	2.60E-03	-2.00E-04	-4.60E-03	-4.20E-03	-3.30E-03	-2.20E-03
Exercise	4.18E-05	1.21E-04	9.73E-05	3.91E-05	1.1E-04 ***	1.46E-04 **
Medical Conditions						
Prevalent CVD	-1.67E-02	-3.63E-02 **	-1.50E-02	-3.99E-02	-6.20E-03	-2.10E-02
Prevalent Diabetes	2.51E-02 **	4.14E-02 ***	3.62E-02 ***	4.76E-02 **	2.11E-02 ***	4.36E-02 ****
Prevalent Osteoarthritis, hip	2.55E-02 **	3.25E-02 ***	1.35E-02	3.82E-02 **	1.90E-02 ***	2.85E-02 **
Prevalent HBP	2.60E-03	9.00E-04	1.70E-03	5.50E-03	4.40E-03	-1.50E-03
Osteopporosis by T-score	-1.36E-01 ****	-1.92E-01 ****	-1.22E-01 ****	-1.71E-01 ****	-6.01E-02 ****	-1.50E-01 ****
Calcium Supplement	-7.90E-03	-1.40E-03	1.78E-02	8.60E-03	5.90E-03	1.22E-02
Vitamin D Supplement	2.98E-02	3.83E-02	5.03E-02 **	6.13E-02	1.11E-02	4.43E-02

Table 4.4: Variation explained (%) by different covariate groups for BMD traits in four ancestry by sex groups

4.4A Caucasian American Male

BMD Trait	Femoral neck	Whole Body	Lumbar Spine	Hip	Mean Arm	Mean Leg
Covariate Group						
allcovariates	18.9	16.7	10.3	26	18.2	22.9
ancestry	0.2	0.1	0.1	0.1	0.2	0.1
demographics	0.1	1.4	1.1	1.9	1	5.3
anthropometrics	13.3	8.7	5.2	15.9	10.5	9.7
lifestyle	0.4	0.6	0.3	0.3	0.5	0.5
medical	5	5.9	3.6	7.7	6	7.2

4.4B Caucasian American Female

BMD Trait	Femoral neck	Whole Body	Lumbar Spine	Hip	Mean Arm	Mean Leg
Covariate Group						
allcovariates	43.9	36.1	26.5	53.1	33.7	40
ancestry	0.1	0.2	0.2	0.1	0.3	0.1
demographics	0.3	0.1	1.4	0.1	0.2	1
anthropometrics	6.4	3.5	2.9	6.4	4.8	5.3
lifestyle	0.7	0.4	0.4	0.3	0.3	0.5
medical	37.2	32.4	22	46.6	28.5	33.6

4.4C African-American Male

BMD Trait	Femoral neck	Whole Body	Lumbar Spine	Hip	Mean Arm	Mean Leg
Covariate Group						
allcovariates	30.7	25.9	22.5	39.3	23.7	29.2
ancestry	1	0.5	0.6	0.3	0.2	0.4
demographics	0.1	3	2.1	1.5	1.5	6.1
anthropometrics	10.7	7.2	6.7	13.1	9.4	6.5
lifestyle	0.4	1.6	0.5	0.7	1	2.5
medical	18.5	13.6	12.6	23.6	11.6	13.6

4.4D African-American Female

BMD Trait	Femoral neck	Whole Body	Lumbar Spine	Hip	Mean Arm	Mean Leg
Covariate Group						
allcovariates	35.7	27.4	25.7	43.7	34.6	36.7
ancestry	1.4	0.9	0.5	1.1	1.3	0.5
demographics	0.3	0.9	2.5	0.6	0.8	2.5
anthropometrics	15.8	6.9	9.4	16.7	14.5	11
lifestyle	1.2	0.6	0.4	0.7	2	1
medical	17	18.2	12.9	24.6	16	21.8

Table 4.5: Longitudinal Analysis of Covariates with BMD traits

****** = p-value < 0.001; ***= 0.001 < p-value < 0.01; **= 0.01< p-value <0.05.**

4.5A Caucasian American Male

GROUP	Caucasian Male					
BMD Traits	Femoral Neck	Hip	Whole Body	Lumbar Spine	Mean Arm	Mean leg
Covariates						
Population Substructure						
PC1	3.90E-02	2.33E-03	-4.48E-02	2.37E-01	9.77E-02	-3.46E-02
PC2	1.74E-01	1.46E-01	4.36E-02	-1.20E-01	1.33E-01	1.64E-01
PC3	-1.25E-01	3.83E-02	-9.80E-03	1.70E-01	1.22E-02	4.17E-02
Demographics						
Site	1.14E-02	3.93E-02 ****	1.15E-02	2.87E-02	-6.04E-03	4.87E-02 ****
Age	-2.84E-03 **	-3.22E-03 **	-3.25E-03 **	1.91E-03	-2.70E-03 ***	-5.03E-03 ****
Anthropometrics						
Height	1.51E-04 **	6.72E-05	3.27E-04 ****	4.77E-04 ****	1.78E-04 ****	4.01E-04 ****
Weight	2.49E-03 ****	3.47E-03 ****	4.15E-04	1.06E-02	-5.90E-03	5.17E-03
Waist Circumference	2.43E-04	1.93E-04	-1.61E-04	1.46E-03 **	-8.04E-04 **	1.23E-04
Time	-7.79E-03 ****	-4.71E-03 ****	-1.59E-02 ****	2.66E-02 ****	-7.97E-03 ****	-2.90E-02 ****
Life style						
Education	5.03E-03	2.67E-03	7.26E-03	1.37E-05	1.17E-05	2.56E-06
Exercise	7.75E-05	6.15E-05	2.26E-05	-4.06E-04 **	-1.88E-04 **	-4.71E-04 ****
Smoking History	-3.47E-04 ***	-4.65E-04 ****	-3.63E-04 ****	9.34E-04	1.14E-03 ****	5.28E-04
Medical Conditions						
Prevalent CHD	1.77E-02 **	1.13E-02	1.16E-02	7.55E-03	2.61E-03	1.29E-02
Prevalent Diabetes	1.69E-02	2.20E-02	6.32E-03	4.89E-02 **	-3.75E-04	1.78E-02
Prevalent Osteoarthritis, knee	1.80E-02	1.26E-02	9.81E-03	1.61E-02	5.92E-03	8.67E-03
Prevalent Osteoarthritis, hip	-1.28E-02	-1.11E-02	5.51E-03	1.28E-02	8.96E-03	8.17E-05
Prevalent HBP	-1.26E-03	2.68E-03	2.43E-03	2.49E-02 **	2.37E-03	-3.14E-03
Osteopporosis by T-score	-1.57E-01 ****	-2.45E-01 ****	-1.59E-01 ****	-1.96E-01 ****	-7.20E-02	-2.16E-01 ****
Calcium Supplement	-2.52E-03	-9.51E-04	1.14E-03	-2.97E-03	2.79E-03	-2.64E-03 *

4.5B Caucasian American Female

GROUP	Caucasian Female					
BMD Traits	Femoral Neck	Hip	Whole Body	Lumbar Spine	Mean Arm	Mean leg
Covariates						
Population Substructure						
PC1	1.01E-02	1.03E-01	-9.27E-02	2.30E-01	-1.79E-02	-1.21E-02
PC2	6.37E-03	-5.50E-02	-6.32E-02	4.36E-02	-1.33E-01	-1.87E-02
PC3	-6.37E-02	-7.83E-02	-5.55E-02	-1.80E-02	7.84E-02	-9.00E-02
Demographics						
Site	-8.93E-03	9.29E-03	-8.19E-03	2.07E-02	-9.52E-03 **	9.68E-03
Age	-4.82E-04	-1.31E-03	-2.75E-05	5.15E-03 **	4.74E-04	-1.62E-03
Anthropometrics						
Height	7.87E-05	4.32E-05	2.93E-04 ****	3.33E-04 ****	1.50E-04 ****	2.61E-04 ****
Weight	2.55E-03 ****	2.98E-03 ****	4.02E-04	1.24E-03 **	9.15E-04 ***	1.27E-03 ****
Waist Circumference	-3.40E-04	-1.59E-04	-2.42E-04	4.31E-04	-2.96E-04	-9.01E-05
Time	-7.12E-04	1.01E-03	-7.51E-03 ****	2.68E-02 ****	-2.53E-03	-1.56E-02 ****
Life style						
Drinking History	-2.88E-05	4.10E-04 **	2.64E-04	2.02E-04	8.55E-05	4.96E-04 ***
Smoking History	7.96E-04	-3.06E-03	-4.81E-04	8.16E-03	8.28E-04	-5.20E-04
Exercise	6.55E-03 **	1.04E-02 ***	5.50E-03	-1.50E-03	4.91E-03	9.51E-03 ***
Medical Conditions						
Prevalent Diabetes	1.27E-02	3.15E-02 **	2.51E-02 **	4.23E-02	1.12E-02	2.28E-02 **
Prevalent Osteoarthritis, knee	1.50E-02 **	1.47E-02 **	1.20E-02	4.50E-02 ****	8.73E-03	1.66E-02 **
Prevalent Osteoarthritis, hip	-8.42E-03	-4.17E-03	-4.26E-03	2.86E-03	-6.22E-03	-8.12E-03
Prevalent Cancer	2.01E-02 **	1.67E-02	1.47E-02	1.79E-02	-2.47E-03	1.11E-02
Prevalent Clinical Osteoporosis	-1.07E-02 ***	-8.61E-03 **	-7.06E-03	-5.53E-03	-1.79E-03	-5.76E-03
Osteopporosis by T-score	-1.32E-01 ****	-1.82E-01 ****	-1.11E-01 ****	-1.51E-01 ****	-5.02E-02 ****	-1.19E-01 ****
Calcium Supplement	-1.04E-03	1.70E-03	-2.47E-03	-4.00E-03	-6.55E-03	-1.20E-03
Vitamin D Supplement	3.04E-03	5.25E-04	-3.14E-03	5.95E-03	-2.45E-03	-2.06E-03
Oral Estrogen	1.46E-02 ****	1.38E-02 ****	3.29E-02 ****	3.42E-02 ****	3.24E-02 ****	2.95E-02 ****

4.5C African-American Male

GROUP	African-American Male					
BMD Traits	Femoral Neck	Hip	Whole Body	Lumbar Spine	Mean Arm	Mean leg
Covariates						
Population Substructure						
PC1	2.24E-01	8.33E-02	-3.02E-02	4.24E-02	1.45E-01	-2.80E-02
PC2	5.45E-01 ***	3.38E-01	4.06E-01 **	7.53E-01 **	1.97E-01	3.62E-01 **
PC3	-3.56E-01	-2.28E-01	8.86E-02	5.13E-01	-6.61E-02	5.02E-02
Demographics						
Site	1.44E-02	4.18E-02 ****	3.38E-02 ***	4.02E-02	1.98E-02 **	5.73E-02 ****
Age	4.89E-04	1.22E-03	1.58E-03	6.52E-03	-9.03E-04	2.29E-03
Anthropometrics						
Height	-3.82E-05	-1.12E-04	1.29E-04	6.76E-05	9.55E-05	1.31E-04
Weight	2.72E-03 ****	2.84E-03 ****	1.59E-03 ***	3.99E-03 ****	1.61E-03 **	1.78E-03 ****
Waist Circumference	4.29E-04	1.37E-03	-7.56E-04	-1.68E-03	-7.01E-04	-5.76E-04
Time	-8.65E-03 ****	-8.89E-03 ****	-1.59E-02 ****	2.98E-02 ****	-6.89E-03 ***	-2.78E-02 ****
Life style						
Drinking History	7.36E-05	7.32E-05	6.12E-04 ***	9.18E-04 **	2.21E-04	5.59E-04 **
Smoking History	-7.06E-03	-9.91E-03	-9.23E-03	-1.26E-02	4.07E-06	-8.97E-03
Exercise	1.12E-03	5.70E-03	1.46E-03	2.27E-02 **	2.52E-03	2.49E-03
Medical Conditions						
Prevalent Diabetes	2.58E-02	2.56E-02	1.32E-02	5.16E-02 **	3.09E-03	8.24E-03
Prevalent Osteoarthritis, knee	8.73E-03	5.26E-03	5.84E-03	4.24E-02	7.83E-05	3.06E-03
Prevalent Osteoarthritis, hip	-4.24E-02 **	-3.52E-02	-2.50E-02	-3.18E-02	-1.96E-02 **	-1.85E-02
Prevalent Cancer	-2.29E-02	-3.02E-02	-3.84E-02 **	-6.26E-02 **	-2.28E-02	-4.32E-02 ***
Prevalent Clinical Osteoporosis	1.06E-02	-1.32E-03	4.81E-04	1.62E-02	-6.54E-03	6.23E-03
Osteopporosis by T-score	-2.03E-01 ****	-2.67E-01 ****	-1.64E-01 ****	-2.45E-01 ****	-9.93E-02 ****	-1.73E-01 ****
Calcium Supplement	-1.35E-02	-9.54E-03	-1.40E-02	-2.50E-02	-2.78E-02 ***	-6.88E-03
Vitamin D Supplement	-2.12E-02 **	-7.96E-03	-6.41E-04	4.27E-03	4.03E-03	3.98E-04
Oral Estrogen	-5.42E-02 ****	3.08E-04	-3.76E-02 ****	1.32E-02	-3.71E-02 ****	-4.27E-02 ****

4.5D African-American Female

GROUP	African-American Female					
BMD Traits	Femoral Neck	Hip	Whole Body	Lumbar Spine	Mean Arm	Mean leg
Covariates						
Population Substructure						
PC1	6.34E-01 ****	5.87E-01 ****	3.66E-01 **	5.82E-01 **	2.45E-01 ***	2.78E-01
PC2	2.20E-02	1.60E-02	7.69E-02	7.05E-02	1.37E-01	9.87E-02
PC3	-3.00E-02	-1.22E-01	-8.81E-03	-1.86E-01	-2.82E-03	-8.10E-02
Demographics						
Site	-4.72E-03	3.57E-02 ****	3.27E-02 ****	6.64E-02 ****	9.98E-03	3.74E-02 ****
Age	-8.56E-04	-3.19E-03 **	-1.05E-03	2.99E-03	-1.61E-03	-2.56E-03
Anthropometrics						
Height	-8.86E-06	-1.12E-04	1.82E-04 ***	2.90E-04 ***	2.17E-04 ****	1.57E-04 **
Weight	2.22E-03 ****	2.81E-03 ****	-1.37E-04	2.00E-03 ***	8.21E-04 ***	1.35E-03 ****
Waist Circumference	1.35E-03 ****	1.45E-03 ****	1.59E-03 ***	1.88E-03 ***	4.37E-04	1.12E-03 ***
Time	-4.90E-03 ****	-2.56E-04	-1.63E-02 ****	2.31E-02 ****	-7.90E-03 ****	-2.52E-02 ****
Life style						
Drinking History	3.29E-04	3.48E-04	3.36E-04	5.94E-04	-1.22E-04	-5.78E-05
Smoking History	-1.54E-03	-2.97E-03	-5.37E-03	-2.91E-03	-1.13E-03	-1.91E-03
Exercise	1.53E-02 **	1.31E-02 **	1.15E-02 **	1.61E-02	5.68E-03	1.47E-02 ***
Medical Conditions						
Prevalent Diabetes	1.83E-02	2.80E-02 **	2.24E-02 **	3.24E-02	1.67E-02 **	3.35E-02 ***
Prevalent Osteoarthritis, knee	-1.11E-02	-9.14E-03	4.49E-03	1.90E-03	1.91E-03	1.05E-03
Prevalent Osteoarthritis, hip	2.15E-02	2.68E-02	1.21E-02	3.63E-02 **	1.31E-02 **	2.05E-02
Prevalent Cancer	-1.15E-02	-1.21E-02	-2.13E-02	-2.15E-02	-1.18E-02	-2.29E-02
Prevalent Clinical Osteoporosis	-1.03E-02	-6.01E-03	-3.68E-03	-2.24E-02	-5.57E-03	-1.59E-03
Osteopporosis by T-score	-1.42E-01 ****	-1.97E-01 ****	-1.41E-01 ****	-1.80E-01 ****	-7.12E-02 ****	-1.64E-01 ****
Calcium Supplement	7.49E-03 **	1.24E-04	1.09E-03	9.13E-03	-1.40E-04	2.55E-03
Vitamin D Supplement	-7.00E-03	4.38E-05	6.73E-04	4.92E-03	-7.75E-04	-1.12E-04
Oral Estrogen	4.45E-03	1.21E-02 ***	8.15E-03	1.61E-02	1.91E-02 ****	1.74E-02 ****

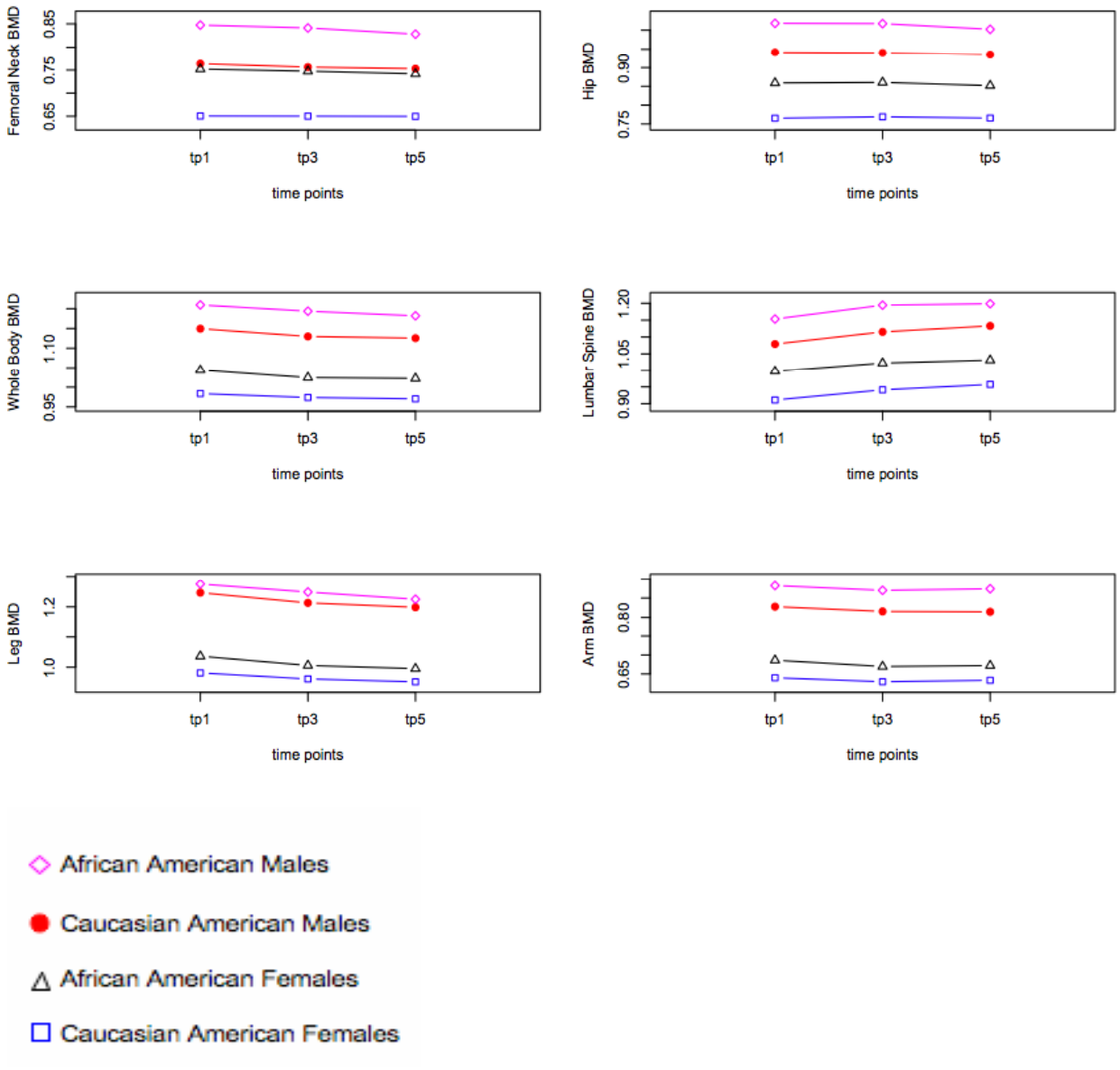


Figure 4.1: Mean BMD across six BMD sites for three time points in all four sex by ancestry group

5.0 ANALYSES OF POSSIBLE ASSOCIATIONS BETWEEN CANDIDATE GENES POLYMORPHISMS AND BONE MINERAL DENSITY TRAITS: SINGLE SNP-SINGLE TRAIT VERSUS SINGLE SNP -MULTIPLE TRAIT COMPARISONS.

5.1 INTRODUCTION

As described in the Introduction to this dissertation (Chapter 1), the goal of the ancillary study for the Health Aging and Body Composition Study (R. Ferrell, P.I.) is to identify genes that influence bone, muscle and fat traits among healthy aged individual (above 68 years at baseline) from African American and Caucasian American groups. As part of my dissertation research, I am investigating whether single nucleotide polymorphisms in specific candidate genes influence bone mineral density (BMD) traits measured at different skeletal sites at three different time-points. BMD is one of the major risk factors for osteoporosis (Jordan et al. 2002). In particular, I wanted to determine whether polymorphisms at these candidate genes influence the genetic architecture of BMD at different skeletal sites in men and women, as well as in different ancestry groups.

The genetic architecture of BMD traits is likely to differ significantly between men and women and between different ancestral groups. As is well-known, women generally tend to lose bone mass more rapidly after menopause than do men. Furthermore, average areal BMD in tibia differs among men and women as well as between age groups above and below 60 years (Khodadadyan-Klostermann et al. 2004). In the HABC study population, both Caucasian American and African American men have higher volumetric BMD than their women counterparts (Taaffe et al. 2003) . In addition to BMD differences between sexes, there are differences in BMD between different ancestry groups. For example, on average,

older African-American women have higher levels of areal BMD at all bone sites than do Caucasian American women (Evans et al. 2005), and lose BMD less rapidly with increasing age (Cauley et al. 2005). This phenomenon leads to higher risk of fracture among Caucasian American women. These observations support the hypothesis that the genetic architecture of BMD differs by sex and ancestry, so we performed our genetic analyses on the four sex by ancestry groups separately.

Although association tests are a powerful way to detect relatively small effects of SNPs on traits (Risch et al. 1996), one of the concerns about association studies is multiple testing (Yongchao et al. 2003). Association analyses done with a single SNP and single phenotype at a time requires a lot of testing. For example, in our study with 15 traits and ~1400 SNPs, we are performing 21000 tests, thus, by chance, at $p = 0.05$, we would expect 1050 false positive results. Several approaches have been suggested to mitigate the concerns about multiple-testing, like Bonferroni (Simes 1986), Sidak (Sidak 1967) , and false discovery rate (Benjamini et al. 1995). The most conservative of these approaches is the Bonferroni method and if we adjusted for multiple testing in our study, we would need a p-value = 2.5E-6. Although the Bonferroni adjustment will reduce the detection of false positives, we may also miss many true positives (Yongchao et al. 2003). Another approach is the false discovery rate (FDR). In this approach the number of false discoveries are controlled, that is the expected proportions of type I errors among the rejections is controlled. It has more power than the Bonferroni method (Yang et al. 2005). The Bonferroni and FDR approaches assume that the tests are independent, however, in our candidate gene study, neither the SNPs nor the BMD traits are independent. Permutation tests, if performed under the null hypothesis in such a way that the correlation structure is preserved can serve as a robust method for identifying the probability of getting a significant p-value. However, this method can be time-consuming. Conneely and Boehnke (Conneely et al. 2007) have proposed a method that is faster than permutation tests and that uses numerical integration to compare the test statistic with their empirical distribution. Still another method by which to reduce the number of tests is to use principal components or factor analysis methods (Nyholt 2004). Although these latter methods offer a way to do dimension reduction testing, they are not based on genetics. Recently, Klei and colleagues (2008) developed a new methodology called

principal components of heritability (PCH). In this methodology, the principal components are devised to maximize the heritability of a set of traits with respect to individual SNP (Klei et al. 2008). In other words, the PCH analyses are a single SNP-multiple trait analysis. In this chapter, we compared association test results obtained using the single SNP-single trait analyses with the single SNP-multiple trait (PCH) analyses.

However, the strongest evidence that a trait is influenced by a particular polymorphism (or one in strong LD with it), is to replicate the original genotype-phenotype association in an independent population (Chanock et al. 2007). This report gives several guidelines to assess and report possible genotype-phenotype associations. These guidelines include: (1) sample size is sufficient in the replicate population to detect the proposed effect, (2) the same phenotype in a similar, independent population should be analyzed, (3) similar magnitude of the effect should be demonstrated, in the same direction and with the same SNP as in the initial study (4) same genetic model should be tested as in the initial study, (5) a strong rationale should be provided in selecting the SNPs for replication.

In our HABC study, we can assess possible replication at several levels. First, if we observe similar results across our four sex by ancestry groups, this result would support a particular genotype-phenotype association. However, because the genetic architecture of each sex by ancestry group may differ, the absence of a consistent result across any two (or more) of the four groups does not indicate a particular association is false. Second, we have three longitudinal measures of BMD obtained two years apart (time points 1, 3, and 5). Thus, a consistent result across all three time points, or 2 of three consecutive time points, would also support the hypothesis that a particular genotype-phenotype result is true. We recognize that a true test of replication would be to look for the same genotype-phenotype association result in a similar, but independent, population. Thus, we also followed up our potentially significant association results using BMD data from an additional population of Caucasian American men, the MrOS study (J. Zmuda, PI of the ancillary study on the genetics of BMD in MrOS). Details of this population are given in the Methods section below.

In this chapter, I present the results of analyses to determine whether genetic variation in 138 candidate genes (1439 total SNPs that passed quality control procedures, see Chapter 3) was associated with BMD in each of four sex by ancestry groups (Caucasian American men, Caucasian American women, African American men and African American women). I also compared the results from two different methods: single SNP/single trait analyses versus single SNP-multiple trait (PCH). In addition, I performed additional analyses on BMD data collected at three different time points to assess whether associations observed at baseline were replicated two and four years later. Finally, I performed replication studies of potentially significant genotype-phenotype relationships in another population of older Caucasian American men.

The inclusion of genetically correlated traits should increase the power of the PCH method to detect a genotype-phenotype association, especially if part of the overall genetic correlation among traits is due to the specific SNP being analyzed (Klei et al. 2008). Therefore, in addition to the 6 BMD traits assessed in the previous chapter, information on an additional 9 skeletal sites is included in the single SNP-multiple trait analysis. However, the Bonferroni significance of the individual single SNP/single trait associations is assessed in two ways – for all 15 correlated BMD traits, as well as for the subset of 6 commonly analyzed BMD traits.

5.2 DATA AND ANALYTICAL METHODS

5.2.1 Health ABC populations

Details of the genotype and phenotype data available on the four sex by ancestry groups were described in detail in Chapters 3 and 4. Briefly, we initially had genotype data on 1742 Caucasian Americans and 1237 African Americans. We did all our genetic association analysis based on the Illumina SNPs only (chapter 5 and chapter 6). The genetic data comprised a total of 1536 SNPs from 131 candidate genes and

gene clusters (the gene clusters comprised multiple genes). We carried out data cleaning across SNPs and individuals. SNPs with call rate of less than 90% were discarded, as were all individuals with a call rate of less than 93%. Each SNP that passed quality control and that was originally denoted as belonging to a specific gene cluster was reassigned to the gene that was closest to that SNP (in terms of base-pair distance). After data cleaning we had complete genotype and phenotype data on 1653 Caucasian Americans (872 men and 781 women) and 1177 African Americans (673 women and 504 men). We had data for 1439 SNPs in 138 genes in these individuals. Both of our analysis programs uses a cut off of $p< 0.001$ for Hardy Weinberg Equilibrium within the four sex by ancestry groups. SNPs that have a p-value lower than 0.001 are discarded from the analyses.

As described in the previous chapter (Chapter 4), I identified a set of anthropometric, demographic, lifestyle, and medical history covariates that influenced each of 15 bone mineral density traits at each of three time-points in each of the four ancestry by sex population groups separately. The BMD residuals obtained from each of these 180 (2 sexes by 2 ancestry groups by 15 skeletal sites by 3 time points) models were used in the single-SNP/single trait and PCH analyses described below.

5.2.2 MrOS population

The Osteoporotic Fractures in Men (MrOS) study is an NIH funded study of risk factors for osteoporosis in men (Orwoll et al. 2005). To participate in the study, men needed to be ≥ 65 years old, able to walk without assistance of another person, could not have had a bilateral hip replacement preventing an assessment of hip BMD, and could not have had a medical condition that, in the judgment of the investigator, would make it unlikely that they would have survived the duration of the study. Similar to the HABC study, at the baseline examination, information was collected on BMD (measured by DXA), medical history, lifestyle and demographic characteristics. As part of an ancillary study entitled "Bone Strength Phenotypes in Men: Genes and Environment", J. Zmuda, PI, genotype data were available on 871 men for the same 1439 SNPs in 138 genes that were assayed in HABC. These data were used to

replicate potential genotype-phenotype associations in Caucasian American-American men group from HABC. The BMD traits from MrOS were also adjusted for significant covariates from the same set of covariates as in HABC cohort.

5.2.3 Single SNP-Single Trait Analyses

For each of the traits a linear regression model was fitted for each SNP, and then we tested the null hypothesis that the slope of the model was equal to zero using an F-test. For a given trait and a given SNP the model is

$$y = \alpha + \beta X + \varepsilon,$$

where y is the quantitative trait (e.g. hip BMD) and X is the genotype of the SNP (coded as 0,1,2, and representing the number of minor alleles in the genotype), and ε is the residual error. So we fitted an additive genetic model with 1 df to test the null hypothesis that the SNP is not associated with the trait. If we have n observation points then the F-statistic is given by mean square of the model divided by the mean square error (i.e., MSM/MSE).

MSM = $\sum(\hat{y}_i - \bar{y})^2 / (1) = SSM / DFM$, DFM = degrees of freedom due to the model,

MSE = $\sum(y_i - \bar{y})^2 / (n-2) = SSE / DFE$. DFE = degrees of freedom due to the error.

5.2.4 Single SNP-Multiple Traits Analyses: Principal Component of Heritability, PCH

We next analyzed the data using the principal components of heritability (PCH) method. Composite measures of the correlated traits were obtained using principal components analyses. However, instead of using the phenotypic correlation among the traits, the principal component was derived by maximizing the heritability of each trait due to the single SNP of interest. This heritability due to a single SNP is a fraction of the total heritability of that trait.

The regression model for testing association of a quantitative trait locus (QTL) or SNP on a set of phenotypes is given by,

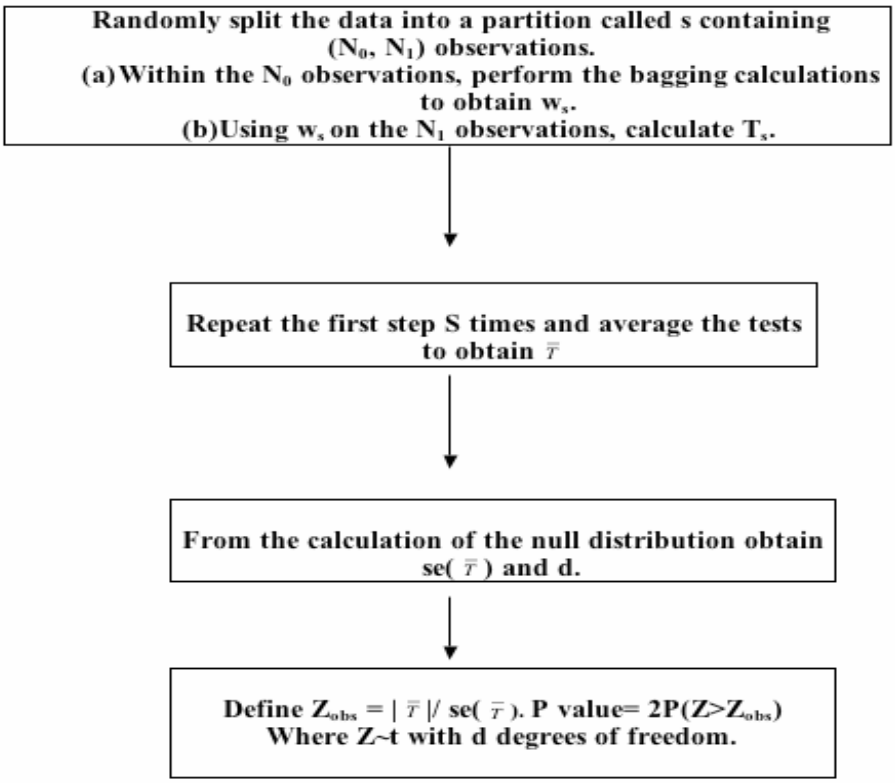
$$w'y = \alpha + \beta x + \varepsilon ,$$

where y is a set of phenotypes, w is the corresponding loading, x is the SNP, ε is the error term. The PCH analysis chooses w such that the SNP-specific heritability of the resulting combination of traits is maximized for the SNP. Here the SNP-specific heritability is a fraction of the total heritability of the trait and is due to the single QTL (or SNP) only. For single SNP-single trait analysis w = (0,0,…,1,0,…0). [In other words, for a single-SNP, single trait, w for the specific trait = 1, whereas for all others it is 0.] For testing the association between the composite trait and the specific SNP, (H_0: $\beta = 0$ vs H_1: $\beta \neq 0$) the test statistic is a t-distribution, T= β /se(β).

The main drawback of this procedure is that the loadings for each SNP have to be estimated and the best way to obtain the estimates is from the data being analyzed for associations. To avoid the inflation of type I error due to estimation and association testing on same data, the PCH procedure follows the algorithm in the flowchart given below. First, the sample is split into two parts (N_0,N_1), as is done in a cross –validation procedure (Hastie et al. 2001). Next, using the first partition of the split sample, N_0 (which is the testing set), the loadings (w_s) are estimated by a bagging (Bishop 2006) technique. Bagging

is a technique which minimizes bias of nonlinear estimates using bootstrap samples (Klei et al. 2008). After the w_s are estimated, they are used on data in the testing set N_1 to calculate T. For each SNP this step is repeated for S times and the average of the T statistics are taken to get $\bar{T}$. From the distribution of $\bar{T}$ the se($\bar{T}$) and d (unspecified degrees of freedom for the t-distribution of $\bar{T}$ using method of moments estimator), is calculated. A variable Z is defined as

$z_{obs}=\bar{T}$ /se($\bar{T}$), and the p-value [p(PCH)] corresponding to a SNP is calculated as $2P(Z>z_{obs})$, Z~t with d degrees of freedom.



Flowchart for PCH procedure

We performed PCH analysis on our traits at each of the three time points separately. Thus, unlike many replication studies in which only the most significant results are analyzed in the replication sets, we analyzed each "replication" set as if it were the original set. This approach is more conservative than that

performed in most association studies. Next, to identify potentially interesting SNPs, we compared the results across each of the time points and within and between each ancestry by sex groups. Because we were able to compare results across multiple population groups and multiple time points, I chose to use a liberal false discovery rate [FDR] = 50%.

5.2.5 Regression analyses

After identifying each potentially interesting SNP, we then plotted the residual values for each BMD trait by SNP genotype and used regression analyses to assess the effect of the SNP genotypes on each trait. We considered a p-value = 0.05 to be significant. To determine whether the effect (magnitude and direction) of a particular genotype was similar on a specific trait across time points, I tested for equality of slopes.

For each potentially interesting SNP the model was

$$y = \alpha + \beta_{tpi}X + \varepsilon,$$

where y is a BMD residual trait. For each significant SNP and BMD trait that was significant at baseline, we fitted this model for all the three time points separately. We next tested the equality of each pair of the regression slopes (for the three time points) using p-value ≤0.05.

H_0: $\beta_{tpi} = \beta_{tpj}$, i,j = 1,3,5;i ≠ j , H_1: The slopes are not equal.

Under H_0 $\dfrac{\hat{\beta}_{tpi} - \hat{\beta}_{tpj}}{\sqrt{(\text{var}(\hat{\beta}_{tpi})/n_i) + (\text{var}(\hat{\beta}_{tpj})/n_j)}}$, follows standard normal. So we tested equality of slopes by comparing it with standard normal probability distribution. Here n_i and n_j are the sample sizes of i^{th} and j^{th} time point.

5.2.6 Haplotype Analyses

We detected an association between multiple SNPs within the GNRHR locus and thus we did further investigations to determine whether (1) each of these SNPs contributed independently to the effect, or (2) the effect of one SNP predominated or (3) the SNPs were marking a specific haplotype. We first calculated the LD structure within the GNRHR locus using all of the genotyped SNPs. Based on the LD structure, we then did haplotype association analysis on SNPs from our most potentially interesting gene, GNRHR, using haplo.stats package in R. The haplotype analysis program first estimates the haplotypes among unrelated people using Expectation Maximization (EM) algorithm and then applies a score test to test for association between the haplotypes and the quantitative trait.

5.3 RESULTS

5.3.1 Single SNP-Single Trait analyses

As can be seen in Figure 5.1 there are many SNP-trait combinations which have - $\log_{10}$(pvalue) > 2 (or p<0.01). However, we have done multiple tests, ~1350 tests per trait for each of the 6 BMD traits (within each sex by ancestry group). Using the conservative Bonferroni adjustment, the -$\log_{10}$(pvalue) for a significant result considering just the 6 traits will be $\geq$ 5.21 (and if we include all 15 BMD traits, the significant p-value should be $\geq$ 5.61). But these levels are very conservative (as described in methods section), especially because the BMD traits are not independent and neither are all of the SNPs, Thus, we are not doing 8100 independent tests across 6 traits (or 20250 tests across 15 traits) on which the Bonferroni critical values are based. We calculated FDR p-values for our tests using Benjamini and Hochberg (Benjamini et al. 1995) approach and used a significance level of 25%, and looked for consistency across time points, BMD skeletal sites, and within candidate genes. Results of single-SNP-

single trait associations for each sex by ancestry group for the 6 skeletal sites at baseline, time point 3, and time point 5 are presented in Figures 5.1, 5.2, and 5.3 respectively. A summary of the most significant results is given in Table 5.1.

In Caucasian American males, a SNP (#rs3828562) within the GNRHR locus on chromosome 4 is associated with hip BMD at baseline (Figure 5.1A) at $-\log_{10}$ pvalue = 4.12. This same SNP is associated with hip BMD at time points 3 and 5 also (Figure 5.2A and Figure 5.3A), although the significance level is not as high ($-\log_{10}$ p = 3.45 and 3.03 respectively), and the result for time point 5 does not meet the FDR cut off value. Two SNPs (#rs344357 and #rs1178435) in IGFALS on chromosome 16 are associated with whole body BMD at baseline ($-\log_{10}$ p = 4.2 and 4.42 respectively).

For Caucasian American females, one SNP (#rs3102724) in the TNFRSF11B locus on chromosome 8 is highly significantly associated with lumbar spine BMD at Bonferroni level at time point 3 (Figure 5.2B) ($-\log_{10}$ p = 5.4) .The same SNP is also associated with lumbar spine at baseline (figure 1B), although the p-value is larger ($-\log_{10}$ p = 4). At time point 5 two SNPs (# rs31476 & #rs25887) from CSF2 gene on chromosome 5 are associated with lumbar spine ($-\log_{10}$ p = 4.28 & 4.02 respectively).

For African-American males two SNPs (#rs2110726 and #rs3917225) in the IL1R1 locus on chromosome 2 are highly significantly associated with arm BMD at time point 5 ($-\log_{10}$ p = 5.82 and 5.31, Figure 5.3C). However, these two SNPs are not significant at any other time points. One SNP (#rs1570070) from the IGF2R locus on chromosome 6 is significantly associated with hip ($-\log_{10}$ p = 4.34) at baseline (Figure 5.1C), as well as with hip BMD at time points 3 and ($-\log_{10}$ p = 2.79) and time point 5 ($-\log_{10}$ p = 3.12), although the results for time points 3 and 5 do not meet the FDR cut off.

Finally, for African-American females, one SNP in the FBXO32 locus on chromosome 8 (#rs2294090 is significantly associated with two traits, arm and whole body BMD, ($-\log_{10}$ p = 5.05 & 5.17, respectively) at time point 3 (Figure 5.2D), but not at any other time points. A SNP (#rs2765880) from TGFBR3 gene on chromosome 1 is associated with lumbar spine at baseline ($-\log_{10}$ p = 4). Another SNP (#rs31473) from CSF2 gene on chromosome 5 is associated with femoral neck BMD ($-\log_{10}$ p =

4.03), but these associations are not found at any of the other time points. At time point 5 a SNP (#rs4648022) from NFKB1 gene on chromosome 4 is associated with arm BMD ($-\log_{10} p = 4.67$).

Although none of the genotype by skeletal site associations were consistent across the four sex by ancestry groups, several of these candidate gene associations are potentially interesting. The association between multiple GNRHR SNPs and hip BMD at two time points in Caucasian American males is especially interesting (Figures 5.1A and 5.2A). Therefore, we further investigated the magnitude and direction of this association, described below after the Single SNP-Multiple Trait Analysis section.

5.3.2 Single SNP-Multiple Trait analyses

As can be seen in Figure 5.2 (A, B, C, D), many SNPs are significantly associated with the composite BMD trait using PCH analysis and assuming an FDR = 50%. Most interestingly the SNPs from GNRHR locus, which were also significant at time points 1 and 3 for the single SNP-single trait analysis in Caucasian American males, were also significant for the composite traits (Figure 5.2A). For African-American males, SNP #rs1570070 from IGF2R was significant at time points 1 and 5 (Figure 5.2D). It was also significant at these two time points, as well as time point 3, for hip BMD for the single SNP-single trait analysis.

In Table 5.2, we list the most significant SNPs (and their genes) that were observed in the results of PCH analyses above. These SNPs were associated with the composite DXA BMD trait at FDR $\geq$ 50%. Across all sex by ancestry groups, there were 21 significant SNPs at time point 1, 20 significant SNPs at time point 3, and 10 significant SNPs at time point 5. Within a time point, we observed no replication of significant SNPs or genes across the sex by ancestry groups. As stated in the introduction, such a result is not unexpected because of known differences in BMD between ancestry groups and between sexes. Likewise, except for Caucasian American males at time points 1 and 3, we did not observe multiple SNPs within the same gene having a significant effect on the composite BMD trait. Within the Caucasian American males at time point 1, 5 (out of 9 total) SNPs within GNRHR, 2 (out of 18 total) SNPs within

TGFB2 and 2 (out of 3 total) SNPs within IGFALS were significantly associated with the composite BMD trait. Most interestingly, these same 5 GNRHR SNPs were also significantly associated with the BMD composite trait at time point 3, but not time point 5. This replication across time points indicates that polymorphisms within GNRHR may affect BMD in older Caucasian American males. Another potentially interesting SNP in Caucasian American males is a SNP in NCOA1 that was significant at time point 1 ($p = 0.002$) and time point 5 ($p = 0.07$), but not at time point 3.

Finally, we also observed replication across time for two SNPs in African American males, a SNP in IGFR2R was significant at time points 1 and 3, and a SNP in NTRK3 was significant at time point 3 ($p = 0.01$) and 5 ($p = 0.01$), but not 1.

Although all of the SNPs that are significantly associated with the composite BMD traits in the sex by ancestry groups are potentially interesting, we followed up on our most significant signals in the GNRHR locus.

5.3.3 Relationship Between Individual GNRHR SNPs on BMD traits in Caucasian American Males

We first plotted the residual BMD values at each skeletal site by genotype for each of the five significant GNRHR SNPs at each of the three time points (Figure 5.5 below).

As can be seen, for #rs972072 (with alleles A/C), the presence of the allele A (minor allele) is associated with decreased BMD at all the sites and this decrease was significant (additive model, nominal $p \leq 0.05$) for hip neck (p-value = 0.003), hip (p-value = 0.007), and whole body BMD (p-value = 0.0452) at baseline (time point 1). At time points 3 and 5 (Figures 5.4 and 5.5, respectively), we observed a similar relationship between this SNP genotype and BMD, although the individual regression analyses were not significant at time point 5. Because the non-significant regression analysis results at time point 5 may be due to a smaller sample size, we also tested for the equality of the three slopes from the three time points. This analysis of equality of slopes was done only for those traits (hip and hip neck) that were

significant at baseline. Table 5.3 gives the result of equality of slope test for all the SNPs in the GNRHR locus. As can be seen, there were no significant differences between slopes at the three time points and the overall estimate of regression was significant. Based on these results, we conclude that the level of BMD decreases as the number of minor allele increases for #rs972072 (Figure 5.3) for BMD of the hip and hip neck, and whole body, but not for BMD traits of the appendicular skeleton.

For the GNRHR SNP #rs974483 (A/G alleles), the presence of the allele G (the minor allele) is associated with an increase the mean level of BMD at all the sites, and is significant for hip (p= 0.0005) and hip neck (p = 0.018). A similar relationship was observed at the year 3 and year 5 time points (Figures 5.4 and 5.5), and overall slopes were significant and equivalent across all three time points (Table 5.3) We checked the equality of regression slope for all the three time points and found that the three slopes were not significantly different from each other. But from inspection of the mean BMD by genotype plots (Figures 5.3, 5.4 and 5.5), the effect of this SNP genotype on BMD does not appear to be additive. So we tested whether the rare allele for #rs974483 had a dominant effect on hip and hip neck BMD. The dominant model was significant for both BMD traits across all the time points; p-value for hip was $2x10^{-5}$, $5x10^{-5}$ and 0.002 for time point1, time point3, and time point5 respectively. The p-values for hip neck were 0.006, 0.0006 and 0.018 for the three time points, respectively. So we conclude from these results that the level of BMD increases in the presence of minor, dominant allele for #rs974483.

For the GNRHR SNP #rs6552115 (A/C alleles), the presence of the allele A (the minor allele) is associated with an increase in the mean level of BMD at all the sites, and is significant for hip (p=0.002) and hip neck (p = 0.002). A similar relationship between hip and hip neck BMD and the SNP genotypes was observed at both year 3 and year 5 time points (Figures 5.4 and 5.5), and this relationship was significant and equivalent across all three time points (Table 5.3). So we can say from our results that the mean BMD of the hip and hip neck increases with the number of minor alleles for #rs6552115.

Similar results were obtained for GNRHR SNPS #rs3796720 and #rs3828562. For each of these SNPs, the presence of the minor allele was associated in an additive increase in the the mean BMD at all

skeletal sites, significant for hip and hip neck, and the slopes were significantly equivalent across all time points for each trait.

Thus, the results of these analyses indicates that five SNPs within the GNRHR locus are associated with increased BMD in older Caucasian American males at all skeletal sites, and are similar across all three time points. However, in general, these five SNPs have the strongest effect on hip and hip neck. For four of these five SNPs, the minor allele is associated with increased BMD, whereas for the remaining SNP, the minor allele is associated with decreased BMD. However, these analyses do not indicate whether a single common haplotype is responsible for the observed effect, or whether each SNP accounts for a small, independent effect on BMD.

5.3.4 Haplotype Analyses of GNRHR SNPs

Because of the consistency associations for all five significant GNRHR SNPs for different skeletal sites across three time points, we next assessed whether these relationships were attributable to a common haplotype or individual effects. As can be seen in Figure 5.6, 8 of the 9 SNPs in GNRHR were in moderate to strong LD (D' $\geq$ 0.80) , and 1 SNP was not in strong LD with any of the other SNPs. The 5 SNPs that were each significantly associated with hip and hip neck BMD are contained within the large block of 8 SNPs.

So we tested for association between hip and hip neck BMD and GNRHR haplotypes analysis based on 8 SNPs and the 5 significant SNPs separately. The analysis with haplotypes generated from all the 8 SNPs does not give any extra input in our understanding of possible association. So we are presenting the analysis from haplotype analysis using the five SNPs only.

Using data on the 5 SNPs that, when analyzed separately, were each significantly associated with hip BMD at baseline, we performed haplotype association analyses with hip BMD. The overall results (Table 5.4) indicated that the 5-SNP haplotypes were significantly associated with hip BMD ($p = 0.035$). As described above,in the single SNP analyses, the minor allele at SNP #rs972072 was associated with a

decreased BMD whereas for the other 4 SNPs, the major allele was associated with decreased BMD. Likewise, in Table 5.4, the haplotype containing the minor allele for rs#972072 and the major alleles for the other SNPs (we call this haplotype L) was also associated with low hip BMD (the Haplotype Score = -3.23, p = 0.001) In contrast, the haplotype with major allele at #rs972072 and minor allele at all other loci was positively associated with hip BMD (we call this haplotype H), (haplotype score = 3.62, p = 0.0003). We also plotted the mean BMD across different dipoltype groups (Figure 5.7A). From the plots we can give the same explanation as above.

Although femoral neck BMD is not significantly associated with overall haplotypes, the association with the H (score = 2.74, p-value = 0.006) and L (score = -2.88, p-value = 0.004) haplotypes were significant and in the same direction as hip BMD. This is also evident from Figure 5.7B.

5.3.5 Analysis of Replicate Sample (MrOS) for Effects of GNRHR and IFGALS loci

As described above, we also had genotype and BMD data on another independent population of 871 Caucasian American males. Because of the apparent strength of our relationship between BMD and SNPs in GNRHR and IGFALS loci, we tried to replicate our results using the MrOS genotype and BMD data. We had data on the same 9 SNPs from GNRHR, in both studies, so we tested for associations between the GNRHR SNPs and hip and hip neck BMD. We also had genotypes on the same 4 SNPs from IGFALS in both studies, so we tested for association between the IGFALS SNPs and arm, leg and whole body BMD. As can be seen in Table 5.5 one of the SNPs from the GNRHR locus (#rs974483) was significantly associated (p = 0.007412) with hip BMD. However, the direction of the association was opposite that observed in HABC Caucasian American men, that is, the minor allele genotype was associated with a decrease in mean hip BMD (slope = - 0.019±0.007). In HABC Caucasian American men, the minor allele was associated with an increase in hip BMD. We also carried out the haplotype tests of five GNRHR SNPs on hip BMD. We found that the L and H haplotype as given in the previous section were associated with the hip trait but the association of the respective haplotypes were in opposite direction when

compared with Caucasian males in HABC data. None of the SNPs from IGFALS was significantly associated with BMD in MrOS. Plot of residual Hip BMD by genotypes in MrOS is given in Figure 5.8.

5.4 DISCUSSION

In this Chapter we performed both single SNP-single trait and single SNP-multiple trait analyses. To date, most of the GWAS studies and candidate gene studies have used the single SNP-single trait analyses, and all have required stringent criteria for significance or have replicated results in multiple populations or both. In our single SNP-single trait analyses of BMD data within the four HABC cohorts, we detected significant associations between BMD and TNFRSF11B in Caucasian American females and IL1R1 in African American males after performing conservative Bonferroni adjustments for multiple testing. The association between lumbar spine BMD and the TNFRSF11B SNP is potentially intriguing because it was observed at time points 1 and 3. However, the association between hip BMD and three SNPS at the GNRHR loci, although not statistically significant at the Bonferroni critical value, is also interesting because it also was observed at two time points at FDR <0.25.

Results of the single SNP-multiple trait analyses revealed several potentially interesting results. Although none of the significant SNPs were similar across the four sex by ancestry groups, three different SNPs within IGF2R were associated with composite BMD in Caucasian American males, and African American males and females. In fact the same IGF2R SNP was associated with composite BMD at two time points in African American males. Furthermore, multiple SNPs in three genes (IGFALS, GNRHR, and TGFB2) were associated with composite BMD in Caucasian American males at baseline. Finally, in addition to the IGR2R SNP, polymorphisms in two other genes GNRHR (in Caucasian American males) and NTRK3 (in African American males) were also associated with composite BMD at two time points.

When we compared the results from the single SNP-single trait analyses to those of the single SNP-multiple trait analyses, we observed some consistent patterns. Multiple SNPs from the GNRHR

locus were significantly associated with hip or composite BMD within Caucasian American males at time points 1 and 3. Similarly, both methods revealed that two SNPs at IGFALS were associated with whole body, leg, or composite BMD in Caucasian American males. Finally, a SNP at the IGF2R locus was associated with composite or hip BMD at time points 1 and 3 in African American males. These results are encouraging and may indicate that these associations are real.

Overall, the results of the single SNP-single trait and single-SNP-multiple trait analyses were somewhat surprising because, across all three time points, we detected more SNPs potentially influencing BMD in males, especially Caucasian American males (n= 22 SNPs total for the single SNP-single trait and multiple trait analyses, Table 5.1 & Table 5.2), than we did in women. In fact, the fewest significant results were obtained in Caucasian American women (n= 7 SNPs, Table 5.1& Table 5.2). Because the sample sizes are similar among the four sex by ancestry groups, this four-fold difference in total number of potentially significant SNPs is unlikely to be due to sample size. Some of this difference is attributable to the observation that 5 SNPs within a single locus influence BMD in Caucasian American males, and this result is consistent across two time points. However, there still are relatively fewer results in women versus men. Another reason for this difference may relate to the covariates included in the current analyses. We included all potentially significant covariates in our model of BMD at each skeletal site and used the residuals from this model in our candidate gene association analyses. Thus, we will not detect any SNPs that have pleiotropic effects on both covariates and the BMD traits. Analyses of the BMD traits using other sets of residuals should also be done.

Because the GNRHR result appeared to be the strongest, we performed follow-up analyses of the SNPs at this locus and their relationship with BMD and different skeletal sites. These analyses revealed that four of the five GNRHR SNPs had a strong additive genetic effect on multiple skeletal sites and the remaining SNP had a strong dominant effect, and these relationships were strongest at the hip BMD sites. Furthermore, the additive or dominant genetic effect of each of the five SNP genotypes was similar across all 3 time points, that is, there were no significant differences among the slopes across the three time points. Additional investigation of GNRHR haplotypes revealed there is significant association between

the haplotypes and hip BMD traits and that all of the SNPs may be marking a single, influential haplotype. The results of these analyses, that is, significant effects of multiple SNPs across multiple skeletal sites and across all three time points, strongly supports the hypothesis that variation in the GNRHR locus (or a locus in strong LD) is associated with variation in BMD in Caucasian American men.

Our finding that the gonadotropin releasing hormone receptor locus (GNRHR) may influence BMD is somewhat unexpected, as there does not appear to be a direct effect of GNRHR on BMD. Several investigators have reported that sex steroids influence bone density and bone metabolism in men. Van den Beld and colleagues (van den Beld et al. 2000) reported that increased testosterone level was associated with increased hip BMD in men aged 73-94 years. This result is consistent with the report that a continuous gradual reduction in the production of sex steroid has a strong negative effect in bone turnover of elderly men (Riggs et al. 2002). Likewise, Khosla et al., (Khosla 2004) report that estrogen and testosterone have roles in maintaining the skeletal structure among males over the age of 60 years. However, we did not see any relationship between polymorphisms in the sex steroids loci and BMD in our study, at least after adjusting for all significant covariates.

The sex steroids are under control of the gonadotropins, such as luteinizing hormone (LH), and follicle stimulating hormone (FSH), which are secreted by the pituitary gland. Karim and colleagues (Karim et al. 2008) reported gonadotrophins and gonadal peptides play a role in bone mass maintenance among Caucasian American males of mean age 57.7 years. The gonadotropin releasing hormones (GNRH1 and GNRH2), as well as the GNRHR regulate the production of gonadotropins, and they theoretically can influence bone mass. We did not see any relationship between SNPs in GNRH1 or GNRH2 on BMD, but we did obtain reasonably strong evidence for an association between BMD in Caucasian American males and genetic variation at the GNRHR locus. Although we did not observe any potentially significant relationships between BMD and other loci in this pathway, our results do not preclude that such relationships exist.

Based on strength and biological plausibility of our result, we tried to replicate our results in another Caucasian American male population, MrOS. Unfortunately, although one of the GNRHR SNPs

was significantly associated with BMD, the direction of the association was not the same. This non-replication of a possibly significant result may be due to several causes. First, there are differences in between the populations. The mean age for MrOS is 65 years, whereas for HABC Caucasian American men, the mean age is 73 years at baseline. Bone density is known to change rapidly among older individuals and this may have influenced our results. Another reason might be population substructure. Although population substructure was minimal among the HABC Caucasian American cohort, we still incorporated a component for substructure in our models. However, we did not have such information available on MrOS and this might have influenced our results. Finally, the result in HABC could be spurious. Further replication studies needed to be done before we can conclusively tell about the role of gonadotropins in bone metabolism**.**

5.5 TABLES AND FIGURES

Table 5.1: Results from single SNP single trait analysis in time points 1, 3 and 5

SNPs coded red in time points 3 and 5 were also significanttime point 1

Time point 1

Group	SNP	GENE	BMD TRAIT	pvalue	(-1)(LOG$_{10}$P)	FDR p-value
Arican American Female	rs2765880	TGFBR3	Lumbar Spine	1.00E-04	4.00	0.14
	rs31473	CSF2	Femoral Neck	9.25E-05	4.03	0.13
African American Male	rs1570070	IGF2R	Hip	4.60E-05	4.34	0.06
	rs1570070	IGF2R	Arm	1.22E-04	3.91	0.12
	rs1570070	IGF2R	Whole Body	1.76E-04	3.75	0.24
Caucasian Female	rs3102724	TNFRSF11B	Lumbar Spine	1.00E-04	4.00	0.18
Caucasian Male	rs3828562	GNRHR	Hip	7.55E-05	4.12	0.10
	rs1178435	IGFALS	Whole Body	3.76E-05	4.42	0.04
	rs344357	IGFALS	Whole Body	6.29E-05	4.20	0.04

Time point 3

Group	SNP	GENE	TRAIT	pvalue	(-1)(LOG$_{10}$P)	FDR p-value
African American Female	rs2294090	FBXO32	Arm	8.96E-06	5.05	0.01
	rs2294090	FBXO32	Whole Body	6.71E-06	5.17	0.01
African American Male	rs1570070	IGF2R	Hip	1.61E-03	2.79	0.80
	rs939348	THRA	Leg	5.54E-05	4.26	0.08
	rs939348	THRA	Whole Body	2.22E-04	3.65	0.30
Caucasian Female	rs3102724	TNFRSF11B	Lumbar Spine	3.96E-06	5.40	0.01
Caucasian Male	rs3828562	GNRHR	Hip	3.53E-04	3.45	0.24

Time point 5

Group	SNP	GENE	TRAIT	pvalue	(-1)(LOG$_{10}$P)	FDR p-value
African American Female	rs4648022	NFKB1	Arm	2.16E-05	4.67	0.03
African American Male	rs1570070	IGF2R	Hip	7.50E-04	3.12	0.73
	rs2110726	IL1R1	Arm	1.53E-06	5.82	0.00
	rs3917225	IL1R1	Arm	4.89E-06	5.31	0.00
	rs948588	SMAD4	Arm	2.16E-05	4.67	0.01
Caucasian Female	rs31476	CSF2	Lumbar Spine	5.20E-05	4.28	0.07
	rs25887	CSF2	Lumbar Spine	9.58E-05	4.02	0.07
Caucasian Male	rs12522630	MEF2C	Arm	7.79E-05	4.11	0.11
	rs2304060	CSF1R	Lumbar Spine	3.61E-05	4.44	0.05
	rs3828562	GNRHR	Hip	9.26E-04	3.03	0.67

Table 5.2: Results from PCH analysis in time points 1, 3 and 5

SNPs coded red in time points 3 and 5 were also significanttime point 1; Snps coded green in time point 5 were also significant at time point 3

Time point 1

Group	SNP	GENE	FULL NAME	p(PCH)	p(PHN)
African American Female	rs31473	CSF2	colony stimulating factor 2	1.42E-03	7.62E-04
	rs2513924	TIEG	Kruppel-like factor 10	2.02E-02	8.34E-04
	rs4149577	TNFRSF1A	tumor necrosis factor receptor superfamily, member 1A	6.61E-01	5.35E-04
	rs2765880	TGFBR3	transforming growth factor, beta receptor III	2.21E-03	5.41E-04
African American Male	rs1570070	IGF2R	insulin-like growth factor 2 receptor	4.49E-02	2.58E-04
Caucasian Female	rs7761846	ESR1	estrogen receptor 1	5.32E-01	1.10E-03
	rs233998	TNFSF10	tumor necrosis factor (ligand) superfamily, member 10	2.69E-04	1.14E-03
	rs2939421	TNFRSF11A	tumor necrosis factor receptor superfamily, member 11A	7.82E-02	9.96E-04
	rs943551	NTRK1	neurotrophic tyrosine kinase, receptor, type 1	1.38E-01	6.19E-04
Caucasian Male	rs235219	TNFRSF1B	tumor necrosis factor receptor superfamily, member 1B	1.14E-01	2.38E-03
	rs3828562	GNRHR	gonadotropin-releasing hormone receptor	2.63E-04	2.96E-04
	rs1178435	IGFALS	insulin-like growth factor binding protein, acid labile subunit	9.49E-03	1.18E-03
	rs952817	TGFB2	transforming growth factor, beta 2	5.98E-02	1.56E-03
	rs7688268	SMAD1	SMAD family member 1	7.72E-04	3.47E-02
	rs3796720	GNRHR	gonadotropin-releasing hormone receptor	1.23E-03	1.65E-03
	rs1531362	NCOA2	nuclear receptor coactivator 2	2.00E-03	5.26E-02
	rs972072	GNRHR	gonadotropin-releasing hormone receptor	2.36E-04	1.18E-03
	rs974483	GNRHR	gonadotropin-releasing hormone receptor	1.78E-04	1.76E-03
	rs344357	IGFALS	insulin-like growth factor binding protein, acid labile subunit	2.23E-02	1.49E-03
	rs10495098	TGFB2	transforming growth factor, beta 2	7.52E-02	1.64E-03
	rs6552115	GNRHR	gonadotropin-releasing hormone receptor	2.20E-03	5.80E-03

Time point 3

Group	SNP	GENE	FULL NAME	p(PCH)	p(PHN)
African American Female	rs2294090	FBXO32	F-box protein 32	7.05E-04	2.12E-04
	rs2198843	CYP1Acluster	chr15, 72788283	4.54E-05	2.33E-02
	rs8191844	IGF2R	insulin-like growth factor 2 receptor	6.05E-04	3.73E-02
	rs4851527	IL1R2	interleukin 1 receptor, type II	1.61E-03	4.15E-03
	rs13894	SHBG	sex hormone-binding globulin	8.49E-04	8.14E-02
African American Male	rs879131	NTRK3	neurotrophic tyrosine kinase, receptor, type 3	1.06E-01	2.82E-04
	rs2230400	IL1R2	interleukin 1 receptor, type II	4.33E-04	4.96E-01
	rs3181139	TNFSF10	tumor necrosis factor (ligand) superfamily, member 10	6.31E-04	8.07E-02
	rs2701653	IRAK3	interleukin-1 receptor-associated kinase 3	3.61E-04	5.49E-02
Caucasian Female	rs3102724	TNFRSF11B	tumor necrosis factor receptor superfamily, member 11B	1.84E-02	3.38E-05
Caucasian Male	rs3828562	GNRHR	gonadotropin-releasing hormone receptor	1.11E-03	1.77E-03
	rs1867348	IGF2R	insulin-like growth factor 2 receptor	2.09E-04	3.21E-04
	rs6179	GHR	growth hormone receptor	8.23E-02	1.80E-03
	rs3796720	GNRHR	gonadotropin-releasing hormone receptor	3.15E-04	5.50E-04
	rs972072	GNRHR	gonadotropin-releasing hormone receptor	9.40E-05	2.67E-04
	rs2830	HSD17B1	hydroxysteroid (17-beta) dehydrogenase 1	2.41E-03	1.70E-02
	rs974483	GNRHR	gonadotropin-releasing hormone receptor	1.55E-03	4.49E-03
	rs4149577	TNFRSF1A	tumor necrosis factor receptor superfamily, member 1A	1.44E-03	1.82E-02
	rs4242119	GHR	growth hormone receptor	1.91E-02	8.50E-04
	rs6552115	GNRHR	gonadotropin-releasing hormone receptor	1.03E-03	3.40E-03

Table 5.2 continued

Time point 5

Group	SNP	GENE	FULL NAME	p(PCH)	p(PHN)
African American Female					
African American Male	rs1470003	SMAD3	SMAD family member 3	4.50E-03	9.71E-04
	rs879131	NTRK3	neurotrophic tyrosine kinase, receptor, type 3	1.14E-01	2.03E-03
	rs1570070	IGF2R	insulin-like growth factor 2 receptor	3.38E-02	7.76E-04
	rs3812429	NCOA2	nuclear receptor coactivator 2	7.01E-02	1.78E-03
	rs3917225	IL1R1	interleukin 1 receptor, type I	2.20E-01	7.33E-05
	rs4477532	LRP6	low density lipoprotein receptor-related protein 6	5.87E-03	3.90E-04
	rs2110726	IL1R1	interleukin 1 receptor, type I	8.99E-04	2.03E-05
	rs2052355	NTF3	neurotrophin 3	1.55E-01	6.20E-04
Caucasian Female					
Caucasian Male	rs12522630	MEF2C	myocyte enhancer factor 2C	2.20E-01	7.61E-06
	rs1531362	NCOA2	nuclear receptor coactivator 2	5.14E-05	4.15E-03

Table 5.3: P-values of equality of slope test between three time points. Hip and Hip Neck BMD traits for SNPs from GNRHR gene.

trait	snp	p13	p15	p35
Hip BMD	rs3796720	0.66	1.00	0.69
	rs3828562	1.00	0.93	0.93
	rs6552115	0.74	0.74	0.54
	rs972072	0.64	1.00	0.66
	rs974483	0.91	0.66	0.61
Hip Neck BMD	rs3796720	0.56	1.00	0.56
	rs3828562	1.00	1.00	1.00
	rs6552115	0.91	0.56	0.48
	rs972072	0.56	0.91	0.48
	rs974483	0.72	0.83	0.59

p13 = p-value when slopes from timepoint 1 and timepoint 3 is compared.

p15 = p-value when slopes from timepoint 1 and timepoint 5 is compared.

p35 = p-value when slopes from timepoint 3 and timepoint 5 is compared

Table 5.4: Results from haplotype analysis of 5 significant SNPs from GNRHR gene on hip and femoral neck BMD in Caucasian male

5.4A Hip BMD

locus	rs3828562	rs3796720	rs972072	rs974483	rs6552115	Hap-Freq	Hap-Score	p-val
	1	1	2	1	1	0.426	-3.233	0.001
	1	2	2	1	1	0.004	-0.940	0.347
	1	2	1	2	1	0.036	-0.890	0.373
	1	2	2	1	2	0.003	-0.562	0.574
	1	2	1	1	2	0.145	-0.427	0.669
haplotypes	1	1	1	1	1	0.009	-0.308	0.758
	1	1	2	1	2	0.047	-0.209	0.834
	1	1	1	2	2	0.027	-0.068	0.946
	1	1	1	2	1	0.003	0.469	0.639
	1	1	2	2	1	0.003	0.744	0.457
	1	2	1	2	2	0.063	0.926	0.354
	2	2	1	2	1	0.031	1.402	0.161
	1	1	1	1	2	0.008	1.548	0.122
	2	2	1	2	2	0.189	3.624	0.000
Over all haplotype analysis				24.91451	df=14			0.035

5.4B Femoral Neck BMD

locus	rs3828562	rs3796720	rs972072	rs974483	rs6552115	Hap-Freq	Hap-Score	p-val
	1	1	2	1	1	0.425	-2.880	0.004
	1	2	2	1	1	0.005	-0.735	0.462
	1	1	2	2	1	0.004	-0.551	0.581
	1	1	1	1	1	0.008	-0.549	0.583
	1	2	1	2	1	0.036	-0.462	0.644
haplotypes	1	1	1	2	2	0.027	-0.202	0.840
	1	1	1	2	1	0.003	-0.193	0.847
	1	2	2	1	2	0.004	-0.186	0.852
	1	1	2	1	2	0.046	0.145	0.885
	1	2	1	1	2	0.146	0.663	0.507
	1	2	1	2	2	0.064	0.759	0.448
	2	2	1	2	1	0.030	0.792	0.429
	1	1	1	1	2	0.008		0.814
	2	2	1	2	2	0.188	2.743	0.006
Over all haplotype analysis				14.20031	df=14			0.435

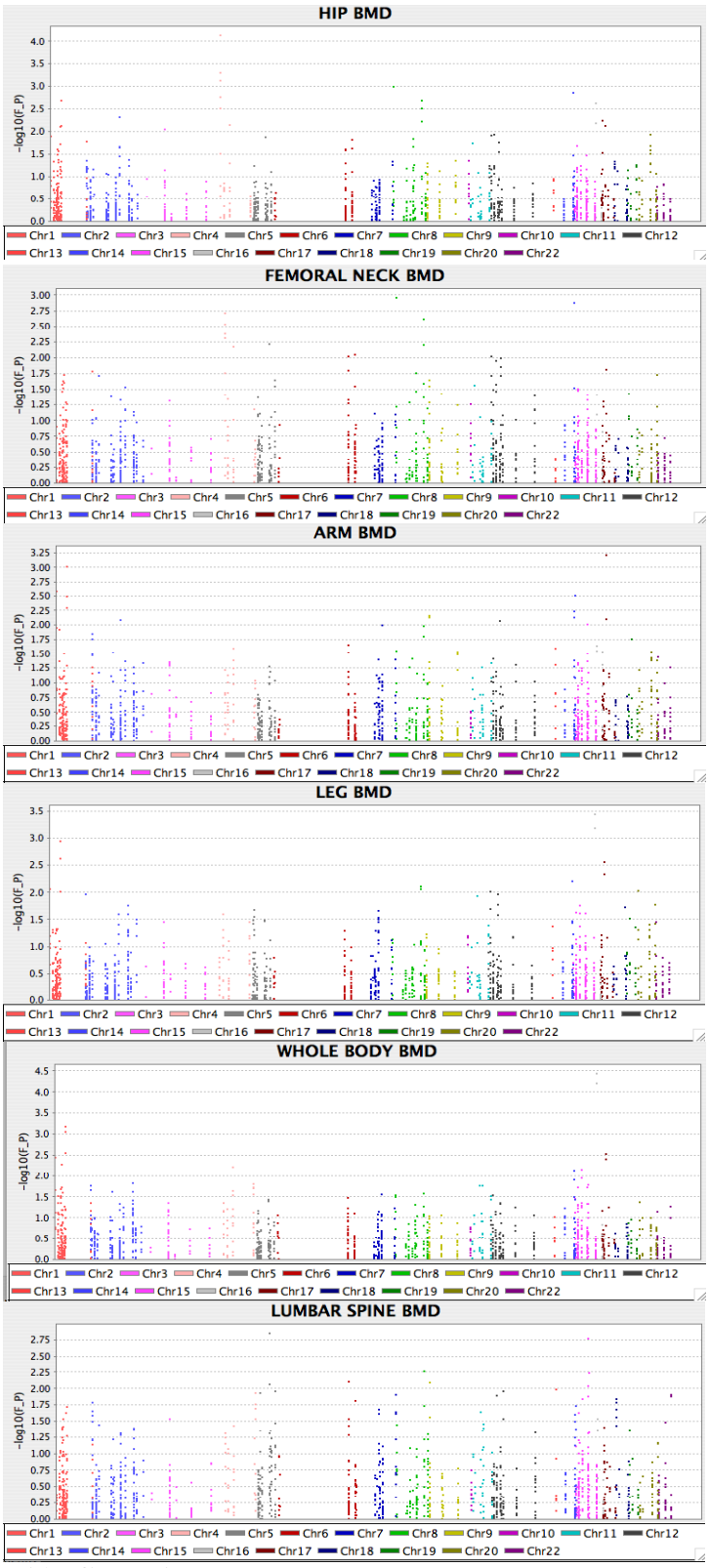
1 at each locus denotes major allele and 2 denotes minor allele

Table 5.5: Results from MrOS analysis on baseline significant SNPs from Caucasian males in HABC cohort

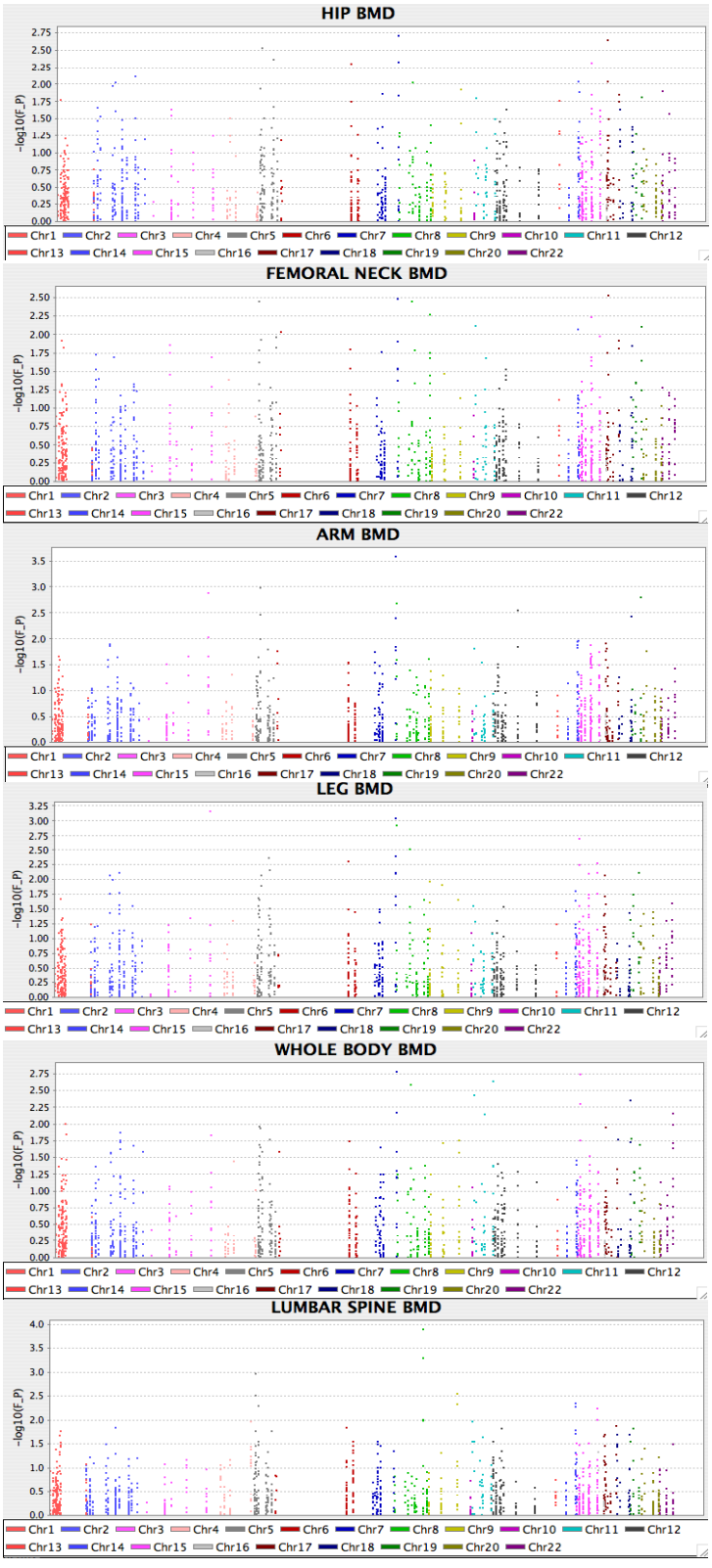
BMD trait	SNP	Gene	Chr	Intercept	p-value
hip	rs1843593	GNRHR	4	-0.009	0.29
	rs1038426	GNRHR	4	-0.008	0.20
	rs3828562	GNRHR	4	-0.009	0.21
	rs3796720	GNRHR	4	-0.011	0.07
	rs974483	GNRHR	4	-0.017	0.01
	rs972072	GNRHR	4	0.012	0.06
	rs6828922	GNRHR	4	0.010	0.40
	rs6552115	GNRHR	4	-0.011	0.08
	rs2711158	GNRHR	4	0.018	0.03
Femoral neck	rs1843593	GNRHR	4	-0.007	0.38
	rs1038426	GNRHR	4	-0.005	0.38
	rs3828562	GNRHR	4	-0.004	0.50
	rs3796720	GNRHR	4	-0.006	0.29
	rs974483	GNRHR	4	-0.011	0.07
	rs972072	GNRHR	4	0.007	0.24
	rs6828922	GNRHR	4	0.008	0.47
	rs6552115	GNRHR	4	-0.006	0.30
	rs2711158	GNRHR	4	0.008	0.29
Mean Arm	rs1178435	IGFALS	16	0.004	0.46
	rs344357	IGFALS	16	-0.001	0.81
	rs4316755	IGFALS	16	-0.004	0.40
	rs2268671	IGFALS	16	0.008	0.11
Mean Leg	rs1178435	IGFALS	16	0.000	0.98
	rs344357	IGFALS	16	-0.007	0.58
	rs4316755	IGFALS	16	0.001	0.95
	rs2268671	IGFALS	16	0.007	0.57
Whole Body	rs1178435	IGFALS	16	0.007	0.36
	rs344357	IGFALS	16	0.000	0.97
	rs4316755	IGFALS	16	-0.008	0.27
	rs2268671	IGFALS	16	0.005	0.50

Table 5.6: Allele frequency of GNRHR SNPs in the two populations Caucasian American Males in HABC and MrOS

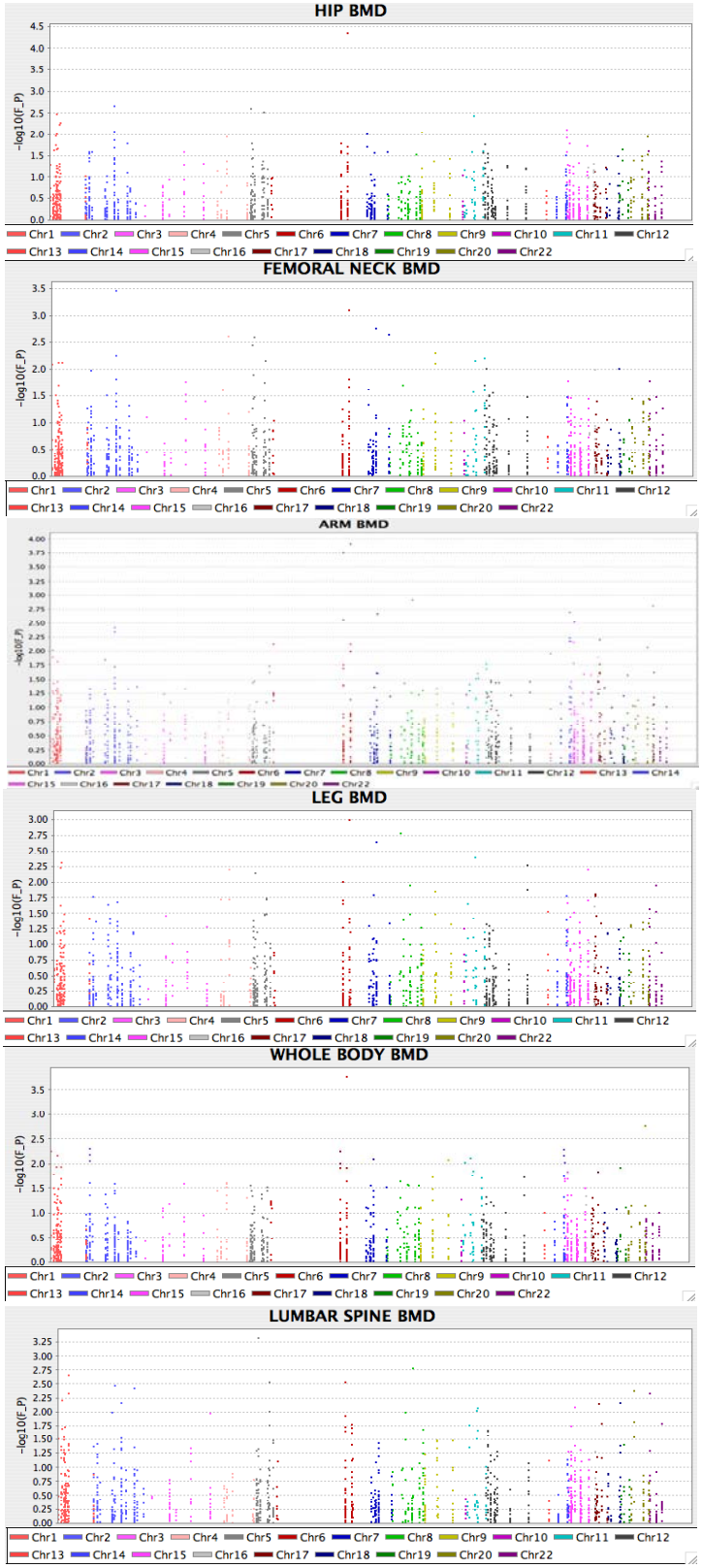
SNP	Minor Allele	Minor Allele Frequency	
		HABC (CAM)	MrOS
rs3828562	A	0.22	0.24
rs3796720	A	0.48	0.47
rs974483	G	0.35	0.36
rs972072	A	0.49	0.49
rs6552115	A	0.49	0.49



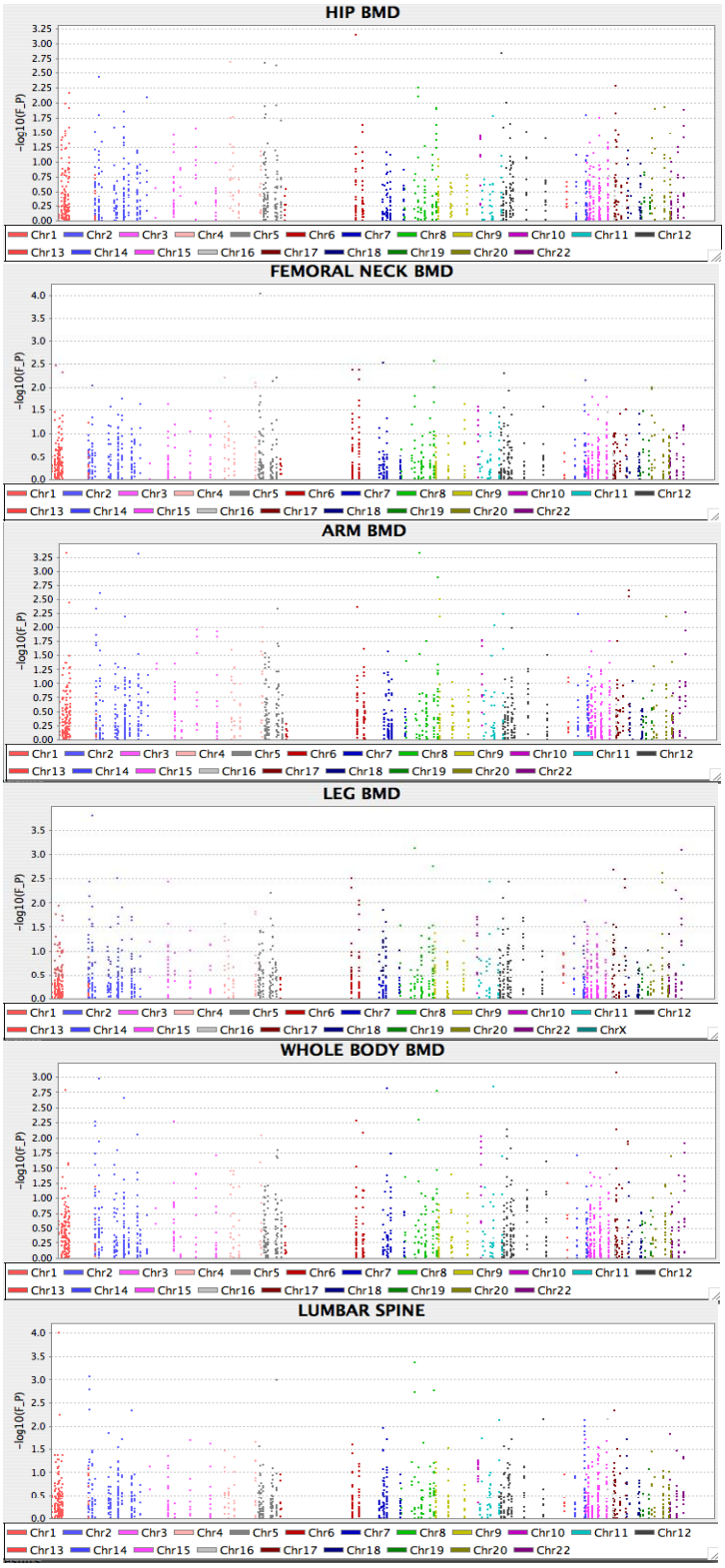
5.1A: Caucasian Males



5.1B: Caucasian Females

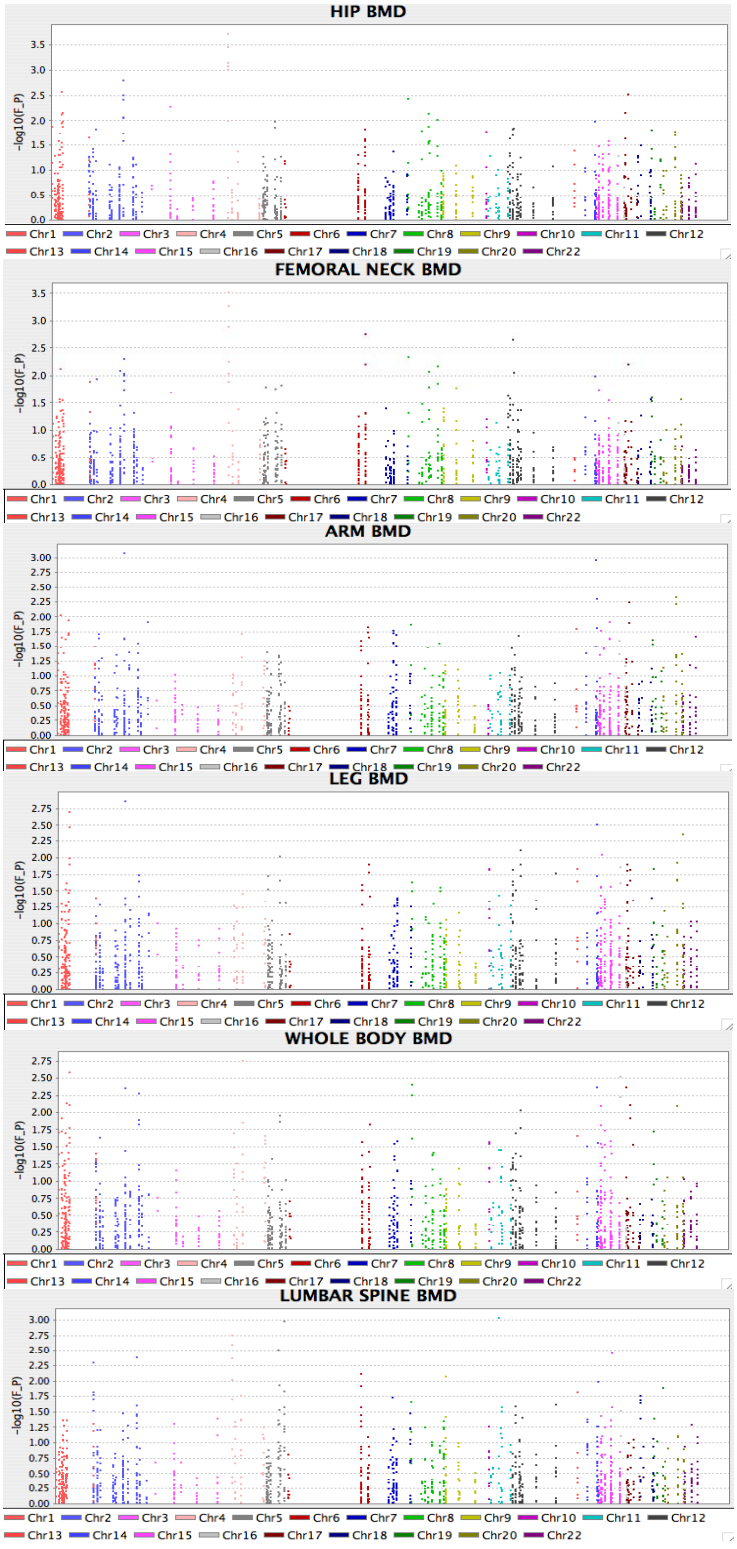


5.1C: African-American Males

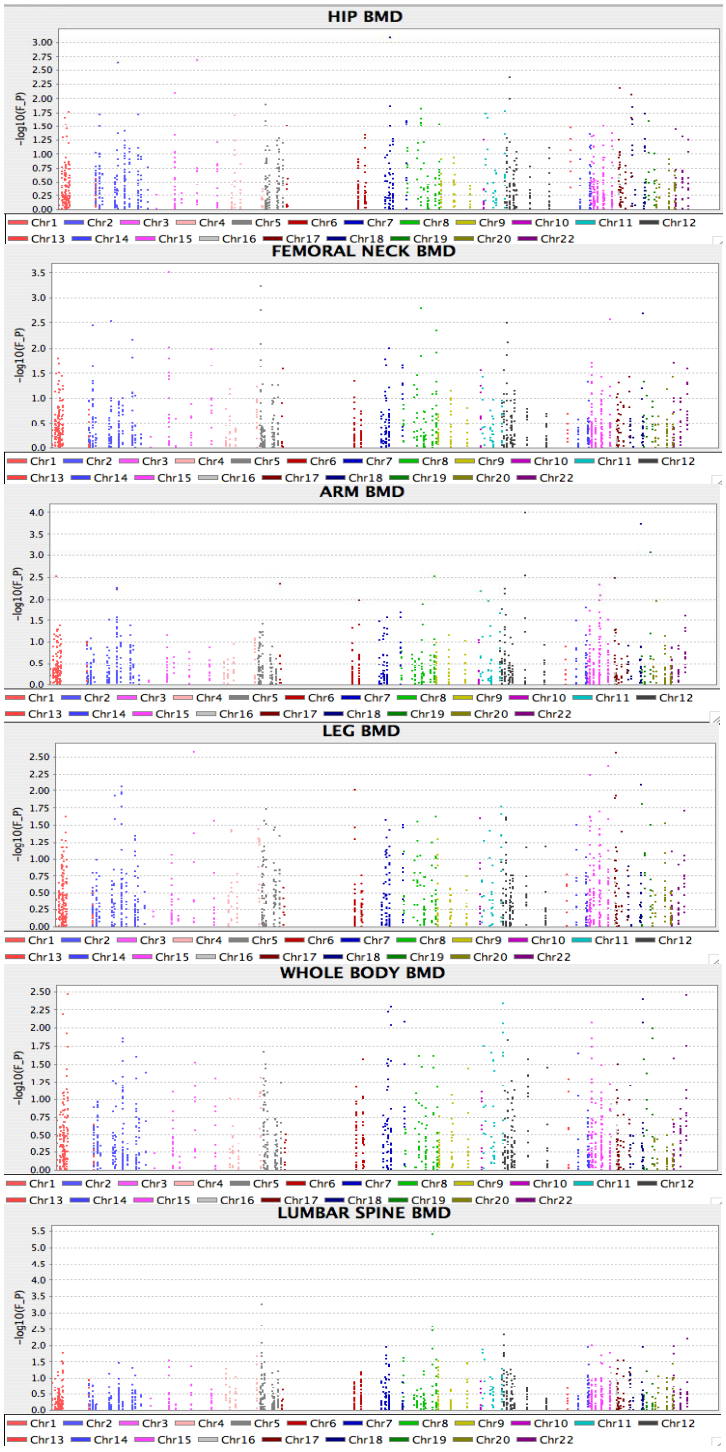


5.1D: African-American Females

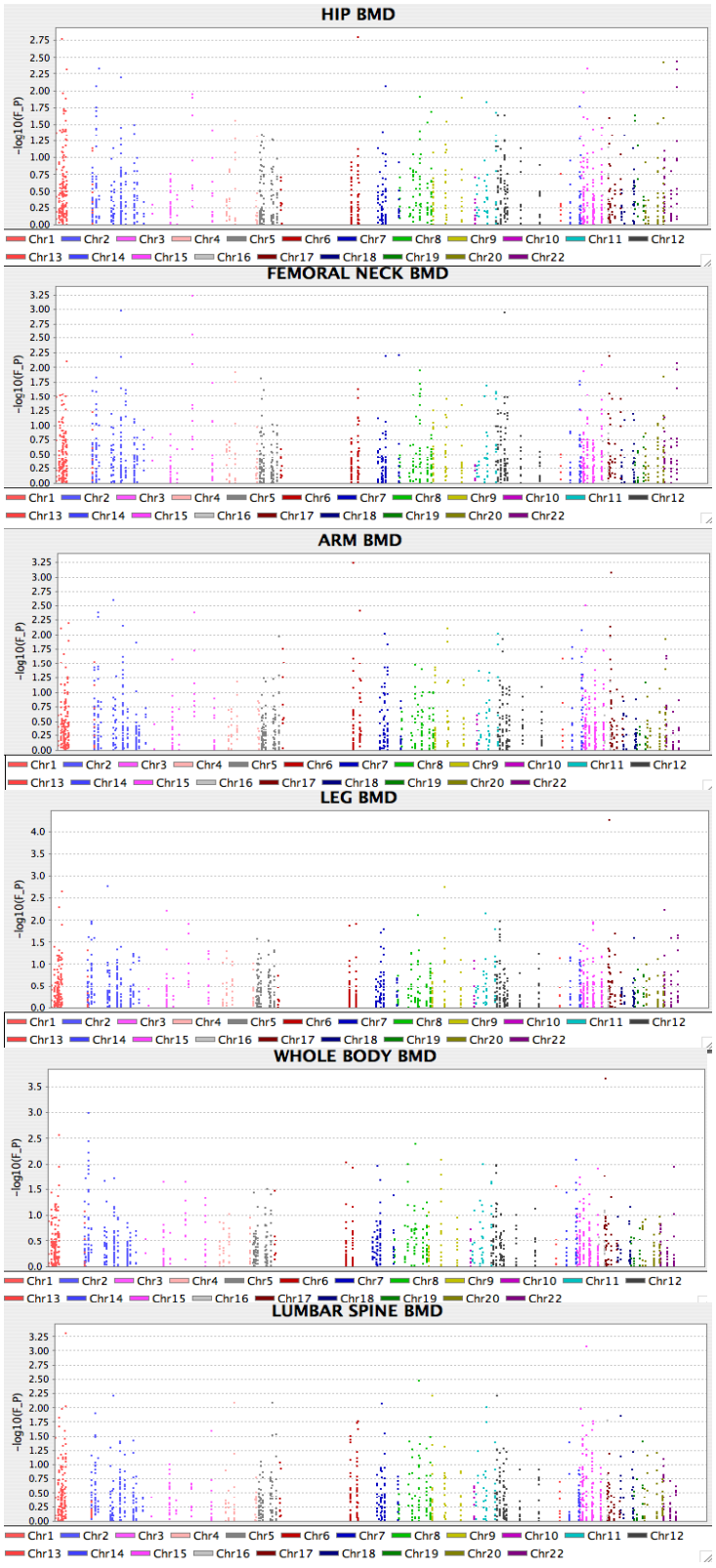
Figure 5.1: Plot of –log10(P) value for each trait (single SNP single trait analysis time point 1)



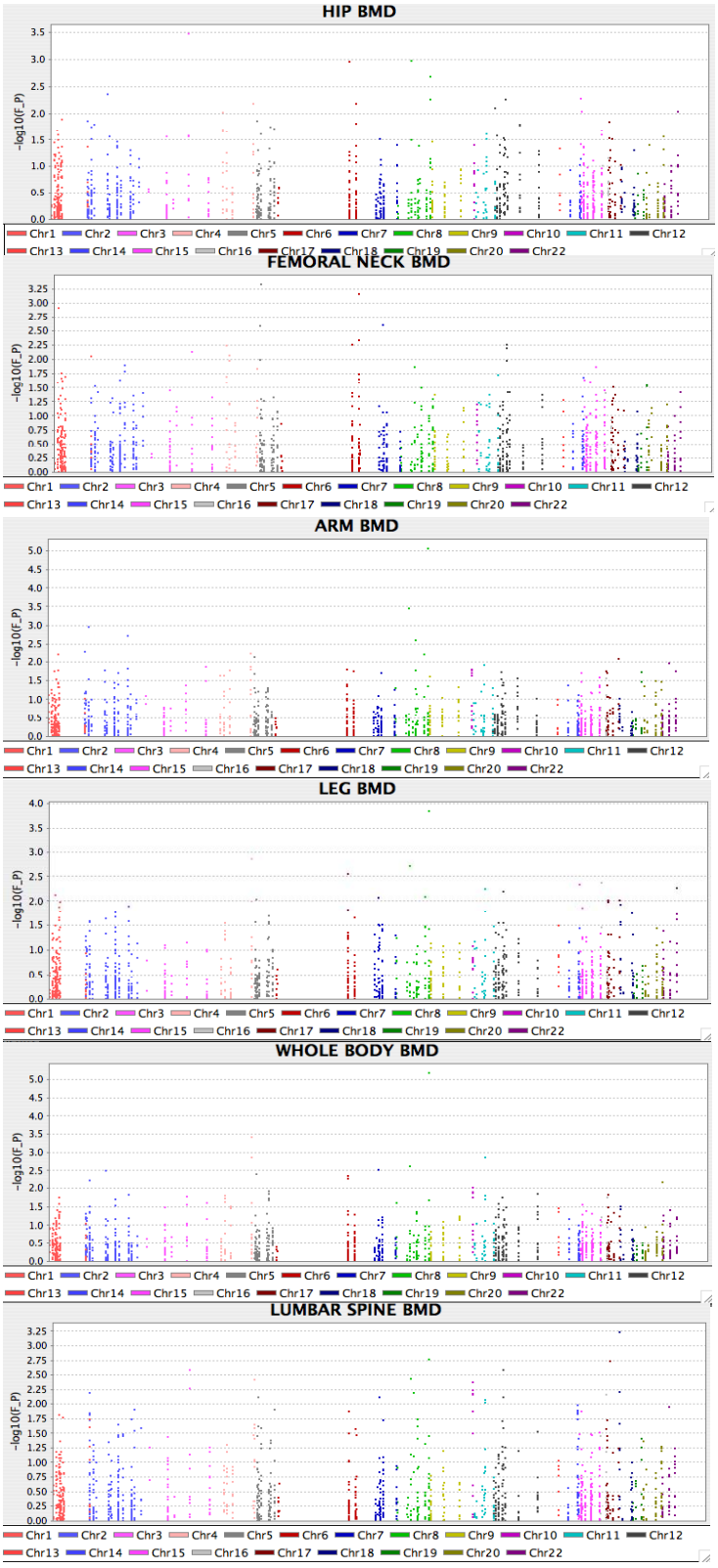
5.2A: Caucasian Males



5.2B: Caucasian Females

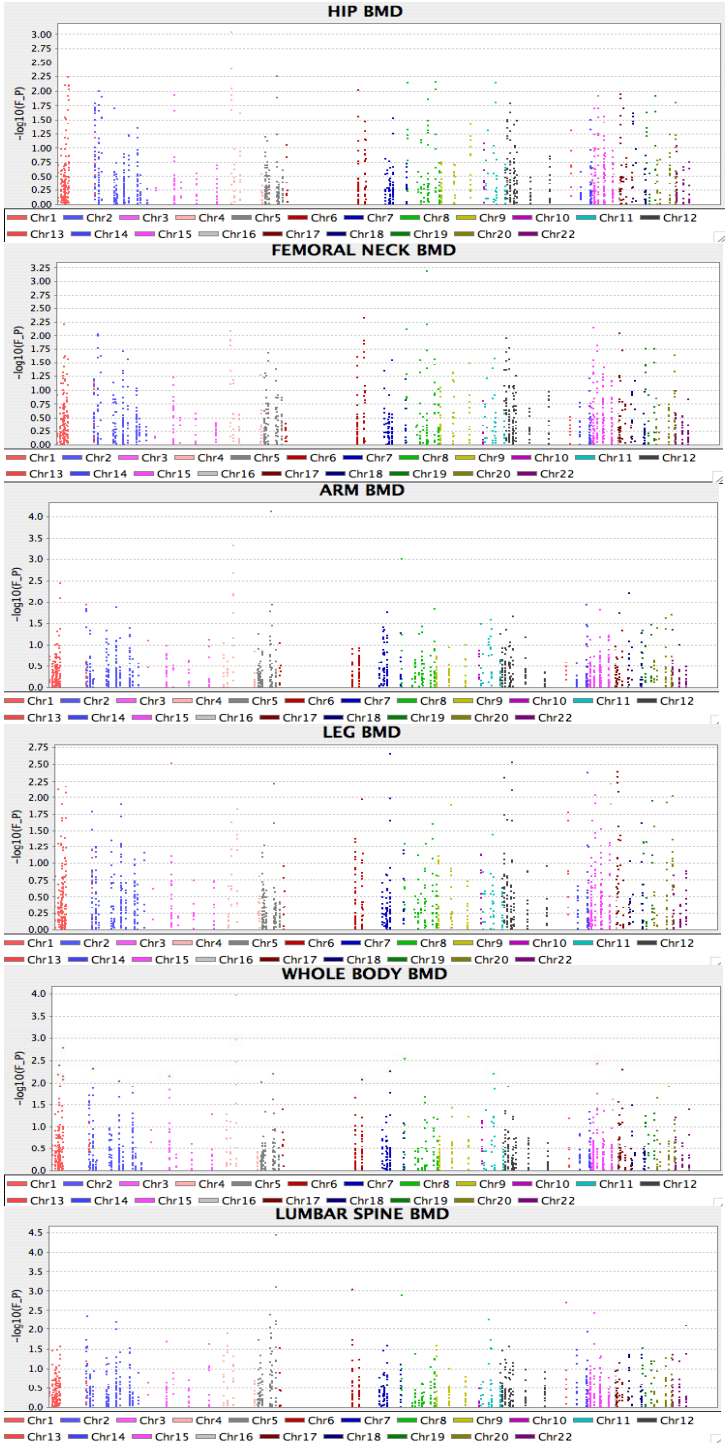


5.2C: African-American Males

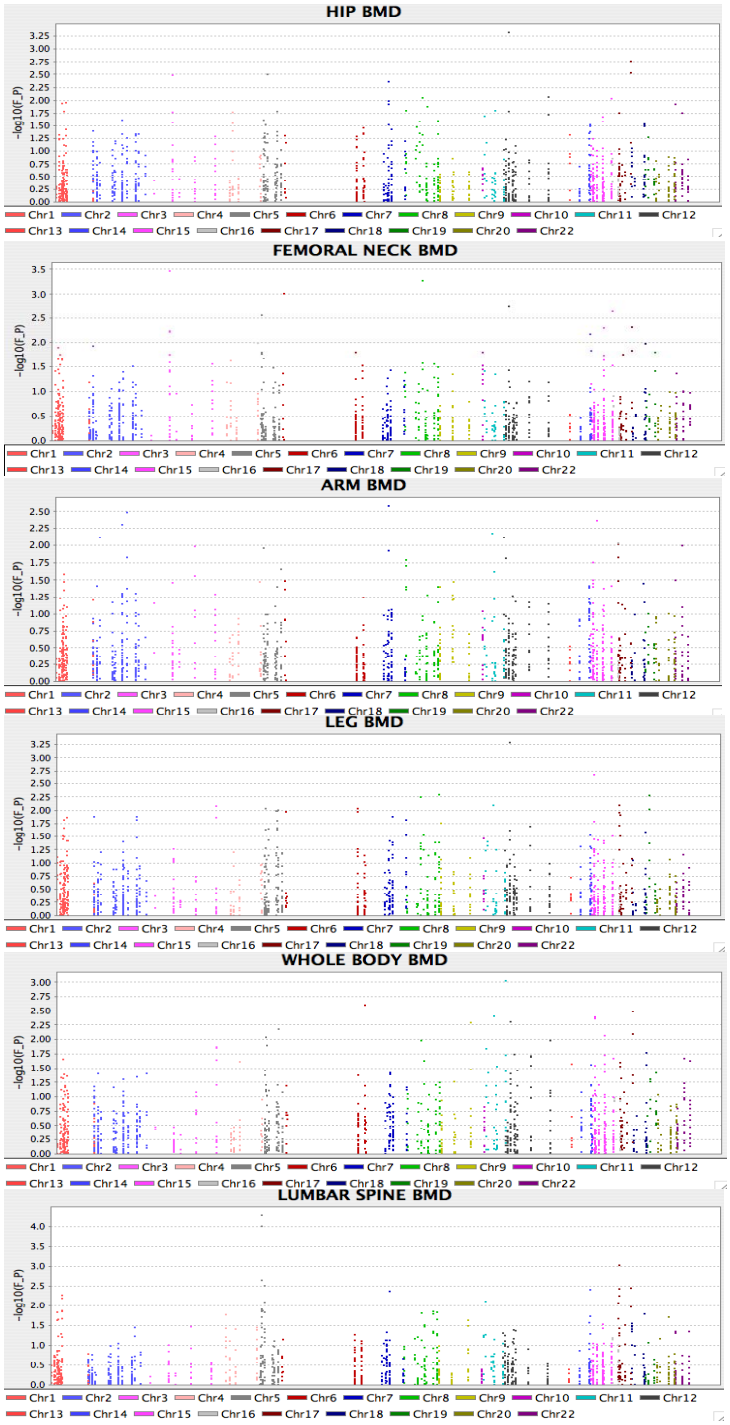


5.2D: African-American Females

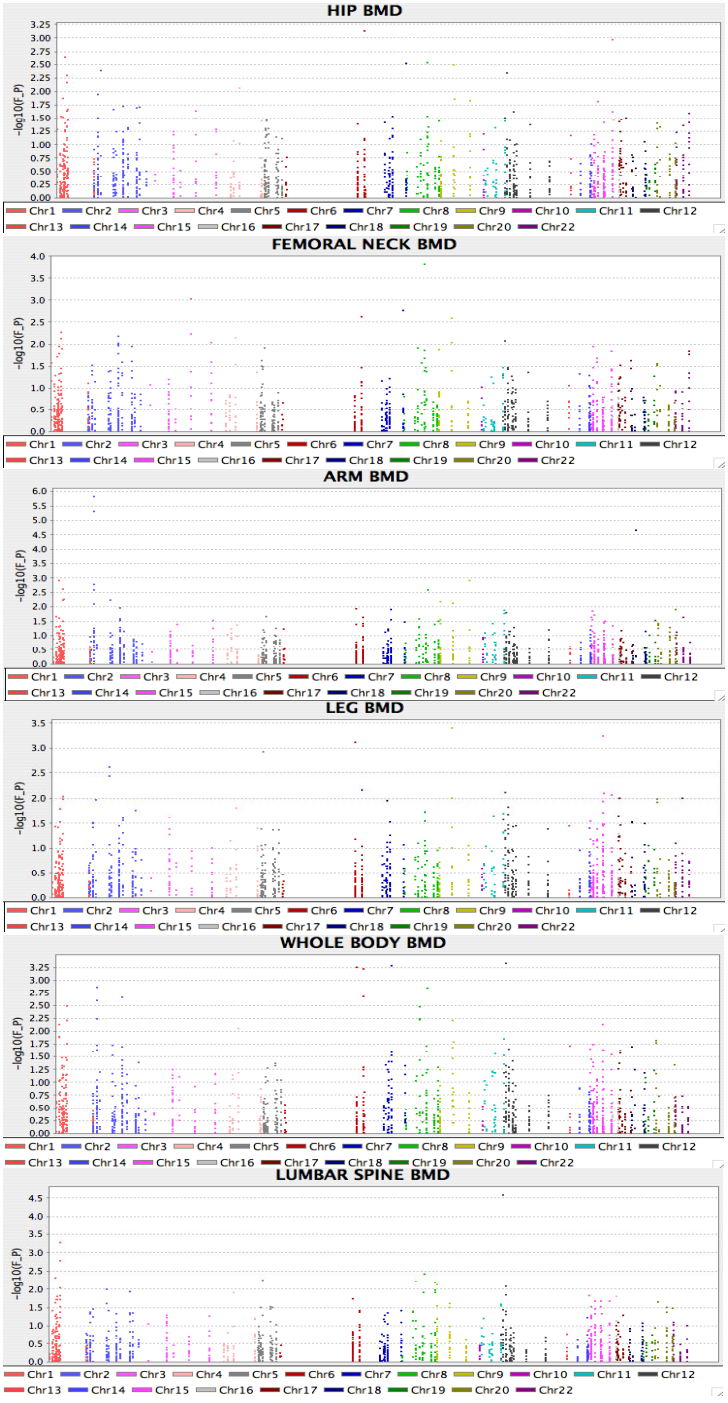
Figure 5.2: Plot of –log10(P) value for each trait (single SNP single trait analysis time point 3)



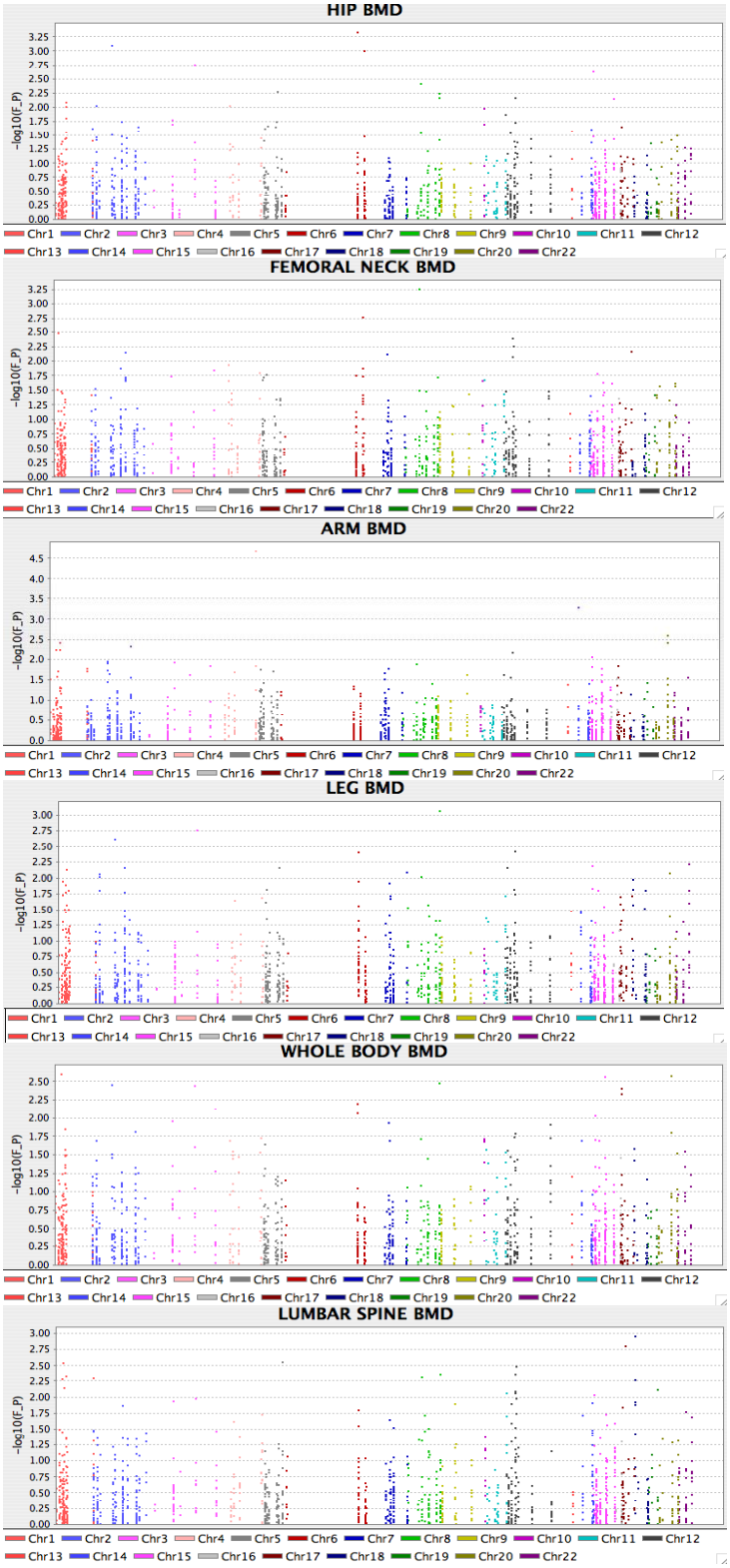
5.3A: Caucasian Males



5.3B: Caucasian Females

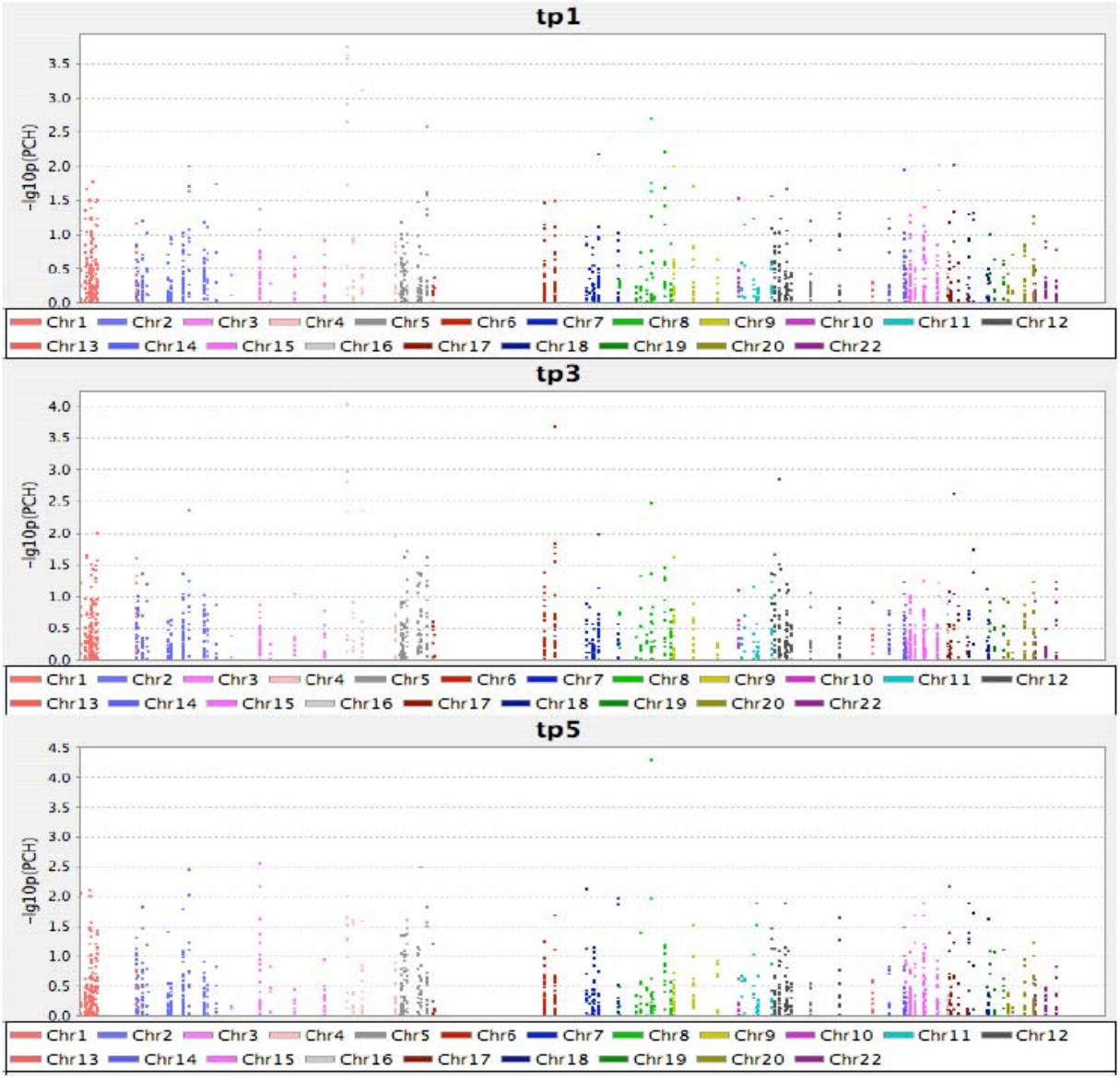


5.3C: African-American Males

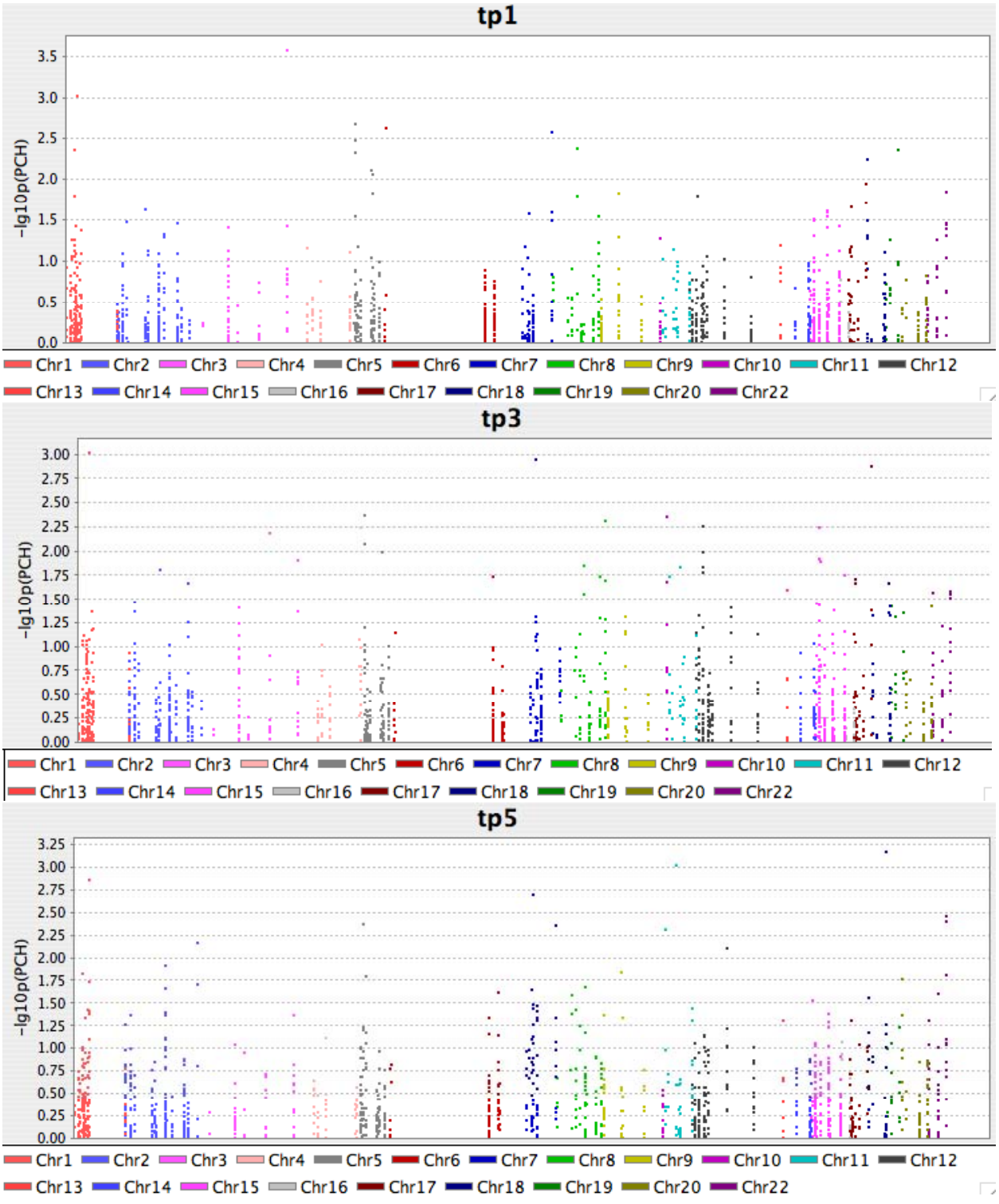


5.3D African American Females

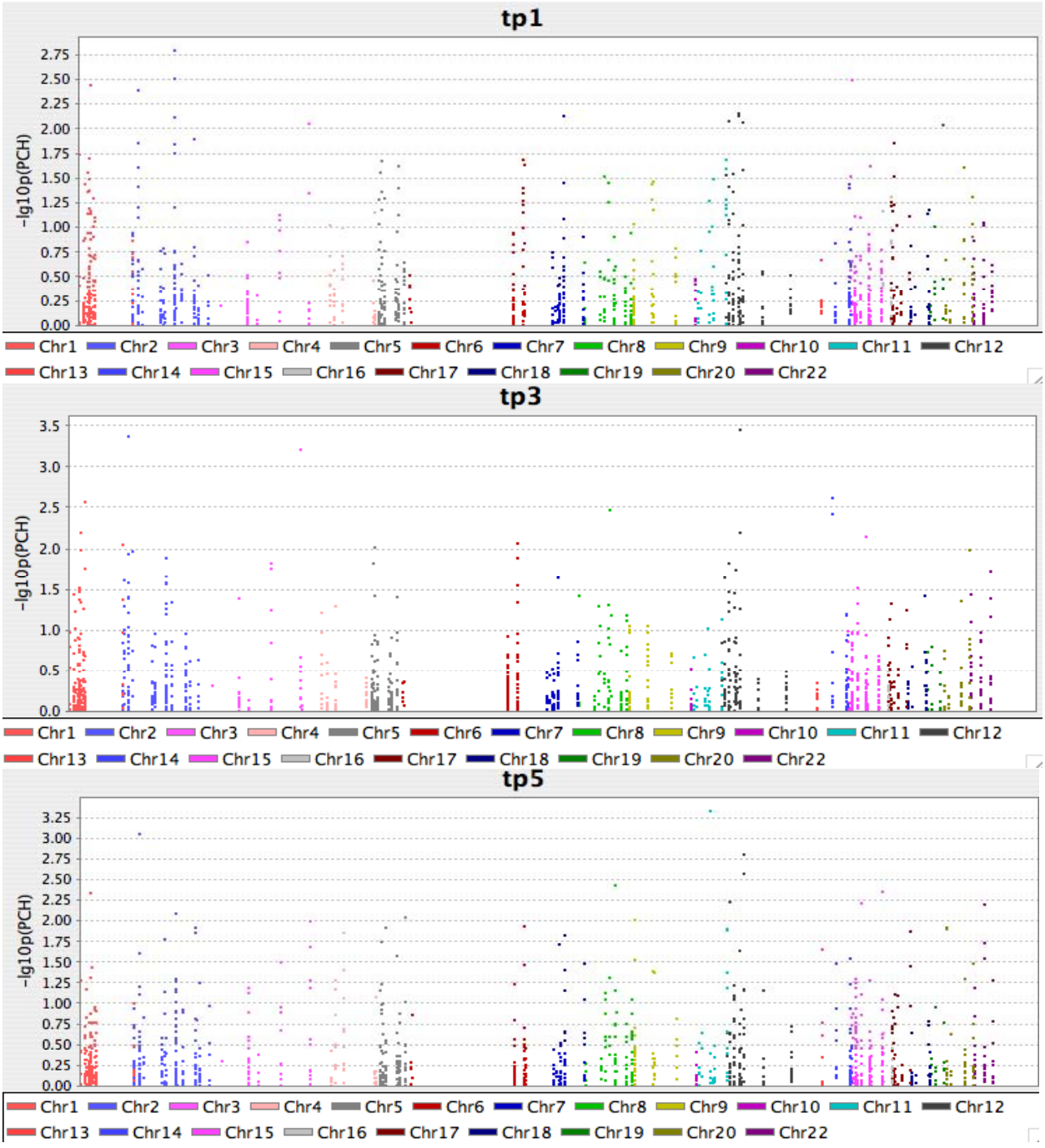
Figure 5.3: Plot of –log10(P) value for each trait (single SNP single trait analysis time point 5)



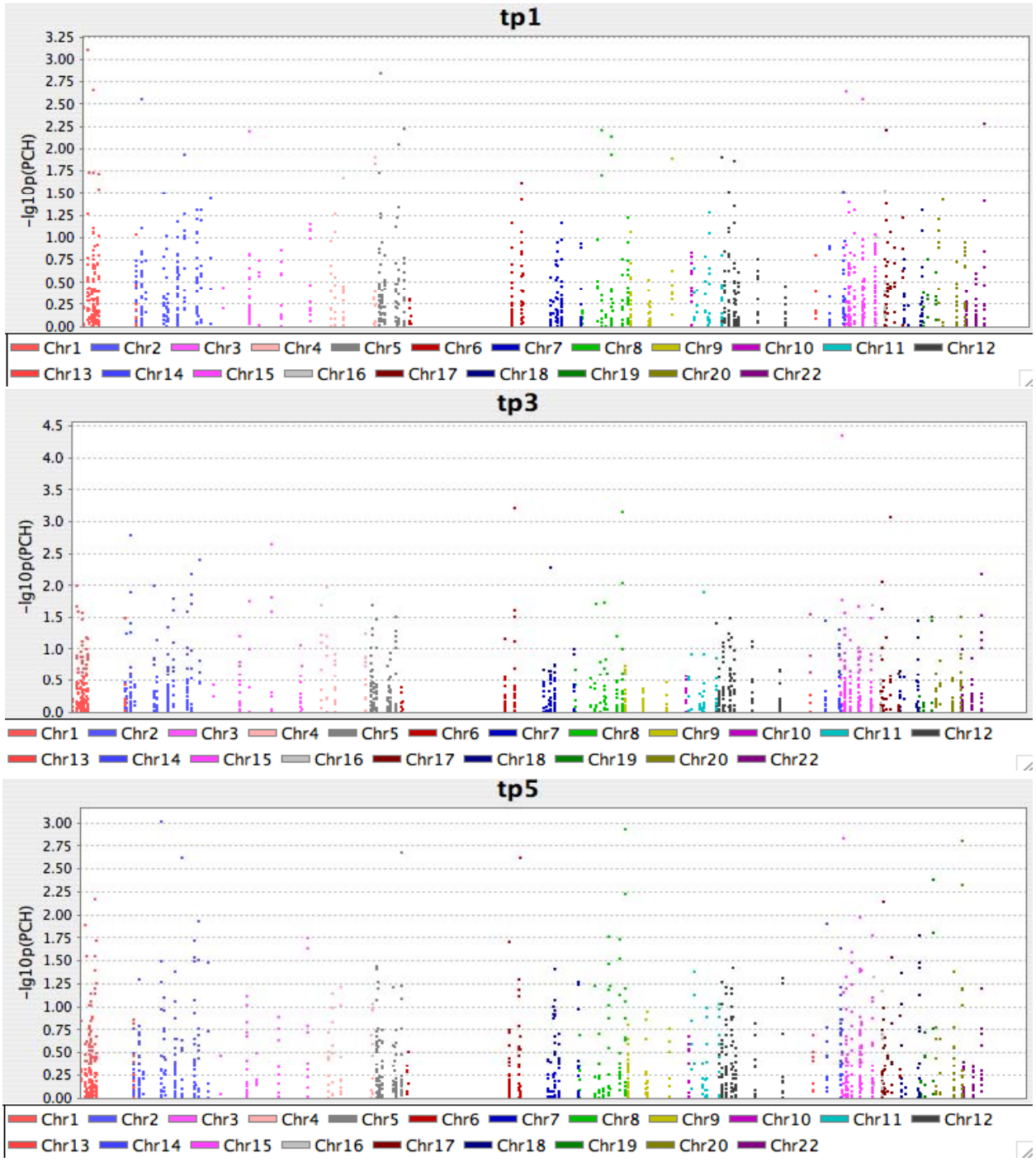
5.4A: Caucasian Males



5.4B: Caucasian Females

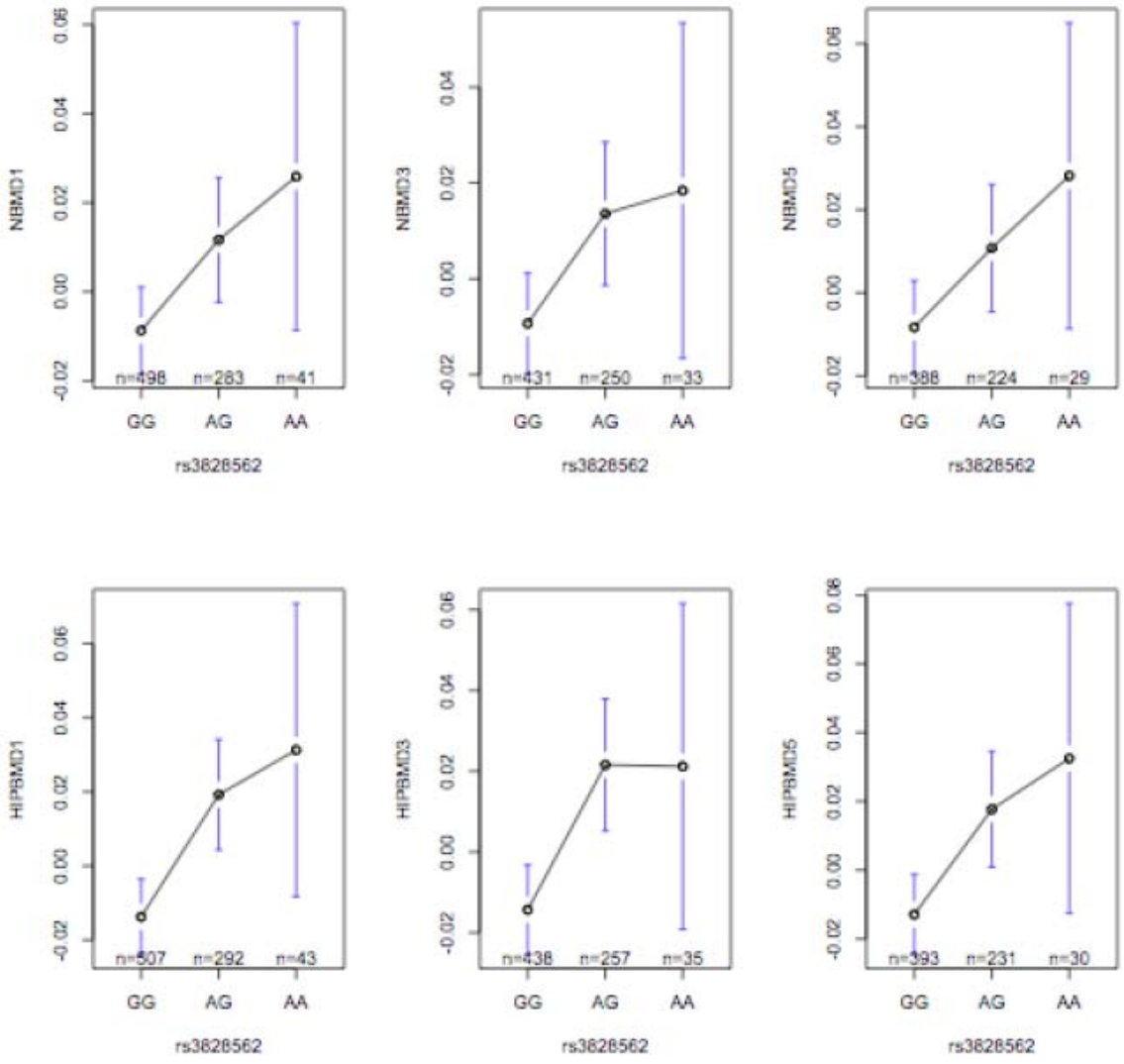


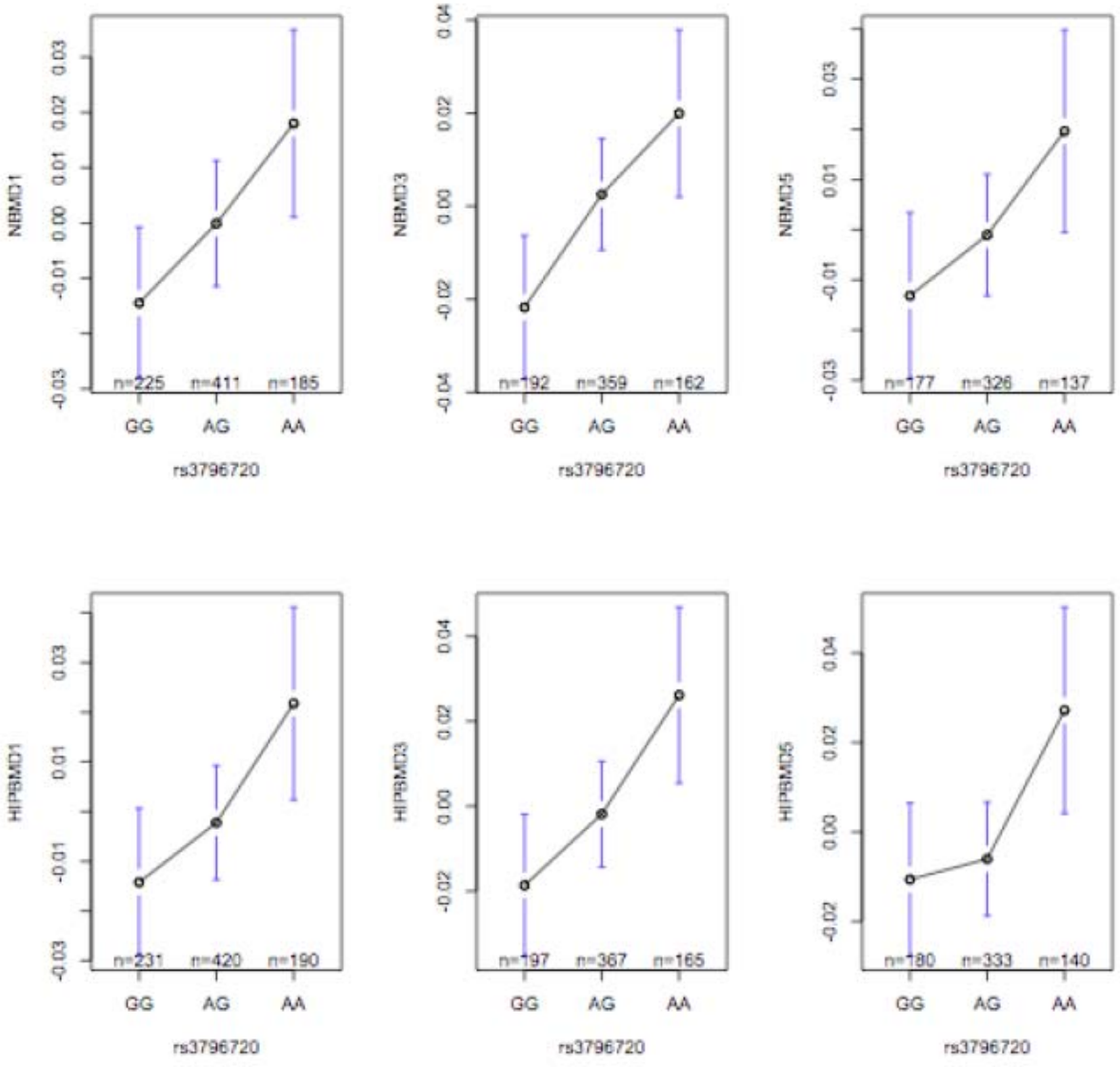
5.4C: African-American Males

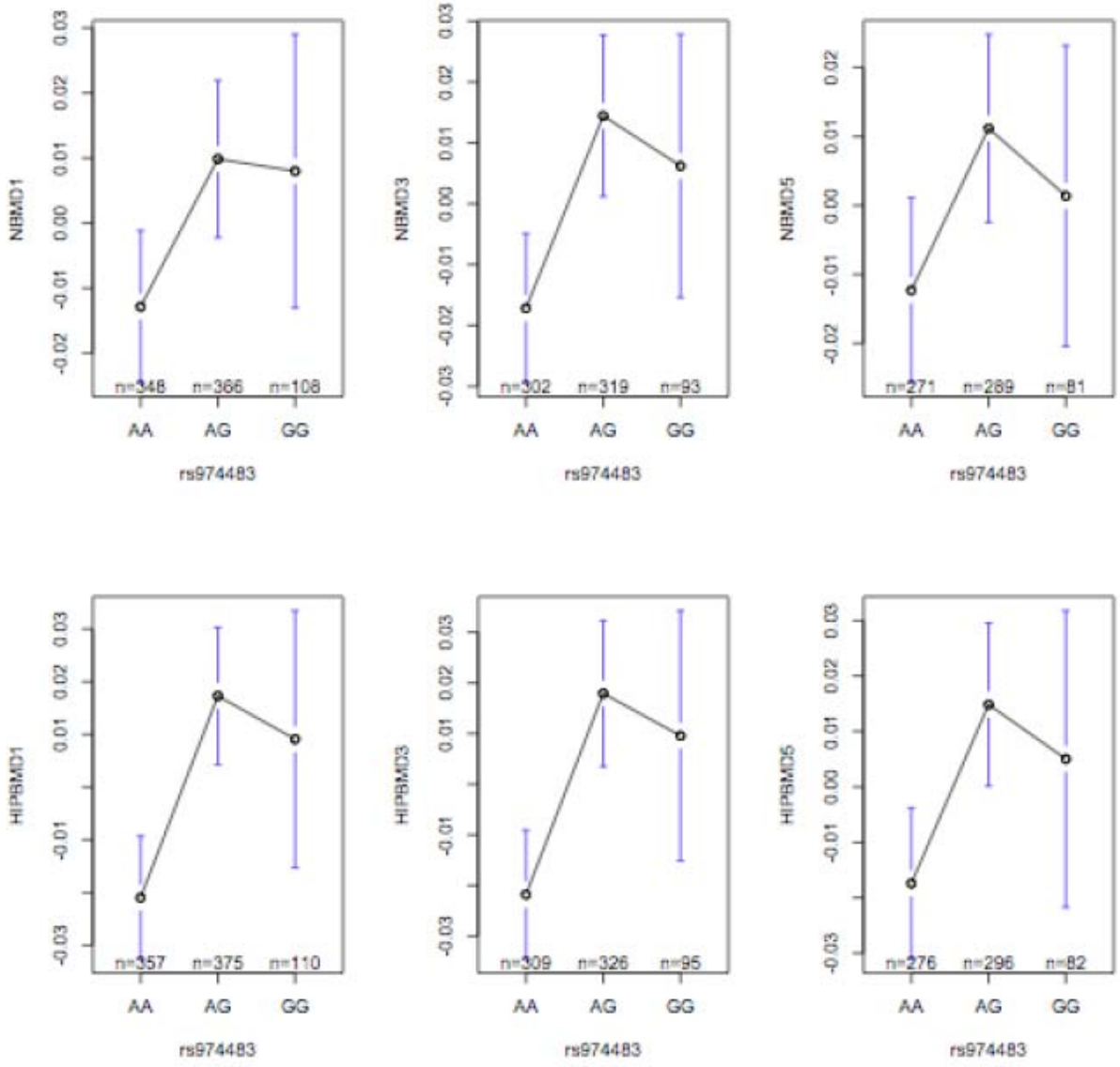


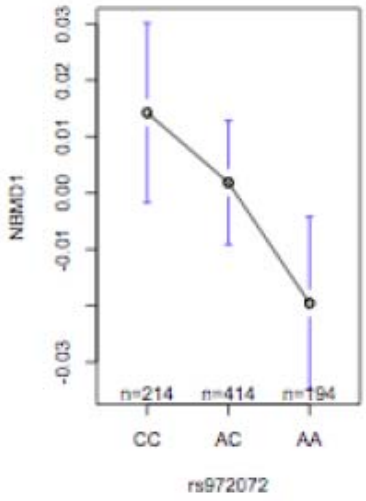
5.4D: African-American Females

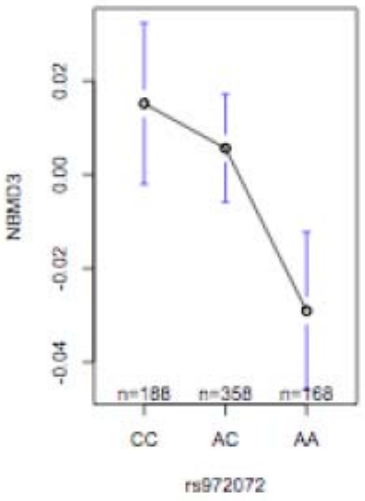
Figure 5.4: Plot of –log10(PCH) value for each group (single SNP multiple trait analysis)

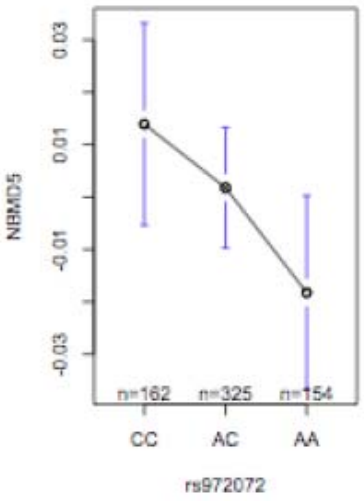

NBMD1
NBMD3
NBMD5
HIPBMD1
HIPBMD3
HIPBMD5
-0.02
0.00
0.02
0.04
0.06
n=498
n=283
n=41
n=431
n=250
n=33
n=388
n=224
n=29
n=507
n=292
n=43
n=438
n=257
n=35
n=393
n=231
n=30
GG
AG
AA
rs3828562

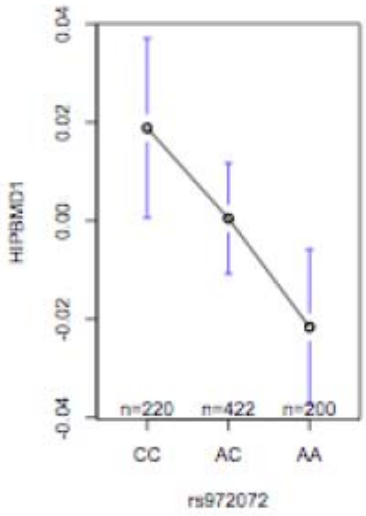

NBMD1
-0.03
-0.01
0.00
0.01
0.02
0.03
n=225
n=411
n=185
GG
AG
AA
rs3796720
NBMD3
-0.04
-0.02
0.00
0.02
0.04
n=192
n=359
n=162
GG
AG
AA
rs3796720
NBMD5
-0.03
-0.01
0.01
0.03
n=177
n=326
n=137
GG
AG
AA
rs3796720
HIPBMD1
-0.03
-0.01
0.01
0.03
n=231
n=420
n=190
GG
AG
AA
rs3796720
HIPBMD3
-0.02
0.00
0.02
0.04
n=197
n=367
n=165
GG
AG
AA
rs3796720
HIPBMD5
-0.02
0.00
0.02
0.04
n=180
n=333
n=140
GG
AG
AA
rs3796720

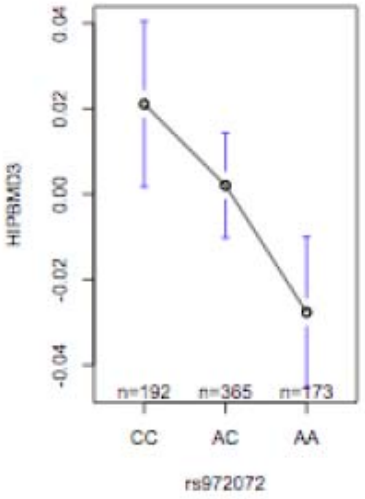

NBMD1
NBMD3
NBMD5
HIPBMD1
HIPBMD3
HIPBMD5
n=348
n=366
n=108
n=302
n=319
n=93
n=271
n=289
n=81
n=357
n=375
n=110
n=309
n=326
n=95
n=276
n=296
n=82
AA
AG
GG
rs974483

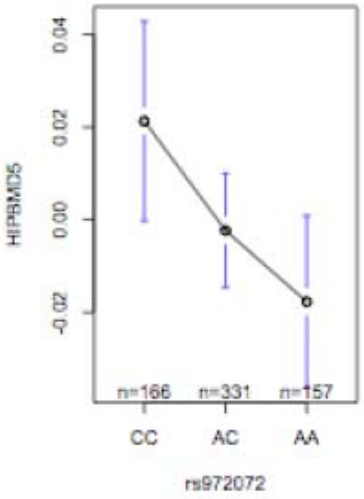

NBMD1
0.03
0.02
0.01
0.00
-0.01
-0.03
n=214
n=414
n=194
CC
AC
AA
rs972072


NBMD3
0.02
0.00
-0.02
-0.04
n=188
n=358
n=168
CC
AC
AA
rs972072


NBMD5
0.03
0.01
-0.01
-0.03
n=162
n=325
n=154
CC
AC
AA
rs972072


HIPBMD1
0.04
0.02
0.00
-0.02
-0.04
n=220
n=422
n=200
CC
AC
AA
rs972072


HIPBMD3
0.04
0.02
0.00
-0.02
-0.04
n=192
n=365
n=173
CC
AC
AA
rs972072


HIPBMD5
0.04
0.02
0.00
-0.02
n=166
n=331
n=157
CC
AC
AA
rs972072

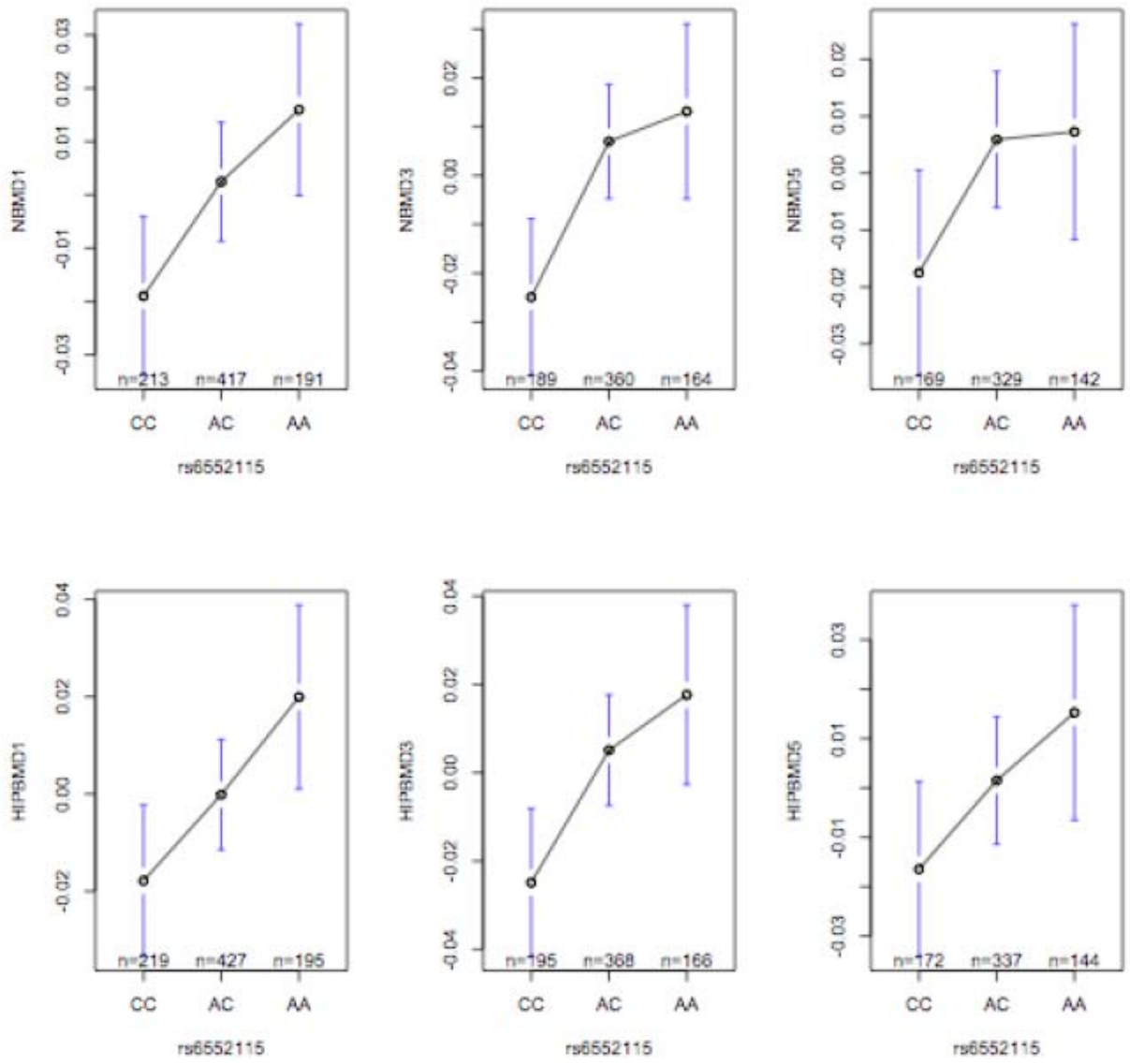


Figure 5.5: Plot of residual mean of Hip and Femoral Neck BMD for genotypes of each significant SNP in the GNRHR gene for Caucasian Males

NBMD1: Femoral Neck BMD and time point 1; NBMD3: Femoral Neck BMD and time point 3; NBMD5: Femoral Neck BMD and time point 5

HipBMD1: Hip BMD and time point 1; HipBMD3: Hip BMD and time point 3; HipBMD5: Hip BMD and time point 5. The notations are same for all the five SNPs.

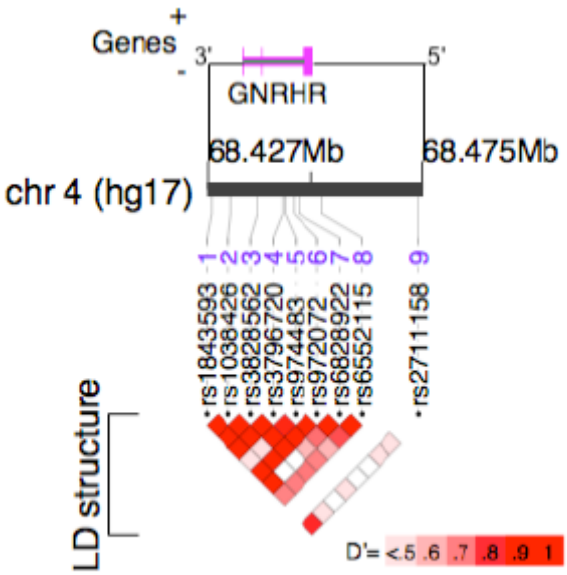


Figure 5.6: LD structure of SNPs from GNRHR gene

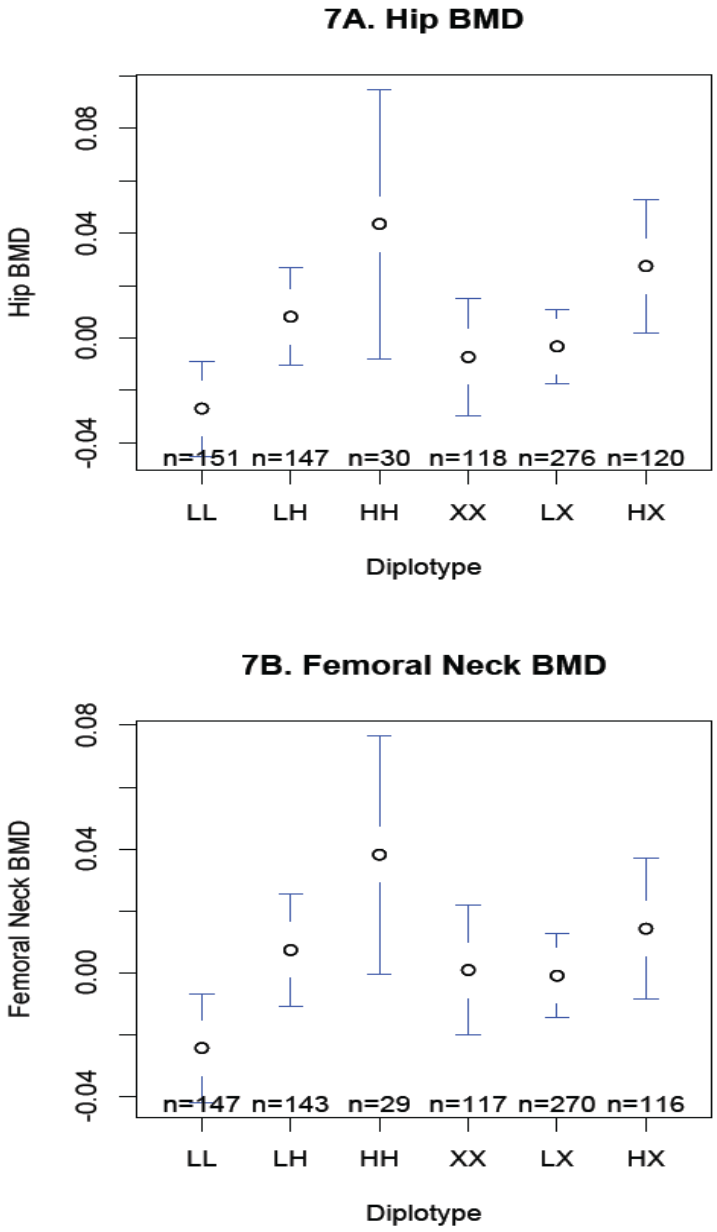


Figure 5.7: Plot of mean BMD at hip and femoral neck skeletal sites across different groups of diplotypes (using the five significant SNPs in GNRHR) for Caucasian American males

H denotes: haplotype that is associated with high level of BMD; L denotes: haplotype that is associated with low level of BMD; X denotes: all haplotypes other than H and L.

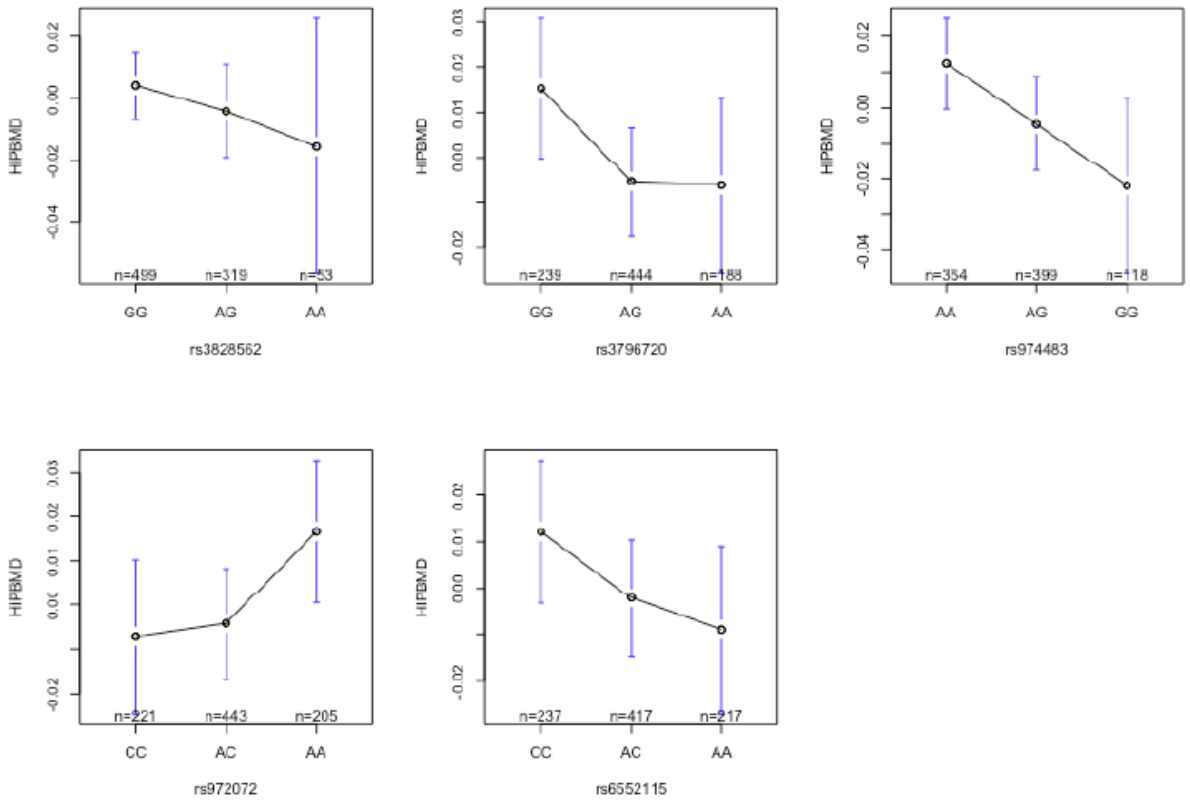


Figure 5.8: Plot of residual mean of Hip BMD for genotypes of each significant SNP in the GNRHR gene for MrOS population

6.0 ASSOCIATION ANALYSIS OF BMD TRAITS WITH GENES IN COMMON PATHWAYS

6.1 INTRODUCTION

Over a decade ago, Risch and Merikangas (Risch et al. 1996) proposed that genome wide association studies (GWA) would have greater power than linkage studies to detect polymorphisms that influence modest disease risk. With the development of high throughput methods for genotyping thousands of markers on thousands of individuals, the identification of millions of SNPs (Smigielski et al. 2000) , as well as the International HapMap project to determine linkage disequilibrium (The International HapMap Consortium, 2007), the GWA approach became feasible. As a result, many GWA studies have been performed resulting in the identification of potentially novel candidate genes for disease (Easton et al. 2007; Saxena et al. 2007). However, not all GWA studies have been successful, for example, no candidate genes were identified in a study of hypertension (2007) and some GWA have identified candidate genes with very small effects on a trait, such as those for height (Zeggini et al. 2007). The mixed success of these studies probably reflects how well each trait being analyzed fits the underlying assumptions regarding the genetic and environmental architecture, as well as other methodological issues pertaining to GWA studies and candidate gene studies in general. Some of these issues were previously discussed in the Introduction (Chapter 1).

Most investigators assume that complex traits are often influenced by many genetic and environmental factors (Zaykin et al. 2005; Freimer et al. 2007), however the current GWA studies report only the most significant results from a single SNP-single trait analytical approach and ignore the possible

effects of the interplay among several genetic factors (Zaykin et al. 2005). One of the reasons most investigators ignore possible gene by gene or gene by environment interactions is that the number of tests would increase exponentially, with a consequent loss of power to detect such effects. An alternate approach to identify genes that influence development of common diseases is to investigate the combined effect of a set of genes, or common pathways, on a specific disease or quantitative trait. Although most of the genes, or SNPs in these genes, might have very small effect on our trait of interest, the combined effect might be significant. One method by which to investigate the combined effect of a set of genes on a trait is pathway-based analysis (Wang et al. 2007). This pathway-based approach is motivated by the gene-set enrichment analyses (GSEA) approach that is used in the analysis of microarray expression data analysis (Subramanian et al. 2005). Wang and colleagues (Wang et al. 2007) modified this GSEA approach for use in GWA studies. We hypothesized that the pathway-based association method might provide additional insights to results from candidate gene association studies. In particular, it might indicate gene pathways (or networks) that would be useful to explore possible genotype by genotype interactions. Therefore, we applied this methodology to our data on BMD in the HABC cohorts. Further details regarding the specifics of the pathway-based approach are presented in the Methods section below.

As described in the Chapter 5, we are using several approaches to determine whether variation in specific candidate genes is associated with variation in BMD at different skeletal sites. First, we performed a single SNP-single trait analysis (within each sex by ancestry group). Next, we performed single SNP-multiple trait analyses by analyzing all the SNPs with a linear combination of traits, the PCH method (Klei et al. 2008). The results of these analyses were described in Chapter 5.

In this chapter, we describe the results of our analyses using the pathway-based approach. Because the pathway-based analyses were originally developed for use with GWA data, that is, for use on $\geq$ 100,000 SNPs among $\geq$ 20,000 loci, we were concerned that the approach might not be suitable for our much smaller sample of ~1500 SNPs among 138 loci. Specifically, we questioned whether (1) we had a sufficient number of loci and/or pathways among our 138 loci to obtain meaningful results. In other words, how does the total number of loci influence the significance of the results? (2) We also wondered

how the definition of a specific pathway influences the results. There are several programs, such as Ingenuity (Ingenuity Systems, www.ingenuity.com) and STRING (Snel et al. 2000), that can be used to identify possibly interacting genes or proteins, and each program will give slightly different results.

We performed a series of analyses using data on the HABC cohorts and the MrOS cohort to address the following specific questions:

1. Which pathways (if any) were associated with BMD at six different skeletal sites across three different time points in each of the four age by ancestry cohorts in HABC?
2. Which pathways are associated with BMD at the same sites in MrOS?
3. Does the number of total genes influence the significance of the pathway-based analyses?

6.2 DATA

As described in previous Chapters, we analyzed genotype and phenotype data from four sex by ancestry cohorts in HABC and an independent group of Caucasian American males from MrOS. Briefly, for the HABC cohort, we had BMD data at 6 skeletal sites on 1653 Caucasian Americans (872 males and 781 females) and 1177 African-Americans (504 males and 673 females). We also had data on 138 candidate genes with a total of 1439 SNPs. These BMD data were residuals (after regressing out the significant covariates, see Chapter 4) and were available at three time points each of which were two years apart: baseline (time point 1), time point 3, and time point 5.

We also had data on the same BMD traits for an independent sample of Caucasian American males from MrOS. Genotype data were available on 401 candidate genes (total 4108 SNPs). All of the 1439 SNPs genotyped as part of the HABC study were also present in MrOS study, plus an additional 2662 SNPs in 262 genes.

6.3 METHOD

As stated above, the main goal of this study was to assess whether a group of candidate genes that comprise a common pathway are more likely to be associated with BMD than are a random set of genes from our gene list. In other words, are genes within a specific pathway 'enriched' for associations with specific BMD measures. To perform these analyses, we used a pathway based association analysis method (Wang et al. 2007), that is a freely provided Pearl-language based program [http://openbioinformatics.org/gengen/].

The pathway-based (or gene enrichment) analysis comprises two parts: (1) a single SNP-single trait analysis and (2) calculation of a normalized gene-enrichment statistic using a specific pathway annotation file. The single SNP-single trait analysis of both the original data and the permutations by phenotype (or case/control status) can be performed within the software package. We chose to permute our data 1000 times and estimate association using regression statistics. Alternately, other statistics (such as chi-square) obtained by any other analytical method can be used, as long as statistics are available for both the observed and permuted data. Thus this method is very flexible and can be used on many types of data and many types of analyses. The result file from the original and permutation analyses is one of the input files necessary for step 2, the pathway-based analyses. The pathway-based step requires results from the association analyses, a SNP-to-gene mapping file, and a pathway annotation file, and calculation of a normalized gene-enrichment statistic, all of which are described below.

6.3.1 Construction of the Pathway Annotation File

As described in the Introduction above, we wanted to determine whether BMD in the HABC cohort was associated with different candidate pathways. We also observed whether the size of the pathways (that is, the number of genes contained in them), influenced the results of the pathway-based analyses. We used Ingenuity (Ingenuity Systems, www.ingenuity.com) to construct the pathways.

To obtain the pathway annotation file for the HABC cohort, we first submitted all the candidate genes for which we had genotypes to the Ingenuity program. Ingenuity is a human curated pathway knowledge base. PhD level scientists extract the information from the text of papers published in several peer- reviewed journals (www.ingenuity.com). Because new information is constantly being added to the databases from which the Ingenuity program constructs pathways, we have also included the date when we derived the pathways (June 1, 2008). We used the default option of including all possible direct as well as indirect relationship between any two genes when developing the pathways. The first 11 pathways listed in Table 6.1 were constructed from the list of 138 genes for which genotypes were available on the HABC cohort. The total number of genes for each of the pathway, constructed by Ingenuity, contain more genes than are available in the HABC data set. Therefore, we list the total number of genes contained in the HABC data set, as these are the genes used in the gene-enrichment analyses. Prior to performing the pathway based analyses, we modified the default options defining the minimum number of genes within a pathway from n = 20 to n = 2 genes per pathway. Therefore, pathways 9 and 11, each of which contained only one gene from our list, were excluded from any further analysis.

We analyzed the HABC and MrOS data using the first 9 pathways and data on only 1439 SNPs (or 138 genes). We also analyzed the MrOS data using the 9 listed pathways and data on 4108 SNPs (and 283 genes) because we also wanted determine whether additional genes in the ranking list (see method description below) influenced our results.

6.3.2 Construction of SNP to Gene Correspondence File

We constructed a SNP-Gene map file, which contained the SNP name, the corresponding gene name and the distance from the gene. SNPs that do not fall within a specific gene are assigned to the gene that is nearest to that SNP. For example, rs2198843 is contained within a cluster of CYP loci and could theoretically influence expression of several CYP loci, however, the distance of the SNP from CYP1A1 gene is 1066 bp and from CYP1A2 gene is 39954 bp. Therefore, I assigned this SNP to the CYP1A1

gene. In general, the default option of the program will drop any SNP that is not within 500 kb of any gene, but all of our SNPs were within 500 kb of a gene.

We wish to note that because the user is ultimately responsible for providing the pathway annotation files and the SNP to gene mapping file, the user can define sets of genes of any sort (perhaps based on function) as well as the 'genes' or gene regions. Thus, this program is very flexible.

6.3.3 Gene-Enrichment or Pathway-based Association Analysis Statistic

After the user constructs files containing the single SNP association and permutation statistics, the pathway annotation file, and the SNP to gene mapping file, the pathway-based gene enrichment statistic is calculated as follows. Essentially, what this enrichment score measures is whether the genes in a specific pathway are nearer the top of the list of all genes that have been ordered by the association statistic of interest. For each gene the most significant statistic value (which may be a p-value, chi-squared, etc) obtained for any SNP within the gene is assigned to that gene, and is denoted r_j where $1 \leq j \leq$ N, the total number of genes. Next the enrichment score (ES) for a specific gene set S (pathway) is calculated using a weighted Kolmogorov-Smirnov-like running sum statistic:

$$ES(S) = \max_{1 \leq j \leq N} \left\{ \sum_{G_{j^*} \in S, j^* \leq j} \frac{|r_{(j^*)}|^p}{N_R} - \sum_{G_{j^*} \notin S, j^* \leq j} \frac{1}{N - N_H} \right\},$$

where N is the total number of genes; $r_{(1)},\ldots,r_{(N)}$ is the sorted statistic value from largest to smallest for the N genes, $r_{(j^*)}$ is j* th ranked statistic value. $N_R = \sum_{G_{j^*} \in S} |r_{(j^*)}|^p$. N_H is the number of genes corresponding to the gene set or pathway S; p is the parameter that gives higher weight to genes with extreme statistic values. Wang et al (Wang et al. 2007) recommend using p = 1 as a default value. As can be seen, the ES statistic depends on the number of genes, so we wanted to know if the results from our candidate gene study (with

only 138 candidate genes) would be meaningful. Thus we did compare results from the pathway-based analyses using information on the total set of 402 genes in MrOS and the subset of 138 genes (representing the HABC total set of genes).

By chance, genes that have more genotyped SNPs will have a higher probability of high statistic values than genes with fewer genotyped SNPs, and genes with more genotyped SNPs are usually larger than genes with fewer genotyped SNPs. Thus, pathways comprising more 'large genes' would have a higher ES than pathways with fewer 'large' genes. To control for the effect of "gene size" in a pathway, a series of permutations are performed to break the relationships between the phenotype and genotypes and the test statistic (in our case, an F-test for regression) is recalculated for all of the SNPs (which equals 1439 in the HABC cohort). Then an enrichment score for a specific gene set S is calculated for each of the permutations (π) and denoted by ES(S,π). In our analyses, we used 1000 permutations. Next, a normalized enrichment score (NES) is calculated as follows:

$$\frac{ES(S) - mean[ES(S,\pi)]}{SD[ES(S,\pi)]}$$

Thus, because the enrichment statistic is normalized for each gene set, the NES obtained for the different gene sets can be directly compared for a specific trait within a specific cohort. The possible significance of the pathway-based associations was assessed in several ways. First, a nominal p-value of a specific pathway is calculated as follows: The formula to calculate the nominal p-value is given by

$$\text{nominal pvalue} = \frac{number\ of\ permutations\ with\ ES(S,\pi) \geq ES(S)}{number\ of\ permutations}$$

where ES(S) is the observed enrichment score for a specific gene set S. Because the nominal p-values are calculated from the ES statistics, they are not adjusted for size of different gene sets or for tests of

multiple pathways. Two other statistics, the FDR (False Discovery Rate) and FWER (Family Wise Error Rate) are calculated using the NES, and thus control for gene size and multiple testing.

To obtain FDR, the normalized enrichment score is calculated for each of the permutations and used to obtain an empirical distribution of NES scores, defined NES(S, π) , for each of the pathways (in HABC there are 9 pathways). Then this set of normalized enrichment scores (from all permutations across all pathways) is compared with the NES of the observed data for a specific pathway, designated as NES*, to obtain the numerator of the FDR (see equation below). The denominator of the FDR is calculated by comparing NES* for a specific pathway, to the set of NES of the observed data for all of the pathways (again, for the HABC cohort there are 9 pathways). This empirical distribution is then used to calculate the FDR (False Discovery Rate) and FWER (Family Wise Error Rate) rates. If for a specific gene set, NES* is the normalized enrichment score of the observed data, then

$$FDR = \frac{\%\ of\ all\ (S,\pi)\ with\ NES(S,\pi) \geq NES*}{\%\ of\ \text{observed S}\ with\ NES(S) \geq NES*} \quad ,$$

Thus, the p-values and FDR statistics will reflect the empirical distribution of NES scores within each data set. Or to put it another way, a similar p-value obtained for the same (or different) pathways in the analyses of two different traits, might have different FDR values. As we discussed in our single SNP-single trait and single SNP-multiple trait analyses in Chapter 5, we used a liberal level of significance for FDR, and looked for consistency of results across different time points and different cohorts. Family wise error rate (FWER) is the probability that the results do not include a single false positive and this is a much more stringent condition than FDR.

6.4 RESULTS

6.4.1 Characteristics of the pathways

Using the Ingenuity program, we obtained 11 pathways derived from our group of 138 genes (Table 6.1) and the number of total genes in these pathways ranged from 2 to 28, with a median of 21 genes. However, not all of the genes in each of the Ingenuity pathways were present in our set among our genes. For example, pathway 9 and 11 as derived by Ingenuity comprised a total of 2 and 4 genes, respectively, however, only one gene in each of these pathways was present in our set of HABC candidate genes. In contrast, some genes within our set of candidate genes were present in more than 1 pathway. Of the total 138 genes, 7 were not represented in any pathways, 130 were present in one pathway and 2 were present in 2 pathways. Pathways 8 and 10 were the smallest pathways and contained 2 or 3 candidate genes. Seven of the eight remaining pathways contained 11-18 genes, and the largest pathway, pathway 1 contained 28 candidate genes. The relative coverage of each Ingenuity pathway, that is the proportion of HABC candidate genes to total genes in each pathway, ranged from 11% for pathway 8 to 100% for pathway 1 (Table 6.1).

6.4.2 Pathway-based Associations for Six BMD Traits among the HABC Cohort

We tested for associations between genes in 9 pathways (among 138 genes; see Table 6.1) for the HABC cohorts using data on baseline BMD trait residuals. All association analyses in both Chapters 5 and 6 are performed using the residuals of the BMD traits after adjusting for all significant covariates (as described in chapter 4). Pathway association results with a nominal p-value ≤ 0.05 for total hip, femoral neck, mean arm and mean leg, whole body and lumbar spine BMD measured at baseline (time point 1) are shown in Table 6.2 for all sex by ancestry groups. Pathway association results for time points 3 and 5 are presented in Tables 6.3 and 6.4, respectively.

For Caucasian American males, pathway 10 (comprising two loci) and pathway 2 (comprising 18 loci) are associated with BMD at the hip, femoral neck and lumbar spine bone sites (nominal p-value ≤ 0.05 for all). Specifically, for hip BMD, pathway 10 is nominally significant at all the three time points (p-value = 0.004, 0.005, 0.005, respectively). For femoral neck BMD, pathway 10 is significant at the baseline and time point 3 (p-value = 0.026 and 0.007), whereas lumbar spine BMD is significantly associated with pathway 2 at baseline and time point 3 (p-value = 0.022 and 0.018). However, the FDR for pathway 10 is high and ranges between 0.32 and 0.60. Both pathway 10 and pathway 2 (Figure 6.1) include the GNRHR locus and SNPs from this locus were significantly associated with hip BMD at all three time points in Caucasian American men using both single SNP-single trait and single SNP-multiple trait analyses. Pathway 2 also involves genes of TGFB family, SMAD family and it is involved in functions like gene expression, cellular function and maintenance and cellular development.

For Caucasian American females, pathway 3 is significantly associated with femoral neck BMD at baseline and time point 3 (p-value = 0 and 0.022, respectively). But no other results were consistent across traits and time points. Genes in pathway 3 are involved in skeletal and muscular disorders, cell-to-cell signaling and interaction, connective tissue development and function and contains the following genes: IGFBP family, IL1, IL6, IRAK family, TNFRSF1B and TNFRSF11B.

For African-American males, pathway 1 is significant associated with lumbar spine BMD at baseline as well as time point 5 (p-value = 0.048 and 0.02). No other pathway shows a consistent relationship with BMD at the same skeletal site across different time points. Pathway 1 comprises 28 loci, including CRH, CYP11A, HSDB17B, MEF2, and NTRK and contains loci involved in endocrine system development and function, lipid metabolism, molecular transport.

Finally, for African-American females, pathway 10 is significantly association with hip BMD at baseline (p = 0.001) and with femoral neck BMD at both baseline and time point 5 (p-value = 0.04 and 0.008). Pathway 7 is significantly associated with whole body BMD at time points 3 and 5 (p-value = 0.012 and 0.018), and also with for BMD at two peripheral sites, mean arm BMD at time point 3 (p = 0.017) and mean leg BMD at time point 5 (p = 0.044).

Cellular development, hematological system development and function, immune and lymphatic system development and function are the functions in which pathway 7 is involved. CSF, CSFR, GHRHR, FBXO32 are some of the genes contained in this pathway.

6.4.3 Exploration of Pathway-based Association Statistics

Because the FDR methodology is relatively new, we wanted to better understand how the statistics NES and FDR statistics were calculated and what their distributions looked like. As an example, we explored in more detail the results for hip BMD in Caucasian American males at time points 1 and 3 (Table 6.7). In this example, pathway 10 is nominally significant at both the time points ($p = 0.001$ and $p = 0.005$, respectively), but the corresponding FDR value is 0.349 and 0.316 respectively, which is somewhat counterintuitive. However, as described in the Methods, the nominal p-values and the FDR statistics are dependent upon the data (observed and permuted) from which they derived. Because these FDR statistics were obtained at different time points (and thus different data), the difference is not unexpected.

6.4.4 Pathway Association Analyses in MrOS – a Replicate Cohort

We also tested whether the 9 pathways derived from the HABC candidate genes were associated with BMD traits measured in the MrOS Caucasian American cohort. We performed these analyses in two different ways. First, we performed the association analyses using only genotype data on the 1439 SNPs (for 138 genes) and the nine pathways that were also available for the HABC cohort. Second, we performed the analyses using genotype data on 4108 SNPs (for 401 genes) and the nine annotated pathways. The first analysis is a 'true' replication in that identical SNPs and BMD traits are assessed. The second analysis may (or may not, see discussion below) increase our power to detect a significant association because we've increased the total number of genes that are used in the ranking statistics.

As can be seen in Table 6.5, pathway analyses of the set of 138 genes revealed pathway 10 was

associated with leg BMD (p-value 0.004), whereas pathway 1 was associated with whole body and arm BMD (p-value $\leq$ 0.05 for both). But none of the pathway-based associations with baseline BMD in Caucasian American males in HABC were replicated in the MrOS cohort. When we analyzed BMD data using the set of 393 genes but the 9 pathways, the results were similar. The same pathways were nominally associated with the same traits. Pathway 10 was associated with leg BMD (p-value 0.023), and pathway 1 was associated with whole body and arm BMD (p-value = 0.031 and 0.017 respectively).

Although the MrOS results did not replicate the HABC results for the multiple gene-single trait analyses, this outcome is consistent with our lack of replication in previous analyses comparing HABC Caucasian American men and MrOS. Although we had very strong and consistent single SNP-single trait and single SNP-multiple trait associations, as well as similar magnitude and direction of effect between GNRHR SNPs and hip BMD across all three time points, this result was not replicated in MrOS.

6.5 DISCUSSION

Overall, we consider the results of our pathway-based analyses of associations between candidate gene pathways and measures of areal BMD at different skeletal sites to be exploratory. Although we obtained some potentially intriguing results, we also recognized some limitations and developed additional questions about the usefulness of this approach for a candidate gene study.

As our results from Chapter 5 indicate, the HABC data set has a sufficient number of individuals within each sex by ancestry cohort, as well as reasonable density of SNPs for each candidate gene, to detect potentially meaningful associations using a candidate association approach. In fact, one of the most interesting results was the relationship between SNPs in the GNRHR locus and BMD at the hip and femoral neck in Caucasian American men. This association was observed using both the single SNP-single trait and the single SNP-multiple trait approaches. Thus we were encouraged when we also observed nominally significant results with hip BMD among Caucasian American men for two pathways

(pathway 2 and pathway 10) containing the GNRHR locus. We were especially encouraged because pathway 10 comprises two genes, one of which is GNRHR, and thus we would expect this pathway to show an association. However, the false discovery rate was high, probably in part because this pathway was so small in size and there are few pathways overall.

In contrast to pathway 10, pathway 2 contains 18 loci (one of which is GNRHR) and it also was associated with BMD in Caucasian American men, although this association was with lumbar spine BMD and not hip or femoral neck BMD. Because the FDR was low and the association was seen at both time points 1 and 3, this pathway should be followed up by investigating both single SNP-lumbar spine BMD association among the genes within the pathway, as well as investigating possible interactions among these genes.

The pathway-based approach (a multiple SNP-single trait association analysis) is the third of the association approaches that we used, and it also indicates that variation in the GNRHR locus influences areal BMD among Caucasian American males. Thus, the GNRHR locus might really be an important gene in bone metabolism of Caucasian American male and additional follow up study is recommended.

With the possible exception of the GNRHR result, however, we recognize several limitations to the candidate-gene pathway-based approach. The first and foremost limitation is that we are using candidate genes, and thus, unlike a whole-genome study, all of the genes are potentially related to the trait. Likewise, all of the pathways are likely related to the trait of interest. Thus, our power is probably reduced because we are looking for a true pathway association among a group of pathways (and genes) of which many are all likely to be associated *versus* looking for a true association among a large group of pathways (and genes) of which very few are likely to be associated. For example, we were looking for the best pathway out of nine pathways, or alternatively the best set of 18 genes out of 138 possible candidate genes rather than the best pathway out of 1000 pathways or the best set of 20 genes out of 20,000 genes. This relative lack of power might be one of the reasons we did not obtain consistent results for similar traits across the different time points. On the other hand, the candidate-gene pathway-based approach could be very informative if the results indicate one pathway is more associated with the trait. These

genes in this 'best among candidates' pathway could be given higher priority for additional followup studies.

We also had several questions about the pathway-based approach, some of which arose from our understanding of the NES statistic and some of which arose from our results. Although we did not perform any formal tests of hypotheses resulting from our questions and concerns, we did perform some comparisons to obtain anecdotal results that might indicate future research directions. First, we were concerned that the set of 138 genes was too small to detect consistent results. Thus, we wondered whether the total number of genes would influence the significance of the pathway-based analyses? We had only 138 candidate genes in the HABC cohort. Because we had data on more genes in the MrOS cohort, we performed analyses using the same set of 9 pathways on data from 1439 SNPs, as well as all 4108 SNPs. The results were similar for both the analyses, although the FDR was lower when more genes were included in the analyses. So we hypothesize that the number of genes does not significantly influence the pathway association, at least not the nominal significance, and if an association is present then it will be detected even with a relatively small set of genes.

Second, we wondered whether pathway size would influence the probability of detecting significant results. This question arose, in part, because of the results with pathway 10, but also because we wondered whether smaller pathways that contained more directly interacting genes would be more informative than larger pathways with fewer directly interacting genes. Thus, we constructed additional pathways using MrOS set of genes for which we obtained larger pathways than for the HABC set of genes. We also used the String program and database (Snel et al. 2000) to construct additional pathways. The String program creates pathways for protein interactions, and thus they are much smaller than the Ingenuity pathways that include interactions among DNA, RNA and proteins. We performed pathway-based analyses using all of these pathways on the MrOS data and obtained similar results to our analyses of just the nine HABC pathways. We do not present these results, because upon reflection we decided that such comparisons of pathways required more careful thought and more detailed investigations. For example, the degree of overlap among the different pathways is important, because if a causal gene is

shared by multiple pathways, the power will decrease (Wang et al. 2007). In addition, whether the pathways involve direct interactions versus indirect interactions, protein interactions, DNA interactions, etc, could also influence the outcome. Finally, not all of the genes that are in a specific constructed pathway are present in the data to be analyzed. Thus, the relative 'coverage' of the genes in a pathway might influence the result. Furthermore, if is not clear how to measure coverage. Should it be the proportion of available genes out of the total number of genes in the pathway? Or the number of directly interacting genes? Or the number of genes whose protein (or RNA products) are known to be critical points in a specific metabolic pathway?

In conclusion, the pathway-based approach using candidate genes is an intriguing addition to our suite of association analysis methods, and it may be very fruitful and indicate possible sets of genes that should be further investigated, such as the pathway containing GNRHR. However, there are many caveats for this approach, and the number of candidate genes and candidate gene pathways is likely to be important.

6.6 TABLES AND FIGURES

Table 6.1: Characteristics of Ingenuity-Derived Pathways Using HABC Gene List

Pathway	1	2	3	4	5	6	7	8	9	10	11
Pathway Name	CRH-CYP-HSDB-NTRK	GNRHR-HOXA-IGF2-TGFB-SMAD	IGF2-IRAK-IGFBP-IL1-TNFRS11B	CNTF-ADIPOR-LRP-IL-WNT	BDNF-COMT-CYP1-TNFSF	BMPR-NCOA-VDR-ESR	CSF-FBXO-SHBG	HOXA3-HOXA13	HOXA4	GNRHR	GC
Total genes	28	21	23	22	19	18	15	27	2	10	4
# Candidate genes	28	18	17	17	17	16	12	2	1	2	1
	CRH	ACVR2B	IGF2	ADIPOR1	BDNF	BMPR2	ARF3	HOXA13	HOXA4	CRH	GC
	CRHBP	CSF1R	IGFALS	CNTF	COMT	CRHR1	CSF2	HOXA3		GNRHR	
	CRHR2	GNRHR	IGFBP1	CNTFR	CYP1A1	CYP24A1	CSF2RB				
	CYP11A1	HOXA2	IGFBP2	CSK	CYP1A2	CYP27B1	CSF3R				
	CYP11B1	IGF2R	IGFBP3	FRZB	CYP1B1	ESR2	CYP17A1				
	CYP21A2	LIF	IGFBP4	IGF1R	CYP3A4	FST	CYP19A1				
	GHRH	LTBP1	IGFBP5	IL6	ESR1	GHR	FBXO32				
	GHRL	LTBP3	IGFBP6	IL6ST	GH1	GNRH2	GHRHR				
	GNRH1	SMAD1	IL1R1	LRP5	HOXA11	HOXA10	HOXA7				
	HSD11B1	SMAD2	IL1R2	LRP6	HOXA9	LHB	IGF1				
	HSD17B1	SMAD3	IL1RN	MYF5	IL1A	NCOA1	SHBG				
	HSD17B2	SMAD4	IL6R	MYF6	NFKB1	NCOA2	TNFRSF1A				
	HSD17B3	TGFB1	IRAK3	MYOD1	NR3C1	NCOA3					
	HSD17B4	TGFB2	LIFR	MYOG	NTRK1	THRA					
	HSD3B1	TGFB3	LTBP2	WNT10B	PRKAG1	TNFRSF11A					
	HSD3B2	TGFBR1	TNFRSF11B	WNT3A	TNFSF10	VDR					
	IL1B	TGFBR2	TNFRSF1B	WNT5A	TNFSF11						
	LHCGR	TGFBR3									
	MC2R										
	MEF2A										
	MEF2B										
	MEF2C										
	MEF2D										
	NTF3										
	NTRK2										
	NTRK3										
	POMC										
	STAR										

Table 6.2: Pathway Association with HABC BMD measures at baseline

Caucasian Male

BMD Trait	Pathway	Size	ES	NES	Nominal pvalue	FDR	FWER
Hip	10	2	0.951	1.482	0.001	0.349	0.324
Femoral Neck	10	2	0.887	1.216	0.025	0.369	0.585
Lumbar Spine	2	18	0.625	1.818	0.032	0.136	0.134

Caucasian Female

BMD Trait	Pathway	Size	ES	NES	Nominal pvalue	FDR	FWER
Femoral Neck	3	17	0.706	2.72	0	0.003	0.003
Mean Leg	1	28	0.525	2.014	0.02	0.082	0.08

African-American Male

BMD Trait	Pathway	Size	ES	NES	Nominal pvalue	FDR	FWER
Whole Body	2	18	0.622	1.913	0.025	0.091	0.089
Hip	1	28	0.495	1.568	0.049	0.25	0.242
Lumbar Spine	1	28	0.514	1.662	0.05	0.215	0.209
	4	17	0.583	1.596	0.015	0.132	0.251

African-American Female

BMD Trait	Pathway	Size	ES	NES	Nominal pvalue	FDR	FWER
Hip	10	2	0.951	1.482	0.001	0.349	0.324
Mean Leg	2	18	0.62	1.81	0.038	0.134	0.132
Femoral Neck	10	2	0.853	1.343	0.03	0.549	0.46
	7	12	0.601	1.268	0.041	0.331	0.531

Table 6.3: Pathway Association with HABC BMD measures at time point 3

Caucasian Male

BMD Trait	Pathway	Size	ES	NES	Nominal pvalue	FDR	FWER
Hip	10	2	0.951	1.311	0.005	0.316	0.521
Femoral Neck	10	2	0.938	1.325	0.007	0.604	0.506
Lumbar Spine	2	18	0.638	2.021	0.018	0.065	0.065
	10	2	0.894	1.343	0.023	0.266	0.446

Caucasian Female

BMD Trait	Pathway	Size	ES	NES	Nominal pvalue	FDR	FWER
Femoral Neck	7	12	0.654	1.692	0.022	0.213	0.207
	3	17	0.572	1.528	0.046	0.163	0.307
Lumbar Spine	7	12	0.636	1.449	0.024	0.348	0.318
Mean Leg	6	16	0.629	2.122	0.008	0.041	0.041

African-American Male

BMD Trait	Pathway	Size	ES	NES	Nominal pvalue	FDR	FWER
Whole Body	8	2	0.893	1.551	0.022	0.291	0.274
Mean Arm	8	2	0.84	1.377	0.049	0.571	0.473

African-American Female

BMD Trait	Pathway	Size	ES	NES	Nominal pvalue	FDR	FWER
Whole Body	7	12	0.664	1.809	0.012	0.109	0.109
Mean Arm	7	12	0.642	1.542	0.017	0.298	0.275

Table 6.4: Pathway Association with HABC BMD measures at time point 5

Caucasian Male

BMD Trait	Pathway	Size	ES	NES	Nominal pvalue	FDR	FWER
Hip	10	2	0.951	1.297	0.005	0.581	0.501
Whole Body	2	18	0.62	1.741	0.03	0.169	0.165

Caucasian Female

BMD Trait	Pathway	Size	ES	NES	Nominal pvalue	FDR	FWER
Hip	8	2	0.836	1.375	0.044	0.47	0.413
	10	2	0.836	1.276	0.047	0.322	0.534
Mean Leg	7	12	0.658	1.586	0.014	0.237	0.223

African-American Male

BMD Trait	Pathway	Size	ES	NES	Nominal pvalue	FDR	FWER
Lumbar Spine	1	28	0.529	1.982	0.02	0.083	0.082
Mean Leg	8	2	0.858	1.449	0.037	0.404	0.364
Hip	3	17	0.564	1.618	0.024	0.223	0.212
	4	17	0.54	1.273	0.043	0.332	0.528

African-American Female

BMD Trait	Pathway	Size	ES	NES	Nominal pvalue	FDR	FWER
Whole Body	7	12	0.645	1.63	0.018	0.213	0.206
Femoral Neck	10	2	0.913	1.447	0.008	0.378	0.342
Mean Leg	7	12	0.596	1.26	0.044	0.632	0.502

Table 6.5: Pathway Association with MrOS BMD measures (using only 138 MrOS genes)

MrOS

BMD Trait	Pathway	Size	ES	NES	Nominal pvalue	FDR	FWER
Leg	10	2	0.926	1.315	0.004	0.567	0.48
Whole Body	1	28	0.507	1.703	0.043	0.228	0.213
Arm	1	28	0.509	1.595	0.038	0.221	0.211

Table 6.6: Pathway Association with MrOS BMD measures (using all 393 MrOS genes)

MrOS

BMD Trait	Pathway	Size	ES	NES	Nominal pvalue	FDR	FWER
Leg	10	4	0.78	1.326	0.023	0.496	0.41
Whole Body	1	28	0.497	1.83	0.031	0.146	0.138
Arm	1	28	0.508	1.832	0.017	0.097	0.094

Table 6.7: Table with association results for all the pathways for Hip BMD in HABC data for Caucasian Males at time point 1 and time point 3

Time Point	Pathway	Size	ES	NES	Nominal pvalue	FDR	FWER
	Pathway10	2	0.951	1.482	0.001	0.349	0.324
	Pathway3	18	0.501	0.886	0.167	0.849	0.902
	Pathway4	17	0.461	0.708	0.187	0.779	0.971
tp1	Pathway2	18	0.527	0.657	0.25	0.626	0.98
	Pathway6	16	0.439	0.125	0.449	0.839	1
	Pathway1	28	0.363	-0.106	0.535	0.827	1
	Pathway5	18	0.262	-0.785	0.875	1	1
	Pathway8	3	-0.638	-0.91	0.766	0.946	1
	Pathway7	12	-0.225	-3.922	0.973	0.99	1
	Pathway6	16	0.569	1.487	0.057	0.387	0.347
tp3	Pathway10	2	0.951	1.311	0.005	0.316	0.521
	Pathway2	18	0.52	0.477	0.32	1	0.992
	Pathway3	18	0.411	0.129	0.452	1	1
	Pathway7	12	0.399	-0.04	0.569	0.948	1
	Pathway1	28	0.338	-0.342	0.646	0.937	1
	Pathway5	18	0.319	-0.376	0.69	0.819	1
	Pathway4	17	0.333	-0.429	0.699	0.737	1
	Pathway8	3	-0.768	-0.974	0.843	0.847	1

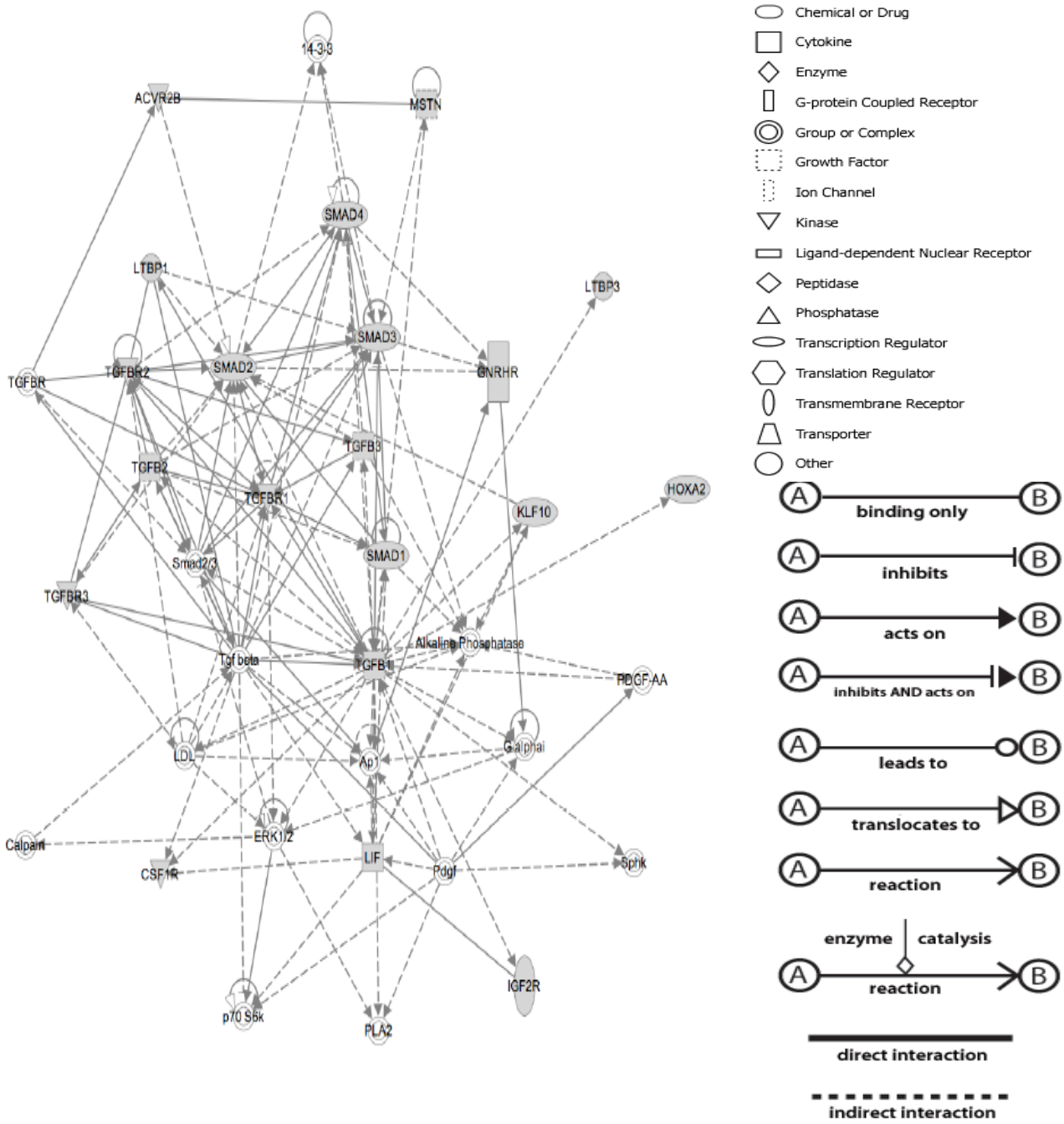


Figure 6.1: Pathway 2 (involving 18 genes from HABC list)

7.0 CONCLUSION

7.1 MOTIVATION

As the number and proportion of older individuals in the U.S. and other countries increases over the next few decades, the prevalence of many diseases associated with older age, including osteoporosis, are also projected to increase (Report of the Surgeon General, US Department of Health and Human Services, 2004). Therefore, considerable research is being done to identify the possible genes and environmental factors that influence development of osteoporosis. The long-term goal of these studies is to use our knowledge of the genes and environmental factors that influence development of osteoporosis to develop therapies that can mitigate or delay onset of bone loss.

The overall goal of my dissertation research was to use genetic association methods to identify variation in candidate genes that influenced areal BMD, a major risk factor for osteoporosis, in a healthy older population of U.S. Caucasian Americans and African Americans who are participants in the Health, Aging and Body Composition (HealthABC) Study. Areal BMD data measured at multiple skeletal sites were available at 3 time points (each two years apart) on 1653 Caucasian Americans and 1177 African Americans men and women >68 years of age at baseline. In addition, data were available on a variety of demographic, anthropometric, lifestyle, and medical history covariates, as well as genotypic data on 1439 SNPs in 138 candidate genes.

Although genetic association methods can be very powerful (Risch et al. 1996; Easton et al. 2007), the strength of these studies is dependent upon the number of SNPs being assessed and the

presence of population structure. Furthermore, a variety of association methods have been and continue to be developed and each method has different strengths and limitations.

Given the above information, we were interested in investigating the following questions:

1. How well do our SNPs cover our candidate genes?

2. Is there evidence of subpopulation structure?

3. What are the environmental factors affecting BMD levels at baseline and as well as two other time points?

4. Are any of the candidate gene SNPs associated with areal BMD at baseline and are these results replicated at other time points or in other populations?

5. Are the association results similar across the three different association methods: single SNP-single trait, single trait, single SNP-multiple traits, or multiple SNP-single trait?

There are numerous strengths to my research approach overall including (1) phenotype data available on a variety of covariates to facilitate identification of genetic signals, (2) data available on multiple BMD traits at three time points to assess internal replication, (3) genotypic data available on multiple SNPs in multiple candidate genes to increase coverage, (4) data available on four groups of reasonable sample size ($n > 500$ for African-American males and $n > 650$ for all other groups) to assess replication across groups, and (5) the use of three association methods that have complementary strengths and limitations.

There are also limitations to this study and the overall limitation is the underlying assumption that variation in areal BMD is due to common polymorphisms, rather than mulyiple rare variants. However, the proportion of variation in areal BMD due to common variation versus rare variation, is not known, so results of this study are useful. There are also strengths and limitations of our results for each of the questions above, and these are discussed in our summary of results below.

7.2 SUMMARY OF RESULTS

7.2.1 SNP coverage

To determine the coverage of our candidate genes, we first built a methodology to assess the coverage of HAPMAP phase two data in Caucasian Americans. We found that tag SNPs chosen from HAPMAP can predict all other unmeasured SNPs, including rare SNPs, at an average of 55.7% which is quite reasonable. We applied this method to our HABC data and determined that although the coverage is good in our genes (52%), we would need an additional 3 SNPs per gene to obtain coverage as good as that available in HAPMAP II. Thus, our model can be used to determine how well a specific set of SNPs covers a specific set of candidate genes. The limitation of our model is that it is less applicable to other groups whose LD structure differs from that of Caucasian Americans.

7.2.2 Assessing population ancestry

Population substructure is a concern for association studies, so we assessed ancestry using eigen analysis on uncorrelated SNPs ($r \leq 0.04$) from our set of candidate genes. Using this method, we were able to classify all individuals into two ancestry categories, including individuals who did not self-report ancestry. We also redid the eigen analysis on each ancestry group to determine additional substructure, and found no strong evidence for additional substructure in Caucasian Americans and possible substructure in African Americans. Therefore, our data on SNPs in 151 candidate genes (including both Illumina and Dr. Ferrell's lab) that were spread across the genome were sufficient to cluster individuals into two ancestry groups, however, they may not have been sufficient to detect more subtle structure. We included the first 3 principal components of ancestry from each ancestry group into our association analyses (described below) to control for more subtle subpopulation structure.

7.2.3 Estimating effects of covariates

Phenotypic variation due to effects of covariates can mask effects of candidate genes, especially if the candidate gene effects are relatively small. Because our sample sizes were large within each sex by ancestry group, we analyzed the effects of multiple demographic, anthropometric, lifestyle, and medical covariates within each group for BMD measured at multiple skeletal sites. We also analyzed the traits at three different time points both cross-sectionally as well as longitudinally. Although some covariates influenced areal BMD in a similar manner across the different sex by ancestry groups, there were differences. Some of these differences were revealed by the total proportion of variance attributable to the different covariate categories. For example, all covariates accounted for a much smaller proportion of total variation in Caucasian American males than any other group. Across all groups, the medical and anthropometric categories accounted for the largest proportion of total variation in areal BMD. Results from longitudinal analysis indicated that that time had a significant effect on all areal BMD traits, but the covariate effects were similar to those estimated from cross-sectional analysis. Residual BMD values were estimated for all BMD traits for all individuals after incorporation of all significant covariates. These residuals were used in the subsequent candidate gene association analyses. A limitation of these analyses is that all significant covariates were included. Thus, if a candidate gene has pleiotropic effects on a significant covariate, as well as areal BMD, we would also be removing effects of the candidate gene. So analysis using different sets of covariates should also be done.

7.2.4 Association analysis results

A major strength of my research was the use of three different association analysis methods: single SNP-single trait, single SNP-multiple trait, and multiple SNP-single trait. Another major strength of my research was availability of areal BMD data at three time points, which allowed assessment of internal replication. Each method has strengths and limitations, for example, all methods assumes that a common

SNP influences the trait, and that the genotyped SNPs correlate strongly with the causal SNP. The single SNP-single trait method is easily interpretable and quick, but multiple testing concerns decrease power. The single SNP-multiple trait method is more powerful than the single SNP-single trait method if the SNP influences multiple traits. Finally, the multiple SNP-single trait method is useful if multiple genes have small effects, but is less powerful if a few genes in different pathways influence a trait. Finally, all methods require follow-up procedures such as regression analyses to determine magnitude and direction of the effect, consistency across time points, BMD traits, and sex by ancestry groups. Results from all methods also require replication in another population or another means of determining biological significance, such as in vitro studies.

The association methods yielded similar results for several SNPs and the most consistent results were for Caucasian American men. Both the single SNP-single trait, and the single SNP-multiple trait analyses indicated that variation within the GNRHR gene influenced areal BMD in Caucasian American men, and this association was apparent at baseline and time point 3. One of SNPs was statistically significant by the single SNP-single trait analysis of hip BMD, and five were significant by the single SNP-multiple trait analysis. Furthermore, two pathways showing nominal significance contained the GNRHR gene. Followup analyses revealed that these 5 SNPs had consistent effects at all time points and haplotype analysis indicate that these 5 SNP are in high LD, and comprise two significant haplotypes. However when we tried to replicate our results in another Caucasian American male population, we were unsuccessful. Thus, we conclude that GNRHR might have a role in bone metabolism of older Caucasian American males, but more research is needed.

We also obtained potentially interesting results in the other sex by ancestry groups, such as a significant association between areal BMD and a single SNP in each of the TNFRSF11B and CSF2 genes. Although some of the associations are highly significant at one time point using a specific association method, they were not highly significant for all the methods and/or time points. None of the candidate gene associations were consistent across the different sex by ancestry groups; however, this is

not unexpected given the previous reports of differences in BMD between the sexes and between different ancestral groups.

The multiple SNP-single trait analyses yielded some potentially interesting results, which could be followed up by doing single additional single SNP analysis, as well as some gene by gene interaction analyses, especially for the GNRHR pathway. A strength of this approach is that it could indicate a subset of genes acting within the same pathway, that could be followed up by performing gene by gene interaction analyses using only this subset of genes. However, as we state in chapter 6, there are several limitations of this method, especially when used with data on candidate genes.

7.3 SIGNIFICANCE OF THE STUDY

Osteoporosis is one of the major public health concerns, in the United States as well as other countries. So knowledge of the possible genetic as well as environmental effects affecting the BMD status and simultaneously the chance of developing osteoporosis is essential. Use of multiple analytical methods on data from multiple traits measured at >1 time point, as well as data on multiple SNPS within traits, can greatly facilitate interpretation of results and possibly identification of genes or biological pathways that influence BMD, a risk factor for osteoporosis. Knowledge of such pathways may lead to additional methods of treatment or cure.

7.4 FUTURE WORK

One of the main areas for future research involves the analyses of BMD data using a different set of covariates. We performed our association studies using residuals from a set of all significant covariates. Because genes influencing BMD are likely to influence other measures of body composition,

removing the effects of such covariates, like weight or presence of osteoporosis by T-score, would reduce our ability to detect such candidate gene effects. In other words, we may be removing, prior to the association analyses, the genetic variation that we are interested in. Thus, additional studies should be performed using different sets of candidate genes, and compare and contrast the results across time points, as well as across the sex by ancestry groups.

Our most exciting result was the association between SNPs at the GNRHR locus and hip BMD in Caucasian American males. However, although we observed a significant association between these SNPs (and haplotypes) in the MrOS population, the direction of the effect was the reverse. This result may imply that the result in HABC is false positive or that the measured SNPs may not be causal, but are in LD with the causal SNP or SNPs. Furthermore, there is the possibility that the potentially causal SNP (allele) might be present on opposite haplotypes for MrOS and HABC Caucasian American males. We need to further investigate this by investigating additional SNPs in GNRHR, both rare and common, and typing them in our study and other studies.

We developed a CART model of SNP coverage for Caucasian Americans. However, additional models should be developed for populations of African and Chinese or Japanese descent because LD structure differs between different ancestry groups. Such a model could be developed by using a database (such as ENCODE) that has complete SNP data on individuals with an African (or Chinese) ancestry.

Finally, additional research should be done on the pathway based analysis approach, especially as regards to use on data from candidate genes. Further investigation of the effects of pathway size (numbers of genes), as well as definition of pathways (for example, direct or indirect interactions), should be done.

Other areas of research include investigation of gene by gene interactions (as obtained from the pathway-based approach) and use of the pathway-based approach on principal components of phenotypes (perhaps a combination of the pathway-based and PCH approaches).

APPENDIX

LIST OF ALL SNPS, COVARIATES AND SIGNIFICANT SNPS BY GROUPS

Table A 1: Association of Covariates with the other nine BMD traits at baseline for four ancestry by sex groups.

GROUP	African-American Male								
BMD Traits	Head	Left Arm	Left Leg	Left Rib	Thoracic Spine	Pelvic	Ward's Triangle	Intertrochanteric	Mean Rib
Covariates									
Population Substructure									
PC1	1.52E-01	1.03E-01	-7.79E-02	1.43E-01	1.73E-01	-4.38E-02	2.15E-01	4.08E-02	8.04E-02
PC2	4.68E-01	1.20E-01	3.25E-01	2.47E-01	4.51E-01	4.60E-01	1.98E-01	3.81E-01	2.53E-01
PC3	9.94E-01	5.05E-02	6.84E-02	6.19E-02	1.36E-02	-3.04E-01	-1.11E-01	-2.39E-02	-1.92E-02
Demographics									
Age	9.70E-03	-1.10E-03	2.10E-03	-1.20E-03	2.80E-03	1.00E-04	2.10E-03	4.00E-03	-5.00E-04
Site	6.73E-02	1.74E-02 **	5.85E-02 ****	2.26E-02 ***	-2.55E-02	3.70E-02 **	3.42E-02 **	7.00E-03	2.39E-02 ***
Anthropometrics									
Height	1.00E-04	0.00E+00	0.00E+00	0.00E+00	-3.00E-04	-2.00E-04	-4.00E-04 ****	-1.00E-04	0.00E+00
Weight	8.40E-03 ***	2.60E-03 ****	4.10E-03 ****	2.90E-03 ****	7.30E-03 ****	8.20E-03 ****	6.30E-03 ****	4.10E-03 ****	3.00E-03 ****
Waist Circumference	-9.40E-03 ****	-1.60E-03 ***	-2.40E-03 ***	-2.40E-03 ****	-4.80E-03 ****	-2.00E-03	-1.30E-03	-1.40E-03	-2.20E-03 ****
Life style									
Drinking History	-4.00E-04	1.10E-03	5.60E-03	5.00E-03	-5.80E-03	-8.00E-04	-4.00E-03	-4.10E-03	5.00E-03
Smoking History	-3.16E-02	-1.90E-03	-8.10E-03	-1.00E-02	-1.38E-02	-8.50E-03	-1.70E-03	-6.40E-03	-7.70E-03
Exercise	2.00E-04	0.00E+00	1.00E-04	0.00E+00	-1.00E-04	1.00E-04	1.00E-04	1.00E-04	0.00E+00
Medical Conditions									
Prevalent CHD	-6.50E-03	5.80E-03	9.40E-03	-2.00E-04	1.48E-02	2.40E-03	1.79E-02	2.80E-03	2.50E-03
Prevalent CVD	-3.00E-02	1.95E-02	3.14E-02	-2.00E-04	1.10E-03	1.55E-02	2.70E-03	-1.15E-02	8.60E-03
Prevalent Diabetes	3.64E-02	-3.70E-03	1.10E-02	1.79E-02	3.10E-02	1.08E-02	2.57E-02	1.47E-02	1.73E-02
Prevalent Osteoarthritis, knee	1.10E-02	4.60E-03	6.20E-03	1.06E-02	3.21E-02	8.90E-03	-2.00E-03	4.00E-03	1.16E-02
Prevalent Osteoarthritis, hip	-7.30E-03	-1.52E-02	-2.15E-02	-1.00E-02	-4.53E-02	-4.82E-02	-3.88E-02	-4.98E-02 **	-1.16E-02
Prevalent HBP	-1.56E-02	-2.70E-03	-6.60E-03	-1.10E-03	-1.27E-02	-1.11E-02	-7.50E-03	-7.00E-03	-3.80E-03
Prevalent Cancer	-2.79E-02	-2.52E-02 **	-3.41E-02 **	-2.14E-02	-7.70E-03	-3.33E-02	-3.33E-02	-2.40E-02	-2.02E-02
Prevalent Clinical Osteoporosis	1.29E-01	8.00E-03	3.01E-02	2.52E-02	3.26E-02	1.51E-02	3.88E-02	7.01E-02	2.20E-02
Osteopporosis by T-score	-3.17E-01 ****	-8.92E-02 ****	-1.51E-01 ****	-9.98E-02 ****	-1.57E-01 ****	-2.16E-01 ****	-2.96E-01 ****	-2.02E-01 ****	-9.11E-02 ****
Any Osteoporosis Drugs	-3.34E-01	-5.07E-02	-6.87E-02	-7.33E-02	-1.40E-01	-1.15E-01	-1.61E-01	-2.51E-01 **	-6.79E-02
Calcium Supplement	-1.36E-02	-1.43E-02	-2.05E-02	-4.44E-02	-6.22E-02	-1.72E-02	-4.60E-03	-3.79E-02	-4.52E-02
Vitamin D Supplement	9.88E-02	-4.20E-03	1.03E-02	1.47E-02	4.34E-02	-4.57E-02	-1.61E-02	-3.06E-02	2.52E-02
Oral Estrogen	-1.00E+00 ***	-1.24E-01	-2.71E-01 **	-1.47E-01	-2.91E-01	-1.85E-01	-1.37E-01	-2.03E-01	-1.88E-01 **

Table A 1 continued

GROUP	African-American Female								
BMD Traits	Head	Left Arm	Left Leg	Left Rib	Thoracic Spine	Pelvic	Ward's Triangle	Intertrochanteric	Mean Rib
Covariates									
Population Substructure									
PC1	9.68E-01	2.08E-01 **	2.10E-01	2.24E-01 **	4.82E-01 **	5.80E-01 ***	6.59E-01 ***	6.57E-01 ***	2.82E-01 ***
PC2	2.87E-01	2.00E-01 **	1.63E-01	5.70E-02	1.14E-01	2.02E-01	1.37E-01	-2.56E-01	5.88E-02
PC3	-3.06E-01	-3.73E-02	-1.58E-01	-1.46E-01	-1.49E-01	-3.55E-01	-1.30E-01	-8.83E-02	-1.24E-01
Demographics									
Age	5.80E-03	-1.80E-03 **	-2.30E-03	-8.00E-04	-1.00E-03	-4.00E-03	-3.50E-03	-1.20E-03	-6.00E-04
Site	-1.34E-02	1.05E-02	3.81E-02 ****	1.85E-02 ***	-3.12E-02 ***	2.14E-02	2.40E-02	-1.13E-02	2.80E-02 ****
Anthropometrics									
Height	3.00E-04	2.00E-04 ****	1.00E-04	1.00E-04	1.00E-04	2.00E-04	-2.00E-04 **	-2.00E-04	1.00E-04
Weight	5.80E-03 ***	1.80E-03 ****	2.60E-03 ****	2.30E-03 ****	3.50E-03 ****	2.50E-03 ****	5.20E-03 ****	3.60E-03 ****	2.30E-03 ****
Waist Circumference	-1.00E-03	-7.00E-04 **	-1.00E-04	-7.00E-04 **	-1.30E-03 **	6.00E-04	-3.00E-04	-1.00E-04	-7.00E-04 **
Life style									
Smoking History	0.00E+00	1.00E-04 ****	1.00E-04 **	1.00E-04	0.00E+00	1.00E-04	1.00E-04	1.00E-04	1.00E-04
Drinking History	2.10E-03	-2.10E-03	-2.30E-03	-1.30E-03	-9.50E-03	-1.00E-03	2.00E-04	3.60E-03	-1.20E-03
Exercise	-2.60E-02	0.00E+00	3.90E-03	3.50E-03	2.60E-03	2.00E-04	4.60E-03	1.00E-03	4.60E-03
Medical Conditions									
Prevalent CVD	2.32E-02	-9.70E-03	-2.86E-02	-9.30E-03	-3.32E-02	-3.17E-02	-4.11E-02 **	-1.73E-02	-5.30E-03
Prevalent Diabetes	6.74E-02	2.32E-02 ****	4.17E-02 ****	1.13E-02	1.21E-02	2.92E-02	4.84E-02 ***	2.70E-02	1.15E-02
Prevalent Osteoarthritis, hip	2.46E-02	1.86E-02 ***	2.99E-02 **	1.83E-02 ***	1.08E-02	3.42E-02 **	3.60E-02 **	3.98E-02 ***	1.84E-02 ***
Prevalent HBP	-1.43E-02	4.40E-03	-4.00E-03	1.27E-02 ***	1.46E-02	9.00E-04	-4.00E-04	-4.00E-03	8.70E-03 **
Osteopporosis by T-score	-2.70E-01 ****	-6.14E-02 ****	-1.54E-01 ****	-3.31E-02 ***	-1.15E-01 ****	-1.91E-01 ****	-2.37E-01 ****	-1.58E-01 ****	-3.95E-02 ****
Calcium Supplement	6.77E-02	3.70E-03	6.00E-03	2.48E-02 ***	1.57E-02	5.00E-03	2.90E-03	-5.90E-03	1.97E-02 **
Vitamin D Supplement	1.95E-01 **	1.52E-02	5.17E-02 **	7.70E-03	4.89E-02	7.72E-02 **	4.09E-02	6.24E-02 **	1.49E-02 *

Table A 1 continued

GROUP	Caucasian Male								
BMD Traits	Head	Left Arm	Left Leg	Left Rib	Thoracic Spine	Pelvic	Ward's Triangle	Intertrochanteric	Mean Rib
Covariates									
Population Substructure									
PC1	-4.54E-01	5.14E-02	-1.26E-01	-5.18E-02	7.38E-02	-1.21E-01	3.35E-02	-5.01E-02	-5.30E-02
PC2	-2.16E-01	6.12E-02	1.18E-01	-1.29E-01	-1.95E-01	1.10E-01	1.52E-01	5.64E-02	-7.70E-02
PC3	1.17E-01	-6.28E-02	-2.04E-02	-8.31E-02	1.35E-01	-1.37E-01	9.73E-02	-1.13E-01	-9.71E-02
Demographics									
Age	2.44E-02	1.02E-03	4.79E-02 ****	2.39E-02 ****	-2.10E-02 **	1.24E-02	3.95E-02 ****	-1.78E-03	2.54E-02 ****
Site	-3.41E-03	-2.05E-03 **	-4.52E-03 ***	-1.79E-03	9.98E-04	-4.28E-03 **	-2.79E-03	-2.97E-03	-1.99E-03 **
Anthropometrics									
Height	8.65E-04 ****	1.66E-04 ****	1.80E-04 **	1.16E-04 **	8.89E-07	-1.37E-06	-1.87E-04	4.60E-05	7.03E-05
Weight	2.85E-03	1.68E-03 ****	3.85E-03 ****	2.04E-03 ****	4.62E-03 ****	5.55E-03 ****	7.29E-03 ****	3.21E-03 ****	2.36E-03 ****
Waist Circumference	-3.12E-03 **	-1.05E-03 ****	-2.07E-03 ****	-1.62E-03 ****	-1.90E-03 ***	-1.24E-03	-2.02E-03 ***	-1.07E-03	-1.68E-03 ****
Life style									
Education	3.24E-02	-8.61E-03 **	2.68E-03	-7.50E-03	-4.83E-03	1.08E-02	2.22E-03	-5.41E-03	-9.12E-03 **
Exercise	-1.64E-04	3.03E-05	-3.14E-05	-5.52E-06	-8.59E-05	1.34E-05	6.90E-05	9.62E-06	1.33E-06
Smoking History	-1.23E-02	3.14E-03	1.02E-03	6.92E-04	-1.14E-03	4.61E-03	6.55E-03	4.99E-03	-6.95E-04
Medical Conditions									
Prevalent CHD	2.59E-02	5.71E-03	2.25E-02 ***	4.05E-03	2.92E-02 ***	5.74E-03	9.67E-03	1.66E-02	5.79E-03
Prevalent Diabetes	-2.66E-02	-1.92E-03	2.26E-02	8.27E-03	4.54E-02 ***	2.26E-02	2.18E-02	-4.21E-03	5.35E-03
Prevalent Osteoarthritis, knee	3.02E-02	4.92E-03	3.61E-03	3.20E-03	1.05E-02	1.98E-02	1.18E-02	1.85E-02	3.82E-03
Prevalent Osteoarthritis, hip	1.02E-02	9.18E-04	-5.89E-03	4.74E-03	3.24E-03	-1.14E-02	-1.71E-02	-1.66E-02	2.54E-03
Prevalent HBP	1.91E-02	1.33E-03	-5.82E-03	3.44E-03	1.77E-02 **	2.28E-03	4.80E-03	4.73E-03	3.97E-03
Osteopporosis by T-score	-3.27E-01 ***	-1.17E-01 ****	-2.00E-01 ****	-7.06E-02 ***	-1.09E-01 **	-2.15E-01 ****	-2.76E-01 ****	-1.61E-01 ****	-7.22E-02 ***
Calcium Supplement	9.79E-02 **	1.39E-02	4.59E-02 ***	1.76E-02	1.50E-02	4.08E-02	2.58E-02	1.56E-02	1.68E-02

Table A 1 continued

GROUP	Caucasian Female								
BMD Traits	Head	Left Arm	Left Leg	Left Rib	Thoracic Spine	Pelvic	Ward's Triangle	Intertrochanteric	Mean Rib
Covariates									
Population Substructure									
PC1	-3.09E-02	-5.39E-02	-1.03E-01	-1.98E-02	1.87E-02	1.07E-01	1.14E-01	-6.25E-02	-1.32E-01
PC2	-1.95E-01	-8.19E-02	-3.43E-02	-5.00E-02	3.52E-02	-1.32E-01	5.64E-03	7.16E-02	-4.37E-02
PC3	-5.60E-01	-3.64E-02	-1.01E-01	-4.07E-02	-7.33E-02	5.47E-02	-9.86E-02	-7.59E-02	1.78E-02
Demographics									
Age	-2.31E-02	-3.65E-03	2.21E-02 ***	8.79E-03	-2.89E-02 ***	-5.84E-03	3.48E-03	-1.06E-02	2.01E-02 ****
Site	5.15E-03	2.53E-04	-1.59E-03	-2.68E-04	3.13E-03 **	-3.11E-03	-1.43E-03	-3.25E-04	-9.08E-04
Anthropometrics									
Height	8.86E-04 ****	1.66E-04 ****	1.78E-04 ***	6.83E-05	1.88E-04 **	2.69E-04 ***	-1.09E-04	7.76E-05	7.15E-05
Weight	5.08E-03 **	1.08E-03 ****	2.39E-03 ****	1.69E-03 ****	1.81E-03 ***	2.82E-03 ****	4.43E-03 ****	2.34E-03 ****	1.45E-03 ****
Life style									
Drinking History	1.49E-02	-1.48E-03	9.82E-04	4.49E-04	2.61E-03	9.25E-03	7.62E-03	9.79E-03	-3.41E-04
Smoking History	-1.47E-02	4.86E-03	4.03E-03	1.90E-04	9.14E-03	5.63E-03	-6.18E-04	3.22E-03	1.25E-03
Exercise	-7.34E-04	-7.27E-05	4.14E-04	-8.38E-05	-4.26E-05	7.83E-05	3.86E-04	-1.59E-04	-4.24E-05
Medical Conditions									
Prevalent Diabetes	2.50E-02	1.30E-02	2.85E-02 **	2.30E-02 **	4.63E-02 ***	5.87E-02 ****	3.49E-02 **	1.36E-03	1.88E-02 **
Prevalent Osteoarthritis, knee	-9.47E-03	7.75E-03 **	1.50E-02 **	5.78E-03	4.64E-03	1.46E-02	1.11E-02	1.60E-02 **	7.39E-03
Prevalent Osteoarthritis, hip	5.42E-03	-8.34E-03 **	-1.46E-02 **	-7.42E-03	-5.44E-03	-9.39E-03	-1.16E-02	-1.72E-02 **	-9.64E-03 **
Prevalent Cancer	3.74E-02	2.06E-03	1.75E-02 **	2.29E-02 ****	1.71E-02	2.80E-02 ***	1.54E-02	2.42E-02 **	2.07E-02 ****
Prevalent Clinical Osteoporosis	-2.81E-02	-1.99E-03	-7.52E-03	3.94E-03	-3.14E-04	-1.51E-02	-8.06E-03	-1.42E-02	1.93E-03
Osteopporosis by T-score	-2.40E-01 ****	-5.47E-02 ****	-1.11E-01 ****	-3.71E-02 ****	-7.57E-02 ****	-1.50E-01 ****	-2.18E-01 ****	-1.41E-01 ****	-3.63E-02 ****
Calcium Supplement	1.71E-02	-7.33E-03	-7.86E-03	3.24E-03	-1.75E-02	-8.82E-03	-2.15E-02	-1.57E-02	6.02E-03
Vitamin D Supplement	-2.70E-02	3.05E-03	2.01E-03	-2.86E-03	1.83E-02	9.94E-03	1.43E-02	2.53E-02 **	-1.74E-03
Oral Estrogen	2.10E-01 ****	3.43E-02 ****	5.06E-02 ****	4.20E-02 ****	6.16E-02 ****	6.12E-02 ****	4.02E-02 ****	3.94E-02 ****	4.26E-02 ****

Table A 2: Table with top 50 SNPs (according to p-value) for single trait single SNP analysis for the four ancestry by sex groups

A.2A African American Male

Hip			Femoral Neck			Whole Body		
SNP	Gene	$(-1)*LOG_{10}P$	SNP	Gene	$(-1)*LOG_{10}P$	SNP	Gene	$(-1)*LOG_{10}P$
rs1570070	IGF2R	4.33	rs817531	LTBP1	3.45	rs1570070	IGF2R	3.75
rs218204	LTBP1	2.65	rs1570070	IGF2R	3.10	rs8124792	CYP24A1	2.76
rs10473100	LIFR	2.61	rs2190242	CRHR2	2.74	rs3917254	IL1R1	2.30
rs442923	HSD17B4	2.52	rs1553009	IGFBP2	2.64	rs2304707	LTBP2	2.29
rs284142	TGFBR3	2.47	rs2289737	SMAD1	2.59	rs1543403	ESR1	2.25
rs2734554	MYOD1	2.43	rs11950140	CRHBP	2.59	rs2282014	GNRH2	2.24
rs10482750	TGFB2	2.28	rs10473100	LIFR	2.43	rs949963	IL1R1	2.18
rs2275608	HSD3B2	2.24	rs4242632	NTRK2	2.30	rs284142	TGFBR3	2.17
rs11638442	CYP11A1	2.11	rs1545550	LTBP1	2.25	rs7148018	LTBP2	2.16
rs1545550	LTBP1	2.07	rs1491851	BDNF	2.20	rs1151503	LTBP3	2.11
rs2026001	HSD17B3	2.05	rs2734554	MYOD1	2.15	rs7384927	GHRHR	2.09
rs2069824	IL6	2.02	rs442923	HSD17B4	2.14	rs3802428	CNTFR	2.07
rs284876	TGFBR3	2.02	rs284876	TGFBR3	2.12	rs2287047	IL1R1	2.06
rs6336	NTRK1	1.96	rs10482750	TGFB2	2.12	rs7156293	TGFB3	2.02
rs4809960	CYP24A1	1.96	rs10780695	NTRK2	2.09	rs948562	CNTF	2.02
rs2289737	SMAD1	1.95	rs2282014	GNRH2	2.09	rs910416	ESR1	2.01
rs2279356	CYP11A1	1.92	rs10772530	LRP6	2.00	rs10482750	TGFB2	1.94
rs817531	LTBP1	1.87	rs7243853	MC2R	2.00	rs6336	NTRK1	1.93
rs3784434	NTRK3	1.80	rs8191044	HSD17B2	1.99	rs8191772	IGF2R	1.92
rs910416	ESR1	1.80	rs4851543	IL1R1	1.97	rs9340835	ESR1	1.91
rs25882	CSF2	1.80	rs25882	CSF2	1.88	rs1982072	TGFB1	1.91
rs4319975	LHCGR	1.78	rs8191844	IGF2R	1.81	rs634918	LRP5	1.85
rs7973583	LRP6	1.76	rs3731575	LTBP1	1.81	rs6259	SHBG	1.83
rs12569286	MEF2D	1.75	rs2239393	COMT	1.79	rs3784434	NTRK3	1.82
rs2033785	SMAD3	1.74	rs2198843	CYP1A1	1.78	rs4844486	HSD11B1	1.77
rs1800796	IL6	1.71	rs815541	WNT5A	1.77	rs2734554	MYOD1	1.76
rs8191772	IGF2R	1.70	rs186233	MEF2C	1.74	rs2302114	LTBP2	1.74
rs3731575	LTBP1	1.69	rs6580928	IGFBP6	1.70	rs1019731	IGF1	1.72
rs235219	TNFRSF1B	1.66	rs2513924	TIEG	1.69	rs10868232	NTRK2	1.72
rs2296621	TGFBR3	1.65	rs1804506	TGFBR3	1.69	rs10501087	BDNF	1.71
rs11950140	CRHBP	1.65	rs8191772	IGF2R	1.67	rs1317681	TGFB2	1.70
rs4605279	TGFB1	1.64	rs2069824	IL6	1.62	rs2279356	CYP11A1	1.69
rs11265622	IL6R	1.63	rs10501087	BDNF	1.60	rs879131	NTRK3	1.69
rs165849	COMT	1.61	rs705117	GC	1.59	rs6455678	IGF2R	1.64
rs1491851	BDNF	1.60	rs4601790	LTBP3	1.56	rs11638442	CYP11A1	1.64
rs1543403	ESR1	1.60	rs2189480	VDR	1.55	rs2513924	TIEG	1.64
rs4675278	BMPR2	1.59	rs218204	LTBP1	1.53	rs2036138	SMAD1	1.60
rs506611	WNT5A	1.59	rs506611	WNT5A	1.51	rs3917332	IL1R1	1.60
rs3917332	IL1R1	1.58	rs10849275	NTF3	1.51	rs3769550	LTBP1	1.58
rs4601790	LTBP3	1.58	rs11265622	IL6R	1.51	rs1795651	WNT5A	1.58
rs1553009	IGFBP2	1.58	rs2372866	IGFBP5	1.51	rs6537355	SMAD1	1.58
rs2073475	CYP11A1	1.58	rs11063703	NTF3	1.48	rs6472528	NCOA2	1.57
rs2190242	CRHR2	1.57	rs2267153	LIF	1.48	rs4846479	TGFB2	1.56
rs7757956	ESR1	1.57	rs8026319	CYP11A1	1.48	rs2285724	HOXA13	1.55
rs6580928	IGFBP6	1.56	rs3729961	IL6ST	1.47	rs2280919	FBXO32	1.55
rs3917254	IL1R1	1.54	rs7148018	LTBP2	1.47	rs10473100	LIFR	1.55
rs10849275	NTF3	1.54	rs2033178	IGF1	1.47	rs2270376	STAR	1.54
rs9457827	IGF2R	1.53	rs1900173	IL6ST	1.47	rs1465693	CSF1R	1.52
rs8124792	CYP24A1	1.53	rs2073475	CYP11A1	1.46	rs1866179	SMAD1	1.51
rs1385499	TNFRSF11B	1.52	rs8025158	NTRK3	1.45	rs1553009	IGFBP2	1.51

A.2A continued

Mean Arm			Mean Leg			Lumbar Spine		
SNP	Gene	$(-1)*LOG_{10}P$	SNP	Gene	$(-1)*LOG_{10}P$	SNP	Gene	$(-1)*LOG_{10}P$
rs1570070	IGF2R	3.92	rs1570070	IGF2R	2.99	rs10940495	IL6ST	3.32
rs9340835	ESR1	3.75	rs10503771	GNRH1	2.78	rs2270376	STAR	2.78
rs2270376	STAR	2.91	rs7384927	GHRHR	2.65	rs10482750	TGFB2	2.66
rs8124792	CYP24A1	2.82	rs2734554	MYOD1	2.39	rs9340835	ESR1	2.53
rs2304707	LTBP2	2.69	rs4846100	TNFRSF1B	2.31	rs32662	HSD17B4	2.52
rs7384927	GHRHR	2.67	rs2033178	IGF1	2.26	rs2164808	POMC	2.47
rs2881766	ESR1	2.56	rs284142	TGFBR3	2.22	rs4319975	LHCGR	2.42
rs12102144	NTRK3	2.52	rs1016792	SMAD1	2.20	rs2050093	GHRH	2.37
rs3769550	LTBP1	2.42	rs2033785	SMAD3	2.19	rs165815	COMT	2.33
rs2305502	LTBP1	2.34	rs11950140	CRHBP	2.14	rs1317681	TGFB2	2.33
rs3825709	LTBP2	2.23	rs1543403	ESR1	2.00	rs2229238	IL6R	2.21
rs939348	THRA	2.20	rs4738070	NCOA2	1.95	rs1893510	TNFRSF11A	2.17
rs7148018	LTBP2	2.17	rs6518682	LIF	1.95	rs218204	LTBP1	2.16
rs11638442	CYP11A1	2.17	rs1019731	IGF1	1.87	rs938671	HSD17B1	2.14
rs879131	NTRK3	2.15	rs10868232	NTRK2	1.84	rs2446419	CYP19A1	2.08
rs3130062	TNFA	2.13	rs8070454	G-CSF	1.81	rs634918	LRP5	2.06
rs8191772	IGF2R	2.12	rs939348	THRA	1.80	rs4757569	MYOD1	2.02
rs2294891	NCOA3	2.07	rs2189239	HOXA13	1.78	rs1143650	HSD17B4	2.01
rs4844486	HSD11B1	2.01	rs2304707	LTBP2	1.78	rs2513924	TIEG	1.99
rs1867348	IGF2R	1.99	rs1042658	G-CSF	1.77	rs1978346	IGFBP5	1.99
rs4531631	TNFSF11	1.95	rs2287047	IL1R1	1.77	rs2041693	TNFSF10	1.97
rs1364286	HSD17B2	1.89	rs3850201	HSD17B4	1.73	rs1543403	ESR1	1.93
rs2235543	HSD11B1	1.88	rs7688268	SMAD1	1.73	rs6032470	GHRH	1.81
rs3755137	IGFBP5	1.84	rs2711158	GNRHR	1.72	rs131842	CSF2RB	1.79
rs284142	TGFBR3	1.81	rs17145454	HSD17B4	1.71	rs1641536	SHBG	1.79
rs3784434	NTRK3	1.78	rs7166081	SMAD3	1.70	rs8191844	IGF2R	1.77
rs10501087	BDNF	1.77	rs910416	ESR1	1.70	rs2560718	HSD17B4	1.76
rs8070454	G-CSF	1.76	rs4569454	LTBP1	1.68	rs948562	CNTF	1.75
rs3853248	ESR1	1.75	rs11638442	CYP11A1	1.67	rs3784434	NTRK3	1.73
rs6865659	CSF1R	1.73	rs9340835	ESR1	1.65	rs2296621	TGFBR3	1.72
rs2015189	LTBP1	1.72	rs948562	CNTF	1.65	rs1884052	ESR1	1.72
rs1543403	ESR1	1.69	rs3755137	IGFBP5	1.63	rs6934383	IGF2R	1.71
rs12273363	BDNF	1.68	rs284876	TGFBR3	1.62	rs11265622	IL6R	1.69
rs1012157	COMT	1.62	rs1364286	HSD17B2	1.61	rs2294090	FBXO32	1.68
rs1465693	CSF1R	1.62	rs4818	COMT	1.57	rs2417086	LRP6	1.67
rs1042658	G-CSF	1.62	rs2070889	LIF	1.53	rs312018	LRP5	1.66
rs255126	CRHR2	1.61	rs2200287	TNFSF11	1.52	rs910416	ESR1	1.65
rs1568400	THRA	1.60	rs727479	CYP19A1	1.51	rs12313200	LRP6	1.65
rs2734554	MYOD1	1.60	rs10482672	NR3C1	1.49	rs1570070	IGF2R	1.62
rs4723025	GHRHR	1.60	rs10746750	NTRK2	1.49	rs4477532	LRP6	1.58
rs871335	IGF1R	1.60	rs1046013	NCOA2	1.49	rs6336	NTRK1	1.54
rs1982072	TGFB1	1.58	rs11740179	HSD17B4	1.49	rs4988492	GHRH	1.53
rs2033785	SMAD3	1.57	rs3892225	TGFB2	1.49	rs4952277	LTBP1	1.52
rs1424151	HSD17B2	1.56	rs258747	NR3C1	1.47	rs2282014	GNRH2	1.51
rs4569454	LTBP1	1.53	rs1530363	IGFBP4	1.45	rs4756914	MYOD1	1.50
rs634918	LRP5	1.52	rs876688	TGFBR2	1.45	rs4734648	TIEG	1.49
rs4601790	LTBP3	1.52	rs3784434	NTRK3	1.44	rs1171549	MEF2D	1.49
rs12424162	NTF3	1.49	rs1012157	COMT	1.42	rs6865659	CSF1R	1.47
rs4792887	CRHR1	1.48	rs1151503	LTBP3	1.42	rs3802428	CNTFR	1.47
rs11950140	CRHBP	1.47	rs13004818	NCOA1	1.40	rs10868232	NTRK2	1.47

A.2B. African American female

	Hip			Femoral Neck			Whole Body	
SNP	Gene	(-1)*LOG$_{10}$P	SNP	Gene	(-1)*LOG$_{10}$P	SNP	Gene	(-1)*LOG$_{10}$P
rs7757956	ESR1	3.14	rs31473	CSF2	4.03	rs7502966	THRA	3.07
rs2280699	IGFBP6	2.84	rs7816339	FBXO32	2.57	rs4851527	IL1R2	2.98
rs3796720	GNRHR	2.70	rs3735529	HOXA6	2.54	rs312018	LRP5	2.85
rs31473	CSF2	2.68	rs1889313	IL6R	2.47	rs3735529	HOXA6	2.82
rs3850651	MEF2C	2.64	rs1003737	IGF2R	2.39	rs2765880	TGFBR3	2.79
rs4851527	IL1R2	2.43	rs7757956	ESR1	2.38	rs2294090	FBXO32	2.77
rs2227321	G-CSF	2.29	rs2799090	TGFB2	2.32	rs11899433	LTBP1	2.66
rs3133287	TIEG	2.25	rs12424162	NTF3	2.30	rs2513924	TIEG	2.30
rs2799090	TGFB2	2.16	rs216124	CSF1R	2.21	rs7757956	ESR1	2.29
rs2513924	TIEG	2.10	rs3796720	GNRHR	2.21	rs315952	IL1RN	2.28
rs2562802	GDF8	2.09	rs1867348	IGF2R	2.17	rs934328	TGFBR2	2.27
rs12424162	NTF3	2.00	rs7149264	TGFB3	2.16	rs315951	IL1RN	2.21
rs2765880	TGFBR3	1.99	rs25640	HSD17B4	2.13	rs2269457	THRA	2.15
rs12522630	MEF2C	1.95	rs4648068	NFKB1	2.09	rs4149577	TNFRSF1A	2.14
rs2662405	CSF2	1.95	rs1030021	IL1R1	2.04	rs1867348	IGF2R	2.09
rs6094716	NCOA3	1.93	rs230498	NFKB1	2.01	rs4075494	LHCGR	2.05
rs4846479	TGFB2	1.92	rs6051545	GNRH2	2.01	rs1609798	NFKB1	2.04
rs2294090	FBXO32	1.91	rs2294090	FBXO32	2.00	rs743572	CYP17A1	2.03
rs4988492	GHRH	1.90	rs2296236	GNRH2	1.97	rs4149570	TNFRSF1A	2.02
rs7816339	FBXO32	1.90	rs10735810	VDR	1.93	rs2070776	GH1	1.95
rs735563	CSF2RB	1.89	rs2513924	TIEG	1.80	rs4851534	IL1R2	1.94
rs2305502	LTBP1	1.85	rs2662405	CSF2	1.79	rs6163	CYP17A1	1.94
rs2227319	G-CSF	1.83	rs2470146	CYP19A1	1.78	rs2413432	CSF2RB	1.92
rs11741953	IL6ST	1.82	rs6494629	SMAD3	1.78	rs2058194	GH1	1.90
rs3218927	IL1R2	1.80	rs163078	CYP1B1	1.73	rs7096475	CYP17A1	1.84
rs2043948	LTBP2	1.79	rs384167	IGF2R	1.71	rs886441	VDR	1.82
rs4988300	LRP5	1.79	rs3739287	FBXO32	1.67	rs25640	HSD17B4	1.81
rs312018	LRP5	1.78	rs25879	CSF2	1.67	rs1866146	POMC	1.80
rs222016	GC	1.77	rs785239	GDF8	1.64	rs2075726	CSF2RB	1.75
rs1038426	GNRHR	1.75	rs3773663	TGFBR2	1.62	rs255102	CRHR2	1.74
rs8038645	IGF1R	1.75	rs4363285	CNTFR	1.62	rs4849127	IL1B	1.74
rs31476	CSF2	1.75	rs2043948	LTBP2	1.61	rs12219246	CYP17A1	1.73
rs216124	CSF1R	1.70	rs8038645	IGF1R	1.61	rs1256061	ESR2	1.71
rs886441	VDR	1.65	rs11957312	FST	1.60	rs4894559	TNFSF10	1.71
rs3739287	FBXO32	1.63	rs1805075	IGF2R	1.59	rs12522630	MEF2C	1.70
rs1867348	IGF2R	1.63	rs3798197	IGF2R	1.59	rs2203877	BDNF	1.69
rs2075726	CSF2RB	1.63	rs13034651	NCOA1	1.57	rs8124792	CYP24A1	1.69
rs1993306	LTBP1	1.60	rs3133287	TIEG	1.57	rs7737181	HSD17B4	1.66
rs2058821	LRP6	1.59	rs743572	CYP17A1	1.57	rs11612899	NTF3	1.65
rs2796820	TGFB2	1.59	rs35767	IGF1	1.56	rs35767	IGF1	1.61
rs2270360	IGFBP2	1.58	rs3764875	TNFRSF1A	1.55	rs3774938	NFKB1	1.60
rs506611	WNT5A	1.57	rs2058194	GH1	1.51	rs6425987	CSF3R	1.59
rs2071369	G-CSF	1.54	rs6163	CYP17A1	1.50	rs315920	IL1RN	1.56
rs284185	TGFBR3	1.53	rs3850170	TNFSF10	1.48	rs4462104	CSF3R	1.56
rs4849125	IL1B	1.52	rs4605279	TGFB1	1.48	rs619824	CYP17A1	1.56
rs1003737	IGF2R	1.52	rs3097237	LIFR	1.48	rs2270360	IGFBP2	1.55
rs1163281	MYF6	1.52	rs7148764	LTBP2	1.48	rs9340835	ESR1	1.52
rs1163285	MYF6	1.51	rs7096475	CYP17A1	1.47	rs939348	THRA	1.49
rs284198	TGFBR3	1.49	rs11119328	HSD11B1	1.46	rs7816339	FBXO32	1.47
rs8124792	CYP24A1	1.48	rs1178435	IGFALS	1.45	rs3796720	GNRHR	1.45

A.2B continued

	Mean Arm			Mean Leg			Lumbar Spine	
SNP	Gene	(-1)*LOG$_{10}$P	SNP	Gene	(-1)*LOG$_{10}$P	SNP	Gene	(-1)*LOG$_{10}$P
rs2765880	TGFBR3	3.32	rs4851527	IL1R2	3.80	rs2765880	TGFBR3	4.00
rs2513924	TIEG	3.32	rs2513924	TIEG	3.14	rs2513924	TIEG	3.36
rs4075494	LHCGR	3.31	rs2413432	CSF2RB	3.09	rs315952	IL1RN	3.07
rs2294090	FBXO32	2.90	rs2294090	FBXO32	2.77	rs216124	CSF1R	3.00
rs2070776	GH1	2.66	rs7502966	THRA	2.68	rs315951	IL1RN	2.79
rs4851527	IL1R2	2.61	rs2294891	NCOA3	2.62	rs2294090	FBXO32	2.77
rs2058194	GH1	2.55	rs7757956	ESR1	2.51	rs3133287	TIEG	2.73
rs2026001	HSD17B3	2.50	rs11899433	LTBP1	2.50	rs4251961	IL1RN	2.36
rs6425987	CSF3R	2.45	rs2070776	GH1	2.50	rs4075494	LHCGR	2.35
rs7757956	ESR1	2.37	rs312018	LRP5	2.45	rs7502966	THRA	2.34
rs315952	IL1RN	2.34	rs10783289	CCDC65	2.45	rs5746026	TNFRSF1B	2.25
rs3850651	MEF2C	2.34	rs315952	IL1RN	2.44	rs1178435	IGFALS	2.16
rs2413432	CSF2RB	2.27	rs934328	TGFBR2	2.44	rs35767	IGF1	2.16
rs2203877	BDNF	2.24	rs6066394	NCOA3	2.42	rs2203877	BDNF	2.14
rs1256061	ESR2	2.24	rs7761846	ESR1	2.32	rs3742793	LTBP2	2.13
rs2066474	HSD17B3	2.20	rs2058194	GH1	2.31	rs4903249	LTBP2	2.00
rs6066394	NCOA3	2.20	rs9608854	LIF	2.26	rs3735529	HOXA6	1.97
rs11899433	LTBP1	2.19	rs12522630	MEF2C	2.20	rs12435481	LTBP2	1.89
rs312018	LRP5	2.04	rs315951	IL1RN	2.14	rs10932667	IGFBP2	1.84
rs4648068	NFKB1	2.01	rs7974186	NTF3	2.11	rs7289747	COMT	1.82
rs10735810	VDR	1.99	rs2075726	CSF2RB	2.08	rs8014087	LTBP2	1.80
rs566926	WNT5A	1.96	rs6455678	IGF2R	2.06	rs1944055	CNTF	1.72
rs4830732	STS	1.96	rs2198843	CYP1A1	2.05	rs1056836	CYP1B1	1.71
rs735563	CSF2RB	1.95	rs1867348	IGF2R	1.97	rs2198843	CYP1A1	1.71
rs9854603	TNFSF10	1.93	rs2765880	TGFBR3	1.94	rs2000709	SMAD2	1.71
rs315920	IL1RN	1.87	rs740044	IL1R2	1.92	rs1152912	IRAK3	1.70
rs4849122	IL1A	1.86	rs162556	CYP1B1	1.90	rs2855658	CYP1B1	1.70
rs4894559	TNFSF10	1.84	rs3735529	HOXA6	1.85	rs6954044	GHRHR	1.70
rs570374	WNT5A	1.83	rs1609798	NFKB1	1.81	rs11918967	WNT5A	1.70
rs743572	CYP17A1	1.78	rs3734181	IGF2R	1.76	rs506611	WNT5A	1.69
rs7096475	CYP17A1	1.77	rs3774938	NFKB1	1.76	rs15974	SMAD3	1.66
rs6163	CYP17A1	1.76	rs1889313	IL6R	1.76	rs4648072	NFKB1	1.64
rs422342	SMAD3	1.76	rs2027566	TGFB2	1.73	rs2043948	LTBP2	1.63
rs7502966	THRA	1.75	rs743572	CYP17A1	1.72	rs2270376	STAR	1.63
rs6472528	NCOA2	1.75	rs12615738	LHCGR	1.71	rs4894559	TNFSF10	1.62
rs230498	NFKB1	1.75	rs1993306	LTBP1	1.71	rs985694	ESR1	1.59
rs315951	IL1RN	1.73	rs1163284	MYF6	1.70	rs1363969	FST	1.55
rs12522630	MEF2C	1.71	rs735563	CSF2RB	1.68	rs12424162	NTF3	1.55
rs4849125	IL1B	1.68	rs3850651	MEF2C	1.68	rs2305502	LTBP1	1.54
rs2546210	HSD17B4	1.67	rs12424162	NTF3	1.68	rs4887364	NTRK3	1.53
rs12219246	CYP17A1	1.67	rs6163	CYP17A1	1.67	rs8024593	IGF1R	1.53
rs1519480	BDNF	1.63	rs315943	IL1RN	1.65	rs2570816	MEF2A	1.52
rs6455678	IGF2R	1.62	rs2301267	LHCGR	1.64	rs3739804	NTRK2	1.51
rs3796720	GNRHR	1.61	rs2009112	TGFB2	1.64	rs4899526	LTBP2	1.51
rs4851522	IL1R2	1.59	rs1163281	MYF6	1.63	rs3890733	VDR	1.51
rs1381112	NTRK3	1.57	rs2284220	CRHR2	1.61	rs86312	HSD17B1	1.50
rs3774938	NFKB1	1.57	rs7149264	TGFB3	1.60	rs2079138	HOXA1	1.49
rs4719885	HOXA6	1.57	rs2289263	SMAD3	1.59	rs9608854	LIF	1.47
rs3097237	LIFR	1.54	rs3828562	GNRHR	1.58	rs2428432	HOXA2	1.47
rs1499886	WNT5A	1.53	rs4851534	IL1R2	1.57	rs4851516	IL1R2	1.46

A.2C. Caucasian American male

Hip			Femoral Neck			Whole Body		
SNP	Gene	$(-1)*LOG_{10}P$	SNP	Gene	$(-1)*LOG_{10}P$	SNP	Gene	$(-1)*LOG_{10}P$
rs3828562	GNRHR	4.12	rs5297	CYP11B1	2.95	rs344357	IGFALS	4.43
rs974483	GNRHR	3.31	rs2043948	LTBP2	2.88	rs1178435	IGFALS	4.20
rs972072	GNRHR	3.13	rs6552115	GNRHR	2.72	rs10495098	TGFB2	3.18
rs5297	CYP11B1	2.99	rs2073617	TNFRSF11B	2.61	rs952817	TGFB2	3.06
rs2043948	LTBP2	2.86	rs972072	GNRHR	2.53	rs6658835	TGFB2	2.55
rs6552115	GNRHR	2.76	rs3796720	GNRHR	2.37	rs86312	HSD17B1	2.52
rs952817	TGFB2	2.69	rs3828562	GNRHR	2.30	rs2274619	GNRH2	2.44
rs10495098	TGFB2	2.68	rs10482672	NR3C1	2.21	rs2830	HSD17B1	2.41
rs1385499	TNFRSF11B	2.68	rs1385499	TNFRSF11B	2.21	rs2279455	TGFBR3	2.28
rs1178435	IGFALS	2.63	rs7688268	SMAD1	2.17	rs2289737	SMAD1	2.21
rs2073617	TNFRSF11B	2.51	rs3798188	IGF2R	2.04	rs2470171	CYP19A1	2.16
rs3796720	GNRHR	2.50	rs2364154	IGFBP6	2.02	rs2043948	LTBP2	2.13
rs162558	CYP1B1	2.32	rs3020368	ESR1	2.02	rs934632	CYP19A1	1.95
rs7502966	THRA	2.25	rs2453477	WNT10B	2.00	rs2028377	LTBP2	1.91
rs6469783	TNFRSF11B	2.23	rs4149577	TNFRSF1A	1.95	rs6545066	LHCGR	1.82
rs344357	IGFALS	2.19	rs833480	WNT10B	1.86	rs230539	NFKB1	1.81
rs7688268	SMAD1	2.14	rs2830	HSD17B1	1.81	rs325381	MEF2A	1.78
rs2830	HSD17B1	2.12	rs2813545	ESR1	1.80	rs4601790	LTBP3	1.77
rs6658835	TGFB2	2.12	rs3087949	MYOG	1.79	rs608343	LRP5	1.77
rs6428828	HSD3B1	2.11	rs1531362	NCOA2	1.76	rs1143643	IL1B	1.75
rs1431131	TGFBR2	2.06	rs974483	GNRHR	1.75	rs2046736	TGFBR3	1.73
rs6018600	NCOA3	1.93	rs235219	TNFRSF1B	1.73	rs7165875	IGF1R	1.73
rs12370969	NTF3	1.93	rs2256501	CYP24A1	1.73	rs4648068	NFKB1	1.73
rs1420731	LRP6	1.91	rs7605442	BMPR2	1.72	rs3784434	NTRK3	1.71
rs2274619	GNRH2	1.89	rs872313	WNT10B	1.72	rs230498	NFKB1	1.69
rs10482672	NR3C1	1.87	rs6580928	IGFBP6	1.71	rs1050316	MEF2D	1.69
rs1380897	NCOA2	1.83	rs2479828	HSD17B3	1.65	rs1171554	MEF2D	1.69
rs3798188	IGF2R	1.82	rs216143	CSF1R	1.65	rs4471211	GNRH2	1.67
rs3087949	MYOG	1.77	rs2046736	TGFBR3	1.63	rs4849125	IL1B	1.65
rs872313	WNT10B	1.76	rs2296621	TGFBR3	1.63	rs11939979	SMAD1	1.64
rs948562	CNTF	1.73	rs496888	TNFRSF1B	1.60	rs12105106	NCOA1	1.62
rs6428830	HSD3B1	1.73	rs6469783	TNFRSF11B	1.59	rs6741840	LHCGR	1.61
rs1369430	NTRK3	1.69	rs1171559	MEF2D	1.58	rs235219	TNFRSF1B	1.60
rs2076549	NCOA3	1.69	rs3782726	TNFRSF1A	1.58	rs2073617	TNFRSF11B	1.57
rs1531362	NCOA2	1.67	rs948562	CNTF	1.56	rs1609798	NFKB1	1.56
rs162555	CYP1B1	1.66	rs1867348	IGF2R	1.55	rs4723025	GHRHR	1.55
rs2447752	CYP1B1	1.63	rs2476923	HSD17B3	1.55	rs1420731	LRP6	1.54
rs1867348	IGF2R	1.62	rs2276983	CSF1R	1.55	rs5297	CYP11B1	1.54
rs2296621	TGFBR3	1.61	rs162558	CYP1B1	1.54	rs11100883	SMAD1	1.54
rs7757956	ESR1	1.60	rs284198	TGFBR3	1.53	rs1925950	MEF2D	1.53
rs6066394	NCOA3	1.58	rs699374	LTBP2	1.52	rs10501087	BDNF	1.52
rs3020368	ESR1	1.58	rs3784416	NTRK3	1.50	rs2274316	MEF2D	1.52
rs235219	TNFRSF1B	1.56	rs1420731	LRP6	1.50	rs742375	HSD11B1	1.51
rs2453477	WNT10B	1.55	rs1369430	NTRK3	1.48	rs4541	CYP11B1	1.50
rs939348	THRA	1.52	rs4149570	TNFRSF1A	1.47	rs699374	LTBP2	1.49
rs1206875	NCOA3	1.51	rs1380897	NCOA2	1.47	rs1285057	ESR1	1.48
rs1038426	GNRHR	1.50	rs1171549	MEF2D	1.46	rs2274317	MEF2D	1.48
rs2046736	TGFBR3	1.49	rs4802533	LHB	1.43	rs11638442	CYP11A1	1.46
rs699374	LTBP2	1.48	rs10868232	NTRK2	1.42	rs6265	BDNF	1.44
rs7165875	IGF1R	1.47	rs1178435	IGFALS	1.41	rs4912911	NR3C1	1.43

A.2C continued

	Mean Arm			Mean Leg			Lumbar spine	
SNP	Gene	(-1)*LOG$_{10}$P	SNP	Gene	(-1)*LOG$_{10}$P	SNP	Gene	(-1)*LOG$_{10}$P
rs86312	HSD17B1	3.20	rs344357	IGFALS	3.44	rs4912911	NR3C1	2.84
rs952817	TGFB2	3.01	rs1178435	IGFALS	3.19	rs2715423	IGF1R	2.77
rs2274619	GNRH2	2.57	rs10495098	TGFB2	2.95	rs1564859	TNFRSF11B	2.27
rs7149264	TGFB3	2.50	rs952817	TGFB2	2.62	rs1808723	MEF2A	2.24
rs10495098	TGFB2	2.49	rs86312	HSD17B1	2.56	rs1285057	ESR1	2.10
rs6658835	TGFB2	2.29	rs2830	HSD17B1	2.33	rs3020368	ESR1	2.10
rs2028377	LTBP2	2.23	rs2043948	LTBP2	2.21	rs2479828	HSD17B3	2.09
rs2479828	HSD17B3	2.16	rs1385499	TNFRSF11B	2.10	rs258747	NR3C1	2.06
rs699374	LTBP2	2.13	rs2274619	GNRH2	2.07	rs7165875	IGF1R	2.03
rs2066485	HSD17B3	2.12	rs2073617	TNFRSF11B	2.07	rs7987864	TNFSF11	1.99
rs2830	HSD17B1	2.10	rs6115763	GNRH2	2.03	rs2304060	CSF1R	1.96
rs219107	LTBP1	2.08	rs1420731	LRP6	2.01	rs2701653	IRAK3	1.95
rs4760648	VDR	2.07	rs6658835	TGFB2	2.01	rs2434651	CRHBP	1.94
rs7165875	IGF1R	2.01	rs1143643	IL1B	1.97	rs4648068	NFKB1	1.92
rs4723025	GHRHR	1.99	rs872313	WNT10B	1.96	rs1542818	IGFBP2	1.90
rs2073617	TNFRSF11B	1.97	rs4601790	LTBP3	1.93	rs131843	CSF2RB	1.90
rs4471211	GNRH2	1.95	rs2453477	WNT10B	1.78	rs5742912	TNFRSF1A	1.90
rs3753519	HSD11B1	1.92	rs4809958	CYP24A1	1.77	rs131842	CSF2RB	1.88
rs1143643	IL1B	1.85	rs4131886	LHCGR	1.76	rs2684811	IGF1R	1.88
rs1385499	TNFRSF11B	1.80	rs2470171	CYP19A1	1.75	rs999480	CYP19A1	1.84
rs4849125	IL1B	1.76	rs948333	MC2R	1.72	rs4940110	SMAD2	1.83
rs3826714	TGFB1	1.75	rs1181332	LRP6	1.69	rs765835	IGF2R	1.81
rs1285057	ESR1	1.64	rs25879	CSF2	1.67	rs419598	IL1RN	1.79
rs344357	IGFALS	1.63	rs929377	CRHR2	1.67	rs948604	SMAD2	1.79
rs2289737	SMAD1	1.58	rs12102144	NTRK3	1.63	rs230539	NFKB1	1.76
rs931273	TNFSF11	1.58	rs727479	CYP19A1	1.62	rs2073617	TNFRSF11B	1.73
rs4541	CYP11B1	1.55	rs7165875	IGF1R	1.60	rs12590673	TGFB3	1.73
rs1178435	IGFALS	1.54	rs7041	GC	1.60	rs10495098	TGFB2	1.72
rs2425955	NCOA3	1.53	rs151256	CYP1B1	1.60	rs230498	NFKB1	1.70
rs1424151	HSD17B2	1.53	rs4374421	LHCGR	1.59	rs706017	HOXA2	1.68
rs3808871	CNTFR	1.52	rs833480	WNT10B	1.57	rs2428432	HOXA2	1.68
rs7757956	ESR1	1.52	rs31473	CSF2	1.55	rs2000709	SMAD2	1.67
rs12105106	NCOA1	1.51	rs2267717	CRHR2	1.52	rs1542176	IL1RN	1.65
rs8038645	IGF1R	1.50	rs3826714	TGFB1	1.51	rs4601790	LTBP3	1.64
rs4849123	IL1B	1.49	rs3762546	GDF8	1.50	rs924140	IGFBP3	1.64
rs652625	TNFRSF1B	1.49	rs4723025	GHRHR	1.50	rs6496454	NTRK3	1.63
rs3763615	CNTFR	1.49	rs10482672	NR3C1	1.48	rs1819698	HSD3B2	1.63
rs315920	IL1RN	1.47	rs6872098	LIFR	1.48	rs6976129	HOXA13	1.61
rs7289747	COMT	1.44	rs4912911	NR3C1	1.47	rs6670	IGFBP3	1.61
rs2076549	NCOA3	1.43	rs699374	LTBP2	1.46	rs3825709	LTBP2	1.60
rs6018600	NCOA3	1.41	rs1078985	TGFBR2	1.46	rs315943	IL1RN	1.58
rs1870393	CRH	1.41	rs230498	NFKB1	1.45	rs2257157	HSD17B3	1.56
rs1420731	LRP6	1.40	rs7289747	COMT	1.45	rs1792655	SMAD2	1.55
rs1542818	IGFBP2	1.40	rs2284220	CRHR2	1.44	rs9661103	TGFBR3	1.54
rs706017	HOXA2	1.39	rs2028377	LTBP2	1.43	rs1152888	IRAK3	1.53
rs6127127	CYP24A1	1.38	rs1444888	LIFR	1.42	rs876687	TGFBR2	1.53
rs2143491	NCOA3	1.38	rs3791783	GDF8	1.42	rs1609798	NFKB1	1.53
rs11939979	SMAD1	1.38	rs6127127	CYP24A1	1.41	rs344357	IGFALS	1.53
rs2286411	LTBP2	1.37	rs6018600	NCOA3	1.40	rs7757956	ESR1	1.53
rs4844880	HSD11B1	1.37	rs162558	CYP1B1	1.40	rs1446510	IL1RN	1.49

A.2D Caucasian American Female

Hip			Femoral Neck			Whole Body		
SNP	Gene	(-1)*LOG$_{10}$P	SNP	Gene	(-1)*LOG$_{10}$P	SNP	Gene	(-1)*LOG$_{10}$P
rs1553009	IGFBP2	2.70	rs1668339	IGFBP4	2.52	rs4619	IGFBP1	2.77
rs1568400	THRA	2.64	rs1553009	IGFBP2	2.48	rs7176429	NTRK3	2.74
rs1700686	CRHBP	2.53	rs2289779	LIFR	2.45	rs1491850	BDNF	2.63
rs11740179	HSD17B4	2.37	rs2436964	TIEG	2.45	rs2436964	TIEG	2.59
rs4619	IGFBP1	2.32	rs4242341	FBXO32	2.27	rs1004446	IGF2	2.43
rs8034284	IGF1R	2.32	rs702497	IGF1R	2.24	rs2939421	TNFRSF11A	2.35
rs827420	ESR1	2.30	rs734351	IGF2	2.12	rs1369430	NTRK3	2.30
rs4131885	LHCGR	2.11	rs2228603	MEF2B	2.10	rs1065780	IGFBP1	2.16
rs12435481	LTBP2	2.04	rs699374	LTBP2	2.07	rs2899276	CSF2RB	2.16
rs2227321	G-CSF	2.03	rs389512	CYP21A2	2.03	rs638051	LRP5	2.14
rs2436964	TIEG	2.03	rs920293	SMAD3	1.97	rs6667450	CSF3R	2.00
rs2164808	POMC	2.02	rs6865659	CSF1R	1.97	rs738149	CSF2RB	1.98
rs11575194	IGFBP5	1.98	rs1700686	CRHBP	1.93	rs31476	CSF2	1.96
rs6872098	LIFR	1.94	rs1171565	MEF2D	1.92	rs1568400	THRA	1.95
rs1080750	CNTFR	1.93	rs2070720	GH1	1.92	rs1613519	CRHBP	1.93
rs1012157	COMT	1.89	rs6670	IGFBP3	1.90	rs218218	LTBP1	1.86
rs2205181	TGFB3	1.89	rs2043136	TGFBR2	1.86	rs2000220	TGFB2	1.84
rs6964896	HOXA5	1.86	rs3826620	TNFRSF11A	1.85	rs233998	TNFSF10	1.83
rs2715423	IGF1R	1.85	rs496888	TNFRSF1B	1.82	rs10423094	LHB	1.77
rs2854184	GH1	1.85	rs1864972	CSF1R	1.82	rs9324916	NR3C1	1.77
rs1065780	IGFBP1	1.83	rs2854184	GH1	1.80	rs1942159	SMAD2	1.76
rs2228603	MEF2B	1.81	rs827420	ESR1	1.80	rs1080750	CNTFR	1.76
rs1004446	IGF2	1.79	rs3099115	LIFR	1.79	rs2117655	NTRK3	1.75
rs708122	WNT3A	1.77	rs2469250	CRH	1.79	rs150703	LTBP1	1.75
rs7987864	TNFSF11	1.76	rs6964896	HOXA5	1.76	rs827420	ESR1	1.74
rs2070720	GH1	1.75	rs1568400	THRA	1.76	rs218223	LTBP1	1.73
rs9340835	ESR1	1.75	rs6988439	FBXO32	1.75	rs3826620	TNFRSF11A	1.72
rs10064000	HSD17B4	1.66	rs3773649	TGFBR2	1.75	rs2075724	CSF2RB	1.72
rs949963	IL1R1	1.66	rs3917296	IL1R1	1.72	rs10868232	NTRK2	1.71
rs702497	IGF1R	1.65	rs2715423	IGF1R	1.70	rs3864277	CSF2	1.68
rs1539871	SMAD2	1.64	rs233998	TNFSF10	1.69	rs2228603	MEF2B	1.68
rs876687	TGFBR2	1.64	rs2164808	POMC	1.69	rs4256003	LHCGR	1.67
rs2701653	IRAK3	1.63	rs676318	LRP5	1.68	rs2285724	HOXA13	1.65
rs1792684	SMAD2	1.63	rs2385277	FBXO32	1.68	rs1700686	CRHBP	1.64
rs920293	SMAD3	1.62	rs10896	MEF2B	1.64	rs713869	CSF2RB	1.64
rs934778	POMC	1.61	rs4966012	IGF1R	1.64	rs150720	LTBP1	1.62
rs9608859	LIF	1.58	rs6872098	LIFR	1.62	rs7721081	GHR	1.58
rs7165875	IGF1R	1.57	rs8034284	IGF1R	1.58	rs3791783	GDF8	1.58
rs2043136	TGFBR2	1.55	rs7761846	ESR1	1.54	rs10249499	IGFBP1	1.58
rs1470004	SMAD3	1.54	rs1065780	IGFBP1	1.54	rs3130062	TNFA	1.58
rs13010656	BMPR2	1.53	rs4619	IGFBP1	1.53	rs1978346	IGFBP5	1.57
rs705125	GC	1.51	rs3771199	IL1R1	1.53	rs3802428	CNTFR	1.57
rs4129472	GHR	1.51	rs2701653	IRAK3	1.52	rs11575194	IGFBP5	1.56
rs2451818	HSD17B4	1.51	rs1662701	NTRK2	1.47	rs2069616	CSF2	1.53
rs10495960	LHCGR	1.50	rs535058	IGFBP4	1.45	rs2684781	IGF1R	1.52
rs222014	GC	1.50	rs876688	TGFBR2	1.45	rs1171554	MEF2D	1.48
rs6265	BDNF	1.50	rs1152918	IRAK3	1.45	rs2796820	TGFB2	1.47
rs584828	IGFBP4	1.49	rs7827149	FBXO32	1.44	rs888415	LTBP2	1.46
rs218178	LTBP1	1.48	rs13010656	BMPR2	1.40	rs6537355	SMAD1	1.45
rs3171845	IL1R1	1.46	rs2230400	IL1R2	1.39	rs2662405	CSF2	1.43

A.2D continued

Mean Arm			Mean Leg			Lumbar Spine		
SNP	Gene	(-1)*LOG$_{10}$P	SNP	Gene	(-1)*LOG$_{10}$P	SNP	Gene	(-1)*LOG$_{10}$P
rs4619	IGFBP1	3.58	rs233998	TNFSF10	3.15	rs3102724	TNFRSF11B	3.89
rs7736209	GHR	2.98	rs4619	IGFBP1	3.04	rs7837123	TNFRSF11B	3.30
rs9854603	TNFSF10	2.87	rs5297	CYP11B1	2.92	rs31476	CSF2	2.97
rs2228603	MEF2B	2.80	rs7176429	NTRK3	2.70	rs3802428	CNTFR	2.54
rs5297	CYP11B1	2.68	rs2436964	TIEG	2.51	rs25887	CSF2	2.50
rs3759084	MYF5	2.54	rs1065780	IGFBP1	2.39	rs7148764	LTBP2	2.35
rs4129472	GHR	2.46	rs412458	MEF2C	2.37	rs1080750	CNTFR	2.33
rs3826620	TNFRSF11A	2.43	rs827420	ESR1	2.31	rs4129472	GHR	2.30
rs1065780	IGFBP1	2.41	rs920293	SMAD3	2.28	rs2028377	LTBP2	2.28
rs9872592	TNFSF10	2.04	rs1369430	NTRK3	2.24	rs6494629	SMAD3	2.24
rs7721081	GHR	2.00	rs622803	MEF2C	2.16	rs2118613	SMAD3	2.01
rs2205181	TGFB3	1.96	rs1553009	IGFBP2	2.12	rs7835846	TNFRSF11B	2.00
rs12435481	LTBP2	1.95	rs2118610	SMAD3	2.11	rs6469783	TNFRSF11B	1.99
rs10445364	CRHR1	1.91	rs2228603	MEF2B	2.11	rs4648068	NFKB1	1.96
rs2164808	POMC	1.90	rs218178	LTBP1	2.11	rs1004446	IGF2	1.96
rs4966008	IGF1R	1.88	rs1962589	IGF1R	2.10	rs2070720	GH1	1.88
rs7569	LTBP2	1.88	rs6670	IGFBP3	2.09	rs2207232	ESR1	1.85
rs934778	POMC	1.86	rs1568400	THRA	2.07	rs150703	LTBP1	1.83
rs6670	IGFBP3	1.85	rs7721081	GHR	2.07	rs3782353	WNT10B	1.82
rs1305277	MYF5	1.84	rs1978346	IGFBP5	2.06	rs1529717	TGFB1	1.82
rs7148018	LTBP2	1.83	rs934778	POMC	1.99	rs699374	LTBP2	1.78
rs1568400	THRA	1.82	rs958829	HSD17B3	1.96	rs2662405	CSF2	1.77
rs1004446	IGF2	1.80	rs1662701	NTRK2	1.91	rs2304070	CSF1R	1.77
rs9324916	NR3C1	1.79	rs4129472	GHR	1.89	rs2000220	TGFB2	1.76
rs2270628	IGFBP2	1.78	rs7569	LTBP2	1.81	rs888415	LTBP2	1.72
rs242944	CRHR1	1.77	rs218223	LTBP1	1.77	rs2000709	SMAD2	1.70
rs1073768	GHRH	1.77	rs11575194	IGFBP5	1.76	rs3826620	TNFRSF11A	1.69
rs3130062	TNFA	1.77	rs2053294	SMAD3	1.75	rs2227321	G-CSF	1.69
rs2118610	SMAD3	1.74	rs4605279	TGFB1	1.74	rs6667450	CSF3R	1.67
rs11766273	IL6	1.74	rs4966008	IGF1R	1.74	rs242939	CRHR1	1.66
rs11247361	IGF1R	1.71	rs1995050	IGFBP1	1.72	rs638051	LRP5	1.64
rs10157853	TGFBR3	1.67	rs939348	THRA	1.71	rs2361634	AR	1.64
rs325381	MEF2A	1.66	rs2684761	IGF1R	1.69	rs10492120	LRP6	1.56
rs232001	TNFSF10	1.66	rs2662405	CSF2	1.68	rs3801776	HOXA9	1.55
rs506611	WNT5A	1.66	rs1171554	MEF2D	1.67	rs998075	IGF2R	1.55
rs218178	LTBP1	1.64	rs1080750	CNTFR	1.66	rs7736209	GHR	1.55
rs1613519	CRHBP	1.64	rs1700686	CRHBP	1.66	rs4320932	IGF2	1.54
rs4586411	IGF1R	1.63	rs1032129	TNFRSF11B	1.65	rs10423094	LHB	1.54
rs7816339	FBXO32	1.61	rs7148018	LTBP2	1.64	rs948562	CNTF	1.54
rs11575194	IGFBP5	1.60	rs394243	HSD17B3	1.60	rs947712	TGFB2	1.52
rs235214	TNFRSF1B	1.59	rs2899276	CSF2RB	1.60	rs11638442	CYP11A1	1.52
rs4736354	CYP11B1	1.59	rs4890114	IGFBP4	1.58	rs1316474	IGF1R	1.51
rs2715423	IGF1R	1.59	rs2270628	IGFBP2	1.57	rs6461992	HOXA11	1.50
rs4264434	CRHR1	1.59	rs219107	LTBP1	1.56	rs738149	CSF2RB	1.50
rs2684761	IGF1R	1.57	rs1004446	IGF2	1.56	rs2372866	IGFBP5	1.50
rs488133	ESR1	1.55	rs31476	CSF2	1.55	rs1792655	SMAD2	1.49
rs1800796	IL6	1.55	rs2117655	NTRK3	1.55	rs7709790	GHR	1.48
rs4601790	LTBP3	1.54	rs7165875	IGF1R	1.55	rs1317681	TGFB2	1.48
rs1553009	IGFBP2	1.53	rs4256003	LHCGR	1.55	rs1369430	NTRK3	1.47
rs3131637	TNFA	1.53	rs1046909	TGFB1	1.55	rs4940110	SMAD2	1.47

Table A 3: SNPs used in the study

SNP	GENE	SNP	GENE	SNP	GENE
rs4407366	ACVR2B	rs165849	COMT	rs216143	CSF1R
rs6779704	ACVR2B	rs174696	COMT	rs216148	CSF1R
rs7373816	ACVR2B	rs740603	COMT	rs740750	CSF1R
rs7374458	ACVR2B	rs933271	COMT	rs1010101	CSF1R
rs782810	ADIPOR1	rs1012157	COMT	rs1077724	CSF1R
rs1539355	ADIPOR1	rs1978058	COMT	rs1465693	CSF1R
rs2185781	ADIPOR1	rs2020917	COMT	rs1549919	CSF1R
rs2275737	ADIPOR1	rs2239393	COMT	rs1864972	CSF1R
rs6672643	ADIPOR1	rs2239394	COMT	rs2276983	CSF1R
rs6688148	ADIPOR1	rs4485648	COMT	rs2304060	CSF1R
rs10920539	ADIPOR1	rs4819850	COMT	rs2304070	CSF1R
rs833830	ARF3	rs7289747	COMT	rs3776075	CSF1R
rs6265	BDNF	rs1870393	CRH	rs6865659	CSF1R
rs908867	BDNF	rs2446432	CRH	rs6874087	CSF1R
rs1008041	BDNF	rs2469250	CRH	rs25879	CSF2
rs1491850	BDNF	rs7013524	CRH	rs25881	CSF2
rs1491851	BDNF	rs1500	CRHBP	rs25882	CSF2
rs1519480	BDNF	rs1613519	CRHBP	rs25887	CSF2
rs1552736	BDNF	rs1700681	CRHBP	rs31473	CSF2
rs2203877	BDNF	rs1700686	CRHBP	rs31476	CSF2
rs10501087	BDNF	rs1875999	CRHBP	rs587759	CSF2
rs12273363	BDNF	rs2434651	CRHBP	rs721121	CSF2
rs1996270	BMPR2	rs7718461	CRHBP	rs2069616	CSF2
rs2228545	BMPR2	rs10474485	CRHBP	rs2069636	CSF2
rs2350807	BMPR2	rs11950140	CRHBP	rs2662405	CSF2
rs4675278	BMPR2	rs110402	CRHR1	rs3864277	CSF2
rs6715945	BMPR2	rs242924	CRHR1	rs131842	CSF2RB
rs7605442	BMPR2	rs242939	CRHR1	rs131843	CSF2RB
rs12621870	BMPR2	rs242944	CRHR1	rs713869	CSF2RB
rs13010656	BMPR2	rs4264434	CRHR1	rs735563	CSF2RB
rs10783289	CCDC65	rs4792887	CRHR1	rs738149	CSF2RB
rs550942	CNTF	rs7225082	CRHR1	rs2072707	CSF2RB
rs948562	CNTF	rs10445364	CRHR1	rs2075724	CSF2RB
rs1944055	CNTF	rs255102	CRHR2	rs2075726	CSF2RB
rs1080750	CNTFR	rs255112	CRHR2	rs2239749	CSF2RB
rs1571401	CNTFR	rs255125	CRHR2	rs2284031	CSF2RB
rs2070802	CNTFR	rs255126	CRHR2	rs2413432	CSF2RB
rs3763613	CNTFR	rs255131	CRHR2	rs2899276	CSF2RB
rs3763615	CNTFR	rs929377	CRHR2	rs3917924	CSF3R
rs3802427	CNTFR	rs973002	CRHR2	rs3917973	CSF3R
rs3802428	CNTFR	rs2190242	CRHR2	rs3917974	CSF3R
rs3808870	CNTFR	rs2240403	CRHR2	rs3917981	CSF3R
rs3808871	CNTFR	rs2267717	CRHR2	rs3917996	CSF3R
rs4363285	CNTFR	rs2284217	CRHR2	rs3917997	CSF3R
rs4395980	CNTFR	rs2284220	CRHR2	rs3918020	CSF3R
rs4472605	CNTFR	rs3779250	CRHR2	rs4462104	CSF3R
rs4680	COMT	rs4723002	CRHR2	rs4571935	CSF3R
rs4818	COMT	rs216123	CSF1R	rs6425987	CSF3R
rs165774	COMT	rs216124	CSF1R	rs6662501	CSF3R
rs165815	COMT	rs216136	CSF1R	rs6667450	CSF3R

Table A.3 continued

SNP	GENE	SNP	GENE	SNP	GENE
rs6667860	CSF3R	rs3784308	CYP19A1	rs926779	ESR1
rs11263901	CSF3R	rs6493494	CYP19A1	rs985694	ESR1
rs11263903	CSF3R	rs1799814	CYP1A1	rs988328	ESR1
rs1378942	CSK	rs2198843	CYP1A1	rs1285057	ESR1
rs6161	CYP11A1	rs2470893	CYP1A1	rs1543403	ESR1
rs1484215	CYP11A1	rs2470890	CYP1A2	rs1709183	ESR1
rs2073475	CYP11A1	rs2472299	CYP1A2	rs1801132	ESR1
rs2279356	CYP11A1	rs10916	CYP1B1	rs1884051	ESR1
rs7172632	CYP11A1	rs151256	CYP1B1	rs1884052	ESR1
rs8026319	CYP11A1	rs162550	CYP1B1	rs1884054	ESR1
rs11638442	CYP11A1	rs162555	CYP1B1	rs2207232	ESR1
rs4534	CYP11B1	rs162556	CYP1B1	rs2234693	ESR1
rs4541	CYP11B1	rs162557	CYP1B1	rs2813545	ESR1
rs5287	CYP11B1	rs162558	CYP1B1	rs2881766	ESR1
rs5288	CYP11B1	rs163078	CYP1B1	rs2982896	ESR1
rs5296	CYP11B1	rs1056827	CYP1B1	rs3020314	ESR1
rs5297	CYP11B1	rs1056836	CYP1B1	rs3020368	ESR1
rs5299	CYP11B1	rs2447752	CYP1B1	rs3020383	ESR1
rs5300	CYP11B1	rs2855658	CYP1B1	rs3020407	ESR1
rs5301	CYP11B1	rs389512	CYP21A2	rs3020411	ESR1
rs5304	CYP11B1	rs389883	CYP21A2	rs3853248	ESR1
rs6410	CYP11B1	rs912505	CYP24A1	rs3853251	ESR1
rs4736354	CYP11B1	rs927650	CYP24A1	rs6557170	ESR1
rs6163	CYP17A1	rs964293	CYP24A1	rs6557171	ESR1
rs9527	CYP17A1	rs1555439	CYP24A1	rs7757956	ESR1
rs619824	CYP17A1	rs2248359	CYP24A1	rs7761846	ESR1
rs743572	CYP17A1	rs2256501	CYP24A1	rs9340835	ESR1
rs2486758	CYP17A1	rs2426498	CYP24A1	rs9340954	ESR1
rs4409766	CYP17A1	rs2762934	CYP24A1	rs12154178	ESR1
rs4919682	CYP17A1	rs4809958	CYP24A1	rs1256030	ESR2
rs7089422	CYP17A1	rs4809960	CYP24A1	rs1256044	ESR2
rs7096475	CYP17A1	rs6023012	CYP24A1	rs1256049	ESR2
rs12219246	CYP17A1	rs6068816	CYP24A1	rs1256061	ESR2
rs10046	CYP19A1	rs6097797	CYP24A1	rs1256063	ESR2
rs700518	CYP19A1	rs6127127	CYP24A1	rs1887994	ESR2
rs700519	CYP19A1	rs8124792	CYP24A1	rs1952586	ESR2
rs727479	CYP19A1	rs1048691	CYP27B1	rs3020450	ESR2
rs764530	CYP19A1	rs2069502	CYP27B1	rs7159462	ESR2
rs934632	CYP19A1	rs2270777	CYP27B1	rs8004842	ESR2
rs934633	CYP19A1	rs8176344	CYP27B1	rs8017441	ESR2
rs936307	CYP19A1	rs8176345	CYP27B1	rs2280915	FBXO32
rs999480	CYP19A1	rs2242480	CYP3A4	rs2280919	FBXO32
rs1004984	CYP19A1	rs2740574	CYP3A4	rs2294090	FBXO32
rs1870049	CYP19A1	rs4986910	CYP3A4	rs2385269	FBXO32
rs2445781	CYP19A1	rs12333983	CYP3A4	rs2385277	FBXO32
rs2446419	CYP19A1	rs488133	ESR1	rs3739286	FBXO32
rs2470146	CYP19A1	rs532010	ESR1	rs3739287	FBXO32
rs2470171	CYP19A1	rs726281	ESR1	rs3739288	FBXO32
rs2899472	CYP19A1	rs827420	ESR1	rs4242341	FBXO32
rs3759809	CYP19A1	rs910416	ESR1	rs4871385	FBXO32

Table A.3 continued

SNP	GENE	SNP	GENE	SNP	GENE
rs4871387	FBXO32	rs3791783	GDF8	rs2274619	GNRH2
rs6470155	FBXO32	rs10931441	GDF8	rs2282014	GNRH2
rs6983670	FBXO32	rs2058194	GH1	rs2296236	GNRH2
rs6988439	FBXO32	rs2070720	GH1	rs4471211	GNRH2
rs7000759	FBXO32	rs2070776	GH1	rs6051545	GNRH2
rs7816339	FBXO32	rs2727308	GH1	rs6051551	GNRH2
rs7821824	FBXO32	rs2854184	GH1	rs6076466	GNRH2
rs7827149	FBXO32	rs6179	GHR	rs6115763	GNRH2
rs7775	FRZB	rs6180	GHR	rs6138982	GNRH2
rs288244	FRZB	rs6182	GHR	rs972072	GNRHR
rs288324	FRZB	rs6183	GHR	rs974483	GNRHR
rs288326	FRZB	rs719756	GHR	rs1038426	GNRHR
rs430350	FRZB	rs1876790	GHR	rs1843593	GNRHR
rs1530056	FRZB	rs2036745	GHR	rs2711158	GNRHR
rs4581857	FRZB	rs3764451	GHR	rs3796720	GNRHR
rs4666865	FRZB	rs4129472	GHR	rs3828562	GNRHR
rs7607737	FRZB	rs4242119	GHR	rs6552115	GNRHR
rs10177746	FRZB	rs4866931	GHR	rs6828922	GNRHR
rs10181110	FRZB	rs7709790	GHR	rs2079138	HOXA1
rs10206992	FRZB	rs7721081	GHR	rs2462907	HOXA1
rs745321	FST	rs7736209	GHR	rs12700794	HOXA1
rs1030227	FST	rs10473282	GHR	rs1859164	HOXA10
rs1363969	FST	rs11744988	GHR	rs2067087	HOXA11
rs1423560	FST	rs12233949	GHR	rs4722670	HOXA11
rs1469101	FST	rs1073768	GHRH	rs6461992	HOXA11
rs2042332	FST	rs2050093	GHRH	rs6968828	HOXA11
rs3797297	FST	rs4988492	GHRH	rs2189239	HOXA13
rs6880850	FST	rs6032470	GHRH	rs2285724	HOXA13
rs11957312	FST	rs740336	GHRHR	rs3807598	HOXA13
rs7041	GC	rs2267721	GHRHR	rs6974705	HOXA13
rs222014	GC	rs2267723	GHRHR	rs6976129	HOXA13
rs222016	GC	rs2302019	GHRHR	rs10486494	HOXA13
rs705117	GC	rs4723025	GHRHR	rs11975202	HOXA13
rs705125	GC	rs4988498	GHRHR	rs706017	HOXA2
rs842881	GC	rs6954044	GHRHR	rs2428432	HOXA2
rs1352844	GC	rs7384927	GHRHR	rs2465275	HOXA2
rs1491709	GC	rs7458593	GHRHR	rs2428433	HOXA3
rs1491711	GC	rs11771444	GHRHR	rs1859166	HOXA4
rs3733359	GC	rs27647	GHRL	rs6964896	HOXA5
rs1042657	G-CSF	rs696217	GHRL	rs3735529	HOXA6
rs1042658	G-CSF	rs6185	GNRH1	rs4719885	HOXA6
rs2071369	G-CSF	rs958579	GNRH1	rs4722665	HOXA7
rs2227319	G-CSF	rs1876281	GNRH1	rs3801776	HOXA9
rs2227321	G-CSF	rs7842899	GNRH1	rs10259620	HOXA9
rs2241245	G-CSF	rs10503771	GNRH1	rs11773804	HOXA9
rs2302775	G-CSF	rs566570	GNRH2	rs742375	HSD11B1
rs8070454	G-CSF	rs633293	GNRH2	rs846906	HSD11B1
rs785239	GDF8	rs872786	GNRH2	rs864028	HSD11B1
rs2562802	GDF8	rs1034063	GNRH2	rs2205985	HSD11B1
rs3762546	GDF8	rs2274617	GNRH2	rs2235543	HSD11B1

Table A.3 continued

SNP	GENE	SNP	GENE	SNP	GENE
rs3753519	HSD11B1	rs2275609	HSD3B1	rs7165875	IGF1R
rs4844484	HSD11B1	rs6428828	HSD3B1	rs7403473	IGF1R
rs4844486	HSD11B1	rs6428830	HSD3B1	rs8024593	IGF1R
rs4844488	HSD11B1	rs879332	HSD3B2	rs8027767	IGF1R
rs4844880	HSD11B1	rs947129	HSD3B2	rs8030950	IGF1R
rs11119328	HSD11B1	rs1819698	HSD3B2	rs8032477	IGF1R
rs2830	HSD17B1	rs2275608	HSD3B2	rs8034284	IGF1R
rs86312	HSD17B1	rs10923805	HSD3B2	rs8038645	IGF1R
rs647397	HSD17B1	rs6214	IGF1	rs10794486	IGF1R
rs938671	HSD17B1	rs6219	IGF1	rs11247361	IGF1R
rs1048183	HSD17B1	rs35745	IGF1	rs11635928	IGF1R
rs723012	HSD17B2	rs35765	IGF1	rs734351	IGF2
rs1364286	HSD17B2	rs35767	IGF1	rs1004446	IGF2
rs1424151	HSD17B2	rs860597	IGF1	rs2070762	IGF2
rs2042429	HSD17B2	rs865927	IGF1	rs4320932	IGF2
rs2317615	HSD17B2	rs1019731	IGF1	rs7924316	IGF2
rs3751860	HSD17B2	rs2033178	IGF1	rs384167	IGF2R
rs4445895	HSD17B2	rs2139573	IGF1	rs435612	IGF2R
rs4889450	HSD17B2	rs2195240	IGF1	rs600324	IGF2R
rs8191044	HSD17B2	rs5742667	IGF1	rs629849	IGF2R
rs8191135	HSD17B2	rs7136446	IGF1	rs635551	IGF2R
rs8191246	HSD17B2	rs10735380	IGF1	rs645851	IGF2R
rs394243	HSD17B3	rs702497	IGF1R	rs765835	IGF2R
rs407179	HSD17B3	rs871335	IGF1R	rs998075	IGF2R
rs958829	HSD17B3	rs907808	IGF1R	rs1003737	IGF2R
rs999269	HSD17B3	rs1316474	IGF1R	rs1570070	IGF2R
rs1324196	HSD17B3	rs1319868	IGF1R	rs1805075	IGF2R
rs2026001	HSD17B3	rs1546713	IGF1R	rs1867348	IGF2R
rs2026002	HSD17B3	rs1879612	IGF1R	rs1883869	IGF2R
rs2066474	HSD17B3	rs1962589	IGF1R	rs1888727	IGF2R
rs2066485	HSD17B3	rs2137683	IGF1R	rs2065396	IGF2R
rs2253502	HSD17B3	rs2229765	IGF1R	rs2230043	IGF2R
rs2257157	HSD17B3	rs2684761	IGF1R	rs2230048	IGF2R
rs2476923	HSD17B3	rs2684777	IGF1R	rs2277070	IGF2R
rs2476927	HSD17B3	rs2684781	IGF1R	rs2297370	IGF2R
rs2479828	HSD17B3	rs2684790	IGF1R	rs2342868	IGF2R
rs8190480	HSD17B3	rs2684792	IGF1R	rs3734181	IGF2R
rs25640	HSD17B4	rs2684796	IGF1R	rs3777404	IGF2R
rs32651	HSD17B4	rs2684811	IGF1R	rs3798188	IGF2R
rs32662	HSD17B4	rs2715423	IGF1R	rs3798197	IGF2R
rs442923	HSD17B4	rs2871864	IGF1R	rs6455678	IGF2R
rs1143650	HSD17B4	rs3743249	IGF1R	rs6455680	IGF2R
rs2451818	HSD17B4	rs3784606	IGF1R	rs6917747	IGF2R
rs2546210	HSD17B4	rs4586411	IGF1R	rs6934383	IGF2R
rs2560718	HSD17B4	rs4965425	IGF1R	rs8191746	IGF2R
rs3850201	HSD17B4	rs4966008	IGF1R	rs8191772	IGF2R
rs7737181	HSD17B4	rs4966012	IGF1R	rs8191797	IGF2R
rs10064000	HSD17B4	rs4966014	IGF1R	rs8191819	IGF2R
rs11740179	HSD17B4	rs4966042	IGF1R	rs8191842	IGF2R
rs17145454	HSD17B4	rs4966050	IGF1R	rs8191844	IGF2R

Table A.3 continued

SNP	GENE	SNP	GENE	SNP	GENE
rs9456484	IGF2R	rs1878318	IL1A	rs315951	IL1RN
rs9457785	IGF2R	rs1878320	IL1A	rs315952	IL1RN
rs9457827	IGF2R	rs3783521	IL1A	rs419598	IL1RN
rs344357	IGFALS	rs3783531	IL1A	rs1446510	IL1RN
rs1178435	IGFALS	rs4848302	IL1A	rs1542176	IL1RN
rs4316755	IGFALS	rs4848303	IL1A	rs1623119	IL1RN
rs4619	IGFBP1	rs4849122	IL1A	rs1688075	IL1RN
rs1065780	IGFBP1	rs1143623	IL1B	rs2234676	IL1RN
rs1995050	IGFBP1	rs1143627	IL1B	rs2234678	IL1RN
rs4988515	IGFBP1	rs1143643	IL1B	rs4251961	IL1RN
rs10249499	IGFBP1	rs2048874	IL1B	rs1546762	IL6
rs1525608	IGFBP2	rs4447608	IL1B	rs1800796	IL6
rs1542818	IGFBP2	rs4849123	IL1B	rs2056576	IL6
rs1553009	IGFBP2	rs4849125	IL1B	rs2069824	IL6
rs2056642	IGFBP2	rs4849127	IL1B	rs2069827	IL6
rs2270360	IGFBP2	rs6735739	IL1B	rs2069832	IL6
rs2270628	IGFBP2	rs12469600	IL1B	rs2069849	IL6
rs2372848	IGFBP2	rs949963	IL1R1	rs2069860	IL6
rs3770473	IGFBP2	rs997049	IL1R1	rs4722167	IL6
rs10932667	IGFBP2	rs1030021	IL1R1	rs7776857	IL6
rs10932669	IGFBP2	rs2110726	IL1R1	rs11766273	IL6
rs6670	IGFBP3	rs2192752	IL1R1	rs1386821	IL6R
rs924140	IGFBP3	rs2228139	IL1R1	rs1889313	IL6R
rs3793345	IGFBP3	rs2287047	IL1R1	rs2229238	IL6R
rs9282734	IGFBP3	rs3171845	IL1R1	rs4129267	IL6R
rs535058	IGFBP4	rs3771199	IL1R1	rs4553185	IL6R
rs584828	IGFBP4	rs3917225	IL1R1	rs8192284	IL6R
rs1530363	IGFBP4	rs3917254	IL1R1	rs10908835	IL6R
rs1668339	IGFBP4	rs3917296	IL1R1	rs11265608	IL6R
rs2015561	IGFBP4	rs3917320	IL1R1	rs11265618	IL6R
rs4890114	IGFBP4	rs3917325	IL1R1	rs11265622	IL6R
rs7214466	IGFBP4	rs3917332	IL1R1	rs715180	IL6ST
rs11650680	IGFBP4	rs4851543	IL1R1	rs1373998	IL6ST
rs3276	IGFBP5	rs6712813	IL1R1	rs1900173	IL6ST
rs888184	IGFBP5	rs11883987	IL1R1	rs3729961	IL6ST
rs888186	IGFBP5	rs13387400	IL1R1	rs10940495	IL6ST
rs1978346	IGFBP5	rs733498	IL1R2	rs11741953	IL6ST
rs2067041	IGFBP5	rs740044	IL1R2	rs289067	IRAK3
rs2072544	IGFBP5	rs2190358	IL1R2	rs1152888	IRAK3
rs2372866	IGFBP5	rs2230400	IL1R2	rs1152912	IRAK3
rs3755137	IGFBP5	rs3218927	IL1R2	rs1152918	IRAK3
rs7426116	IGFBP5	rs3218979	IL1R2	rs1183892	IRAK3
rs11575194	IGFBP5	rs4851516	IL1R2	rs1732886	IRAK3
rs2277329	IGFBP6	rs4851519	IL1R2	rs1882200	IRAK3
rs2280699	IGFBP6	rs4851522	IL1R2	rs2043668	IRAK3
rs2364154	IGFBP6	rs4851527	IL1R2	rs2118137	IRAK3
rs6580928	IGFBP6	rs4851534	IL1R2	rs2701653	IRAK3
rs7974876	IGFBP6	rs7589525	IL1R2	rs7972963	IRAK3
rs17561	IL1A	rs315920	IL1RN	rs1062708	LHB
rs1800794	IL1A	rs315943	IL1RN	rs3795044	LHB

Table A.3 continued

SNP	GENE	SNP	GENE	SNP	GENE
rs4802533	LHB	rs2731968	LIFR	rs589403	LTBP1
rs6509412	LHB	rs2914344	LIFR	rs609277	LTBP1
rs10423094	LHB	rs2921153	LIFR	rs711235	LTBP1
rs1464728	LHCGR	rs3097237	LIFR	rs817531	LTBP1
rs1949778	LHCGR	rs3099115	LIFR	rs897509	LTBP1
rs2293275	LHCGR	rs3099124	LIFR	rs1020636	LTBP1
rs2301267	LHCGR	rs4869585	LIFR	rs1377945	LTBP1
rs3884615	LHCGR	rs6872098	LIFR	rs1454312	LTBP1
rs4075494	LHCGR	rs6893525	LIFR	rs1463312	LTBP1
rs4131885	LHCGR	rs10473100	LIFR	rs1545550	LTBP1
rs4131886	LHCGR	rs11737983	LIFR	rs1812485	LTBP1
rs4256003	LHCGR	rs312018	LRP5	rs1869452	LTBP1
rs4319975	LHCGR	rs312788	LRP5	rs1902049	LTBP1
rs4352270	LHCGR	rs314756	LRP5	rs1993306	LTBP1
rs4374421	LHCGR	rs587397	LRP5	rs2015189	LTBP1
rs4453732	LHCGR	rs608343	LRP5	rs2123770	LTBP1
rs4519576	LHCGR	rs634918	LRP5	rs2167973	LTBP1
rs4597581	LHCGR	rs638051	LRP5	rs2290427	LTBP1
rs4953616	LHCGR	rs676318	LRP5	rs2305502	LTBP1
rs4953617	LHCGR	rs923346	LRP5	rs3731575	LTBP1
rs6545066	LHCGR	rs3781600	LRP5	rs3769550	LTBP1
rs6714440	LHCGR	rs4988300	LRP5	rs4569454	LTBP1
rs6741840	LHCGR	rs1012672	LRP6	rs4952277	LTBP1
rs6755901	LHCGR	rs1181332	LRP6	rs6705057	LTBP1
rs7579411	LHCGR	rs1420731	LRP6	rs6719824	LTBP1
rs10495958	LHCGR	rs1894984	LRP6	rs6737948	LTBP1
rs10495959	LHCGR	rs2058821	LRP6	rs6742198	LTBP1
rs10495960	LHCGR	rs2302685	LRP6	rs6751758	LTBP1
rs11125179	LHCGR	rs2417086	LRP6	rs7566302	LTBP1
rs11691408	LHCGR	rs3741792	LRP6	rs7568681	LTBP1
rs12615738	LHCGR	rs4477532	LRP6	rs7578053	LTBP1
rs12713012	LHCGR	rs4763794	LRP6	rs7583121	LTBP1
rs715605	LIF	rs4763797	LRP6	rs10432656	LTBP1
rs737921	LIF	rs7973583	LRP6	rs11886226	LTBP1
rs877549	LIF	rs10492120	LRP6	rs11895736	LTBP1
rs929271	LIF	rs10743980	LRP6	rs11899433	LTBP1
rs929273	LIF	rs10772530	LRP6	rs7569	LTBP2
rs2070889	LIF	rs11054721	LRP6	rs699374	LTBP2
rs2078856	LIF	rs12313200	LRP6	rs754634	LTBP2
rs2267153	LIF	rs12425946	LRP6	rs862046	LTBP2
rs3753082	LIF	rs150703	LTBP1	rs888414	LTBP2
rs3761427	LIF	rs150720	LTBP1	rs888415	LTBP2
rs6006427	LIF	rs218178	LTBP1	rs2028377	LTBP2
rs6518682	LIF	rs218204	LTBP1	rs2043948	LTBP2
rs9608854	LIF	rs218218	LTBP1	rs2286411	LTBP2
rs9608859	LIF	rs218223	LTBP1	rs2302114	LTBP2
rs1444888	LIFR	rs218226	LTBP1	rs2304707	LTBP2
rs1562137	LIFR	rs219107	LTBP1	rs3742793	LTBP2
rs2289779	LIFR	rs219143	LTBP1	rs3825709	LTBP2
rs2561805	LIFR	rs219210	LTBP1	rs4899526	LTBP2

Table A.3 continued

SNP	GENE	SNP	GENE	SNP	GENE
rs4903249	LTBP2	rs1171565	MEF2D	rs6472528	NCOA2
rs7148018	LTBP2	rs1750305	MEF2D	rs6985009	NCOA2
rs7148764	LTBP2	rs1750307	MEF2D	rs7018096	NCOA2
rs7150576	LTBP2	rs1925950	MEF2D	rs10504470	NCOA2
rs8014087	LTBP2	rs2274316	MEF2D	rs10504477	NCOA2
rs10047892	LTBP2	rs2274317	MEF2D	rs11786991	NCOA2
rs12435481	LTBP2	rs6685228	MEF2D	rs13260857	NCOA2
rs947791	LTBP3	rs12569286	MEF2D	rs1206875	NCOA3
rs1151503	LTBP3	rs1163263	MYF5	rs1569438	NCOA3
rs3741380	LTBP3	rs1305277	MYF5	rs2026401	NCOA3
rs4244811	LTBP3	rs1305281	MYF5	rs2076549	NCOA3
rs4601790	LTBP3	rs3759084	MYF5	rs2143491	NCOA3
rs948333	MC2R	rs10862186	MYF5	rs2230782	NCOA3
rs1941088	MC2R	rs3121	MYF6	rs2235734	NCOA3
rs3760534	MC2R	rs1163254	MYF6	rs2281279	NCOA3
rs7243853	MC2R	rs1163281	MYF6	rs2294891	NCOA3
rs325381	MEF2A	rs1163284	MYF6	rs2425955	NCOA3
rs325383	MEF2A	rs1163285	MYF6	rs6018572	NCOA3
rs325403	MEF2A	rs10506833	MYF6	rs6018600	NCOA3
rs325411	MEF2A	rs2249104	MYOD1	rs6018611	NCOA3
rs1808723	MEF2A	rs2734554	MYOD1	rs6066394	NCOA3
rs2164058	MEF2A	rs3858512	MYOD1	rs6094716	NCOA3
rs2570816	MEF2A	rs4756914	MYOD1	rs6094723	NCOA3
rs8036677	MEF2A	rs4757569	MYOD1	rs230498	NFKB1
rs11247113	MEF2A	rs3087949	MYOG	rs230539	NFKB1
rs10896	MEF2B	rs6702345	MYOG	rs747559	NFKB1
rs741706	MEF2B	rs1804645	NCOA1	rs997476	NFKB1
rs1050483	MEF2B	rs2119117	NCOA1	rs1609798	NFKB1
rs2040562	MEF2B	rs3731628	NCOA1	rs1882949	NFKB1
rs2228603	MEF2B	rs6739622	NCOA1	rs3774934	NFKB1
rs7360000	MEF2B	rs6749833	NCOA1	rs3774938	NFKB1
rs10424365	MEF2B	rs7567564	NCOA1	rs3774968	NFKB1
rs167345	MEF2C	rs7582565	NCOA1	rs4648022	NFKB1
rs186233	MEF2C	rs12105106	NCOA1	rs4648037	NFKB1
rs216057	MEF2C	rs12474894	NCOA1	rs4648068	NFKB1
rs304162	MEF2C	rs13004818	NCOA1	rs4648072	NFKB1
rs410216	MEF2C	rs13034651	NCOA1	rs10489113	NFKB1
rs412458	MEF2C	rs1046013	NCOA2	rs6190	NR3C1
rs618298	MEF2C	rs1380897	NCOA2	rs6196	NR3C1
rs622803	MEF2C	rs1531362	NCOA2	rs258747	NR3C1
rs770189	MEF2C	rs2926700	NCOA2	rs1866388	NR3C1
rs3850651	MEF2C	rs2926707	NCOA2	rs4607376	NR3C1
rs4446500	MEF2C	rs2957094	NCOA2	rs4912911	NR3C1
rs12522630	MEF2C	rs2977983	NCOA2	rs7701443	NR3C1
rs750439	MEF2D	rs3812429	NCOA2	rs9324916	NR3C1
rs942964	MEF2D	rs4512409	NCOA2	rs10482672	NR3C1
rs1050316	MEF2D	rs4545135	NCOA2	rs11740792	NR3C1
rs1171549	MEF2D	rs4623463	NCOA2	rs6332	NTF3
rs1171554	MEF2D	rs4738070	NCOA2	rs2052355	NTF3
rs1171559	MEF2D	rs6472527	NCOA2	rs4074967	NTF3

Table A.3 continued

SNP	GENE	SNP	GENE	SNP	GENE
rs7974186	NTF3	rs1461214	NTRK3	rs7688268	SMAD1
rs10744685	NTF3	rs1530310	NTRK3	rs11100883	SMAD1
rs10849275	NTF3	rs1834573	NTRK3	rs11939979	SMAD1
rs11063703	NTF3	rs2117655	NTRK3	rs13109195	SMAD1
rs11612899	NTF3	rs2162266	NTRK3	rs948604	SMAD2
rs12370969	NTF3	rs3784416	NTRK3	rs1539871	SMAD2
rs12424162	NTF3	rs3784434	NTRK3	rs1792655	SMAD2
rs12424856	NTF3	rs3825884	NTRK3	rs1792684	SMAD2
rs6334	NTRK1	rs3825885	NTRK3	rs1942159	SMAD2
rs6336	NTRK1	rs3903308	NTRK3	rs2000709	SMAD2
rs928392	NTRK1	rs4887337	NTRK3	rs3923814	SMAD2
rs943551	NTRK1	rs4887346	NTRK3	rs4940110	SMAD2
rs1800879	NTRK1	rs4887348	NTRK3	rs4940140	SMAD2
rs1888861	NTRK1	rs4887351	NTRK3	rs15974	SMAD3
rs1998977	NTRK1	rs4887364	NTRK3	rs266332	SMAD3
rs2644596	NTRK1	rs4887391	NTRK3	rs266347	SMAD3
rs4661229	NTRK1	rs6496454	NTRK3	rs422342	SMAD3
rs10908521	NTRK1	rs6496469	NTRK3	rs718663	SMAD3
rs11264554	NTRK1	rs7164531	NTRK3	rs745103	SMAD3
rs12145540	NTRK1	rs7170215	NTRK3	rs920293	SMAD3
rs1221	NTRK2	rs7176429	NTRK3	rs991157	SMAD3
rs681329	NTRK2	rs7176520	NTRK3	rs1465842	SMAD3
rs1002261	NTRK2	rs8025158	NTRK3	rs1470003	SMAD3
rs1187321	NTRK2	rs8038245	NTRK3	rs1470004	SMAD3
rs1211443	NTRK2	rs8042990	NTRK3	rs1498506	SMAD3
rs1662701	NTRK2	rs10520671	NTRK3	rs2033785	SMAD3
rs3739804	NTRK2	rs10520676	NTRK3	rs2053294	SMAD3
rs3758317	NTRK2	rs12102144	NTRK3	rs2118610	SMAD3
rs3780634	NTRK2	rs12594095	NTRK3	rs2118612	SMAD3
rs4242632	NTRK2	rs12595249	NTRK3	rs2118613	SMAD3
rs10116287	NTRK2	rs934778	POMC	rs2289263	SMAD3
rs10512156	NTRK2	rs1866146	POMC	rs2414937	SMAD3
rs10746750	NTRK2	rs2118404	POMC	rs3809572	SMAD3
rs10780695	NTRK2	rs2164808	POMC	rs6494629	SMAD3
rs10868232	NTRK2	rs4665777	POMC	rs6494636	SMAD3
rs11140714	NTRK2	rs6545976	POMC	rs7166081	SMAD3
rs744993	NTRK3	rs2293445	PRKAG1	rs7167838	SMAD3
rs744994	NTRK3	rs6259	SHBG	rs7183244	SMAD3
rs878646	NTRK3	rs6260	SHBG	rs948588	SMAD4
rs879131	NTRK3	rs13894	SHBG	rs10502913	SMAD4
rs894290	NTRK3	rs858521	SHBG	rs12455792	SMAD4
rs898706	NTRK3	rs1641536	SHBG	rs2270376	STAR
rs920067	NTRK3	rs1641544	SHBG	rs2843745	STAR
rs922232	NTRK3	rs2955617	SHBG	rs6474491	STAR
rs1017412	NTRK3	rs1016792	SMAD1	rs1046909	TGFB1
rs1105693	NTRK3	rs1866179	SMAD1	rs1529717	TGFB1
rs1365288	NTRK3	rs1874572	SMAD1	rs1982072	TGFB1
rs1369423	NTRK3	rs2036138	SMAD1	rs3826714	TGFB1
rs1369430	NTRK3	rs2289737	SMAD1	rs4605279	TGFB1
rs1381112	NTRK3	rs6537355	SMAD1	rs8110090	TGFB1

Table A.3 continued

SNP	GENE	SNP	GENE	SNP	GENE
rs947712	TGFB2	rs2043138	TGFBR2	rs2269457	THRA
rs952817	TGFB2	rs3773643	TGFBR2	rs3934886	THRA
rs1317681	TGFB2	rs3773649	TGFBR2	rs7502966	THRA
rs1891467	TGFB2	rs3773663	TGFBR2	rs11078936	THRA
rs2000220	TGFB2	rs4522809	TGFBR2	rs1434277	TIEG
rs2009112	TGFB2	rs6550005	TGFBR2	rs1434278	TIEG
rs2027566	TGFB2	rs11129419	TGFBR2	rs1978752	TIEG
rs2796814	TGFB2	rs284142	TGFBR3	rs2436964	TIEG
rs2796820	TGFB2	rs284156	TGFBR3	rs2511660	TIEG
rs2799090	TGFB2	rs284170	TGFBR3	rs2511719	TIEG
rs3121580	TGFB2	rs284180	TGFBR3	rs2511724	TIEG
rs3892225	TGFB2	rs284185	TGFBR3	rs2513924	TIEG
rs4846479	TGFB2	rs284198	TGFBR3	rs3133287	TIEG
rs6658835	TGFB2	rs284876	TGFBR3	rs4734648	TIEG
rs9308385	TGFB2	rs428815	TGFBR3	rs2255798	TNFA
rs10482750	TGFB2	rs883873	TGFBR3	rs2857605	TNFA
rs10495098	TGFB2	rs1555890	TGFBR3	rs2857713	TNFA
rs10863388	TGFB2	rs1804506	TGFBR3	rs3130062	TNFA
rs2205181	TGFB3	rs1805113	TGFBR3	rs3131637	TNFA
rs2284792	TGFB3	rs1926261	TGFBR3	rs4947324	TNFA
rs3917148	TGFB3	rs2029356	TGFBR3	rs11574936	TNFA
rs3917192	TGFB3	rs2046736	TGFBR3	rs884205	TNFRSF11A
rs7149264	TGFB3	rs2046737	TGFBR3	rs1805034	TNFRSF11A
rs7156293	TGFB3	rs2129975	TGFBR3	rs1893510	TNFRSF11A
rs8008060	TGFB3	rs2279455	TGFBR3	rs1942858	TNFRSF11A
rs8011635	TGFB3	rs2296621	TGFBR3	rs2939421	TNFRSF11A
rs12590673	TGFB3	rs2297142	TGFBR3	rs2957127	TNFRSF11A
rs597457	TGFBR1	rs2391068	TGFBR3	rs2957137	TNFRSF11A
rs928180	TGFBR1	rs2765880	TGFBR3	rs2980964	TNFRSF11A
rs2416666	TGFBR1	rs2770186	TGFBR3	rs2980965	TNFRSF11A
rs4742761	TGFBR1	rs3103332	TGFBR3	rs2980984	TNFRSF11A
rs7034462	TGFBR1	rs6604050	TGFBR3	rs3826620	TNFRSF11A
rs10512263	TGFBR1	rs6604054	TGFBR3	rs4303637	TNFRSF11A
rs12346650	TGFBR1	rs6665312	TGFBR3	rs4436867	TNFRSF11A
rs304822	TGFBR2	rs6677523	TGFBR3	rs1032129	TNFRSF11B
rs744751	TGFBR2	rs6684753	TGFBR3	rs1385499	TNFRSF11B
rs876687	TGFBR2	rs7514724	TGFBR3	rs1485286	TNFRSF11B
rs876688	TGFBR2	rs7517044	TGFBR3	rs1564859	TNFRSF11B
rs905456	TGFBR2	rs7526590	TGFBR3	rs2073617	TNFRSF11B
rs934328	TGFBR2	rs9661103	TGFBR3	rs3102724	TNFRSF11B
rs995435	TGFBR2	rs10157853	TGFBR3	rs3134073	TNFRSF11B
rs1012192	TGFBR2	rs10493858	TGFBR3	rs4876872	TNFRSF11B
rs1019856	TGFBR2	rs10493859	TGFBR3	rs6469783	TNFRSF11B
rs1036095	TGFBR2	rs11165354	TGFBR3	rs6987559	TNFRSF11B
rs1078985	TGFBR2	rs11165787	TGFBR3	rs7835846	TNFRSF11B
rs1155705	TGFBR2	rs12562433	TGFBR3	rs7837123	TNFRSF11B
rs1431131	TGFBR2	rs12566180	TGFBR3	rs3764874	TNFRSF1A
rs1835538	TGFBR2	rs939348	THRA	rs3764875	TNFRSF1A
rs1864616	TGFBR2	rs1568400	THRA	rs3782726	TNFRSF1A
rs2043136	TGFBR2	rs2071570	THRA	rs4149570	TNFRSF1A

Table A.3 continued

SNP	GENE	SNP	GENE
rs4149576	TNFRSF1A	rs2238136	VDR
rs4149577	TNFRSF1A	rs2239179	VDR
rs4149579	TNFRSF1A	rs2239182	VDR
rs5742912	TNFRSF1A	rs2239186	VDR
rs10849446	TNFRSF1A	rs2254210	VDR
rs235214	TNFRSF1B	rs3782905	VDR
rs235219	TNFRSF1B	rs3890733	VDR
rs496888	TNFRSF1B	rs4334089	VDR
rs522807	TNFRSF1B	rs4516035	VDR
rs597519	TNFRSF1B	rs4760648	VDR
rs630542	TNFRSF1B	rs7968585	VDR
rs641941	TNFRSF1B	rs7970314	VDR
rs652284	TNFRSF1B	rs10735810	VDR
rs652625	TNFRSF1B	rs11168307	VDR
rs671106	TNFRSF1B	rs11608702	VDR
rs755398	TNFRSF1B	rs12831006	VDR
rs816060	TNFRSF1B	rs833480	WNT10B
rs1061624	TNFRSF1B	rs833839	WNT10B
rs1148459	TNFRSF1B	rs833843	WNT10B
rs1800621	TNFRSF1B	rs872313	WNT10B
rs2275416	TNFRSF1B	rs2453477	WNT10B
rs3766730	TNFRSF1B	rs3782353	WNT10B
rs4846100	TNFRSF1B	rs7311091	WNT1B
rs5746026	TNFRSF1B	rs708121	WNT3A
rs5746051	TNFRSF1B	rs708122	WNT3A
rs5746065	TNFRSF1B	rs752107	WNT3A
rs6697733	TNFRSF1B	rs3121310	WNT3A
rs231983	TNFSF10	rs4653894	WNT3A
rs231985	TNFSF10	rs7539664	WNT3A
rs232000	TNFSF10	rs472631	WNT5A
rs232001	TNFSF10	rs506611	WNT5A
rs233998	TNFSF10	rs556874	WNT5A
rs2041693	TNFSF10	rs566926	WNT5A
rs3136594	TNFSF10	rs570374	WNT5A
rs3181139	TNFSF10	rs597437	WNT5A
rs3850170	TNFSF10	rs815541	WNT5A
rs4894559	TNFSF10	rs1499886	WNT5A
rs9854603	TNFSF10	rs1795651	WNT5A
rs9872592	TNFSF10	rs11918967	WNT5A
rs931273	TNFSF11		
rs1054016	TNFSF11		
rs1325793	TNFSF11		
rs2200287	TNFSF11		
rs3742257	TNFSF11		
rs4531631	TNFSF11		
rs7987864	TNFSF11		
rs886441	VDR		
rs1544410	VDR		
rs2107301	VDR		
rs2189480	VDR		

BIBLIOGRAPHY

(1993). Consensus development conference: diagnosis, prophylaxis, and treatment of osteoporosis. Am J Med 94(6): 646-50.

(1994). Assessment of fracture risk and its application to screening for postmenopausal osteoporosis. Report of a WHO Study Group. World Health Organ Tech Rep Ser 843: 1-129.

(2003). The International HapMap Project. Nature 426(6968): 789-96.

(2005). A haplotype map of the human genome. Nature 437(7063): 1299-320.

(2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447(7145): 661-78.

Andrew, T., L. Antioniades, K. J. Scurrah, A. J. Macgregor and T. D. Spector (2005). Risk of wrist fracture in women is heritable and is influenced by genes that are largely independent of those influencing BMD. J Bone Miner Res 20(1): 67-74.

Araujo, A. B., T. G. Travison, S. S. Harris, M. F. Holick, A. K. Turner and J. B. McKinlay (2007). Race/ethnic differences in bone mineral density in men. Osteoporos Int 18(7): 943-53.

Barnholtz-Sloan, J. S., B. McEvoy, M. D. Shriver and T. R. Rebbeck (2008). Ancestry estimation and correction for population stratification in molecular epidemiologic association studies. Cancer Epidemiol Biomarkers Prev 17(3): 471-7.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series* B, **57**, 289–300.

Bishop, C. M. (2006). Pattern recognition and machine learning. New York, Springer.

Blair, H. C., P. H. Schlesinger, F. P. Ross and S. L. Teitelbaum (1993). Recent advances toward understanding osteoclast physiology. Clin Orthop Relat Res(294): 7-22.

Brieman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). Classification and regression trees. Wadsworth Publishing, Belmont, CA.

Brown, L. B., E. A. Streeten, J. R. Shapiro, D. McBride, A. R. Shuldiner, P. A. Peyser and B. D. Mitchell (2005). Genetic and environmental influences on bone mineral density in pre- and post-menopausal women. Osteoporos Int 16(12): 1849-56.

Brown, L. B., E. A. Streeten, A. R. Shuldiner, L. A. Almasy, P. A. Peyser and B. D. Mitchell (2004). Assessment of sex-specific genetic and environmental effects on bone mineral density. Genet Epidemiol 27(2): 153-61.

Brown, M. A., M. A. Haughton, S. F. Grant, A. S. Gunnell, N. K. Henderson and J. A. Eisman (2001). Genetic control of bone density and turnover: role of the collagen 1alpha1, estrogen receptor, and vitamin D receptor genes. J Bone Miner Res 16(4): 758-64.

Burger, J., M. Kirchner, B. Bramanti, W. Haak and M. G. Thomas (2007). Absence of the lactase-persistence-associated allele in early Neolithic Europeans. Proc Natl Acad Sci U S A 104(10): 3736-41.

Campbell, C. D., E. L. Ogburn, K. L. Lunetta, H. N. Lyon, M. L. Freedman, L. C. Groop, D. Altshuler, K. G. Ardlie and J. N. Hirschhorn (2005). Demonstrating stratification in a European American population. Nat Genet 37(8): 868-72.

Carlson, C. S., M. A. Eberle, M. J. Rieder, J. D. Smith, L. Kruglyak and D. A. Nickerson (2003). Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. Nat Genet 33(4): 518-21.

Carter, D. R., M. L. Bouxsein and R. Marcus (1992). New approaches for interpreting projected bone densitometry data. J Bone Miner Res 7(2): 137-45.

Cauley, J. A., L. Y. Lui, K. L. Stone, T. A. Hillier, J. M. Zmuda, M. Hochberg, T. J. Beck and K. E. Ensrud (2005). Longitudinal study of changes in hip bone mineral density in Caucasian and African-American women. J Am Geriatr Soc 53(2): 183-9.

Chanock, S. J., T. Manolio, M. Boehnke, E. Boerwinkle, D. J. Hunter, G. Thomas, J. N. Hirschhorn, G. Abecasis, D. Altshuler, J. E. Bailey-Wilson, L. D. Brooks, L. R. Cardon, M. Daly, P. Donnelly, J. F. Fraumeni, Jr., N. B. Freimer, D. S. Gerhard, C. Gunter, A. E. Guttmacher, M. S. Guyer, E. L. Harris, J. Hoh, R. Hoover, C. A. Kong, K. R. Merikangas, C. C. Morton, L. J. Palmer, E. G. Phimister, J. P. Rice, J. Roberts, C. Rotimi, M. A. Tucker, K. J. Vogan, S. Wacholder, E. M. Wijsman, D. M. Winn and F. S. Collins (2007). Replicating genotype-phenotype associations. Nature 447(7145): 655-60.

Cheverud, J. M. and E. J. Routman (1995). Epistasis and its contribution to genetic variance components. Genetics 139(3): 1455-61.

Chrischilles, E. A., C. D. Butler, C. S. Davis and R. B. Wallace (1991). A model of lifetime osteoporosis impact. Arch Intern Med 151(10): 2026-32.

Collins, F. S., E. D. Green, A. E. Guttmacher and M. S. Guyer (2003). A vision for the future of genomics research. Nature 422(6934): 835-47.

Collins-Schramm, H. E., B. Chima, T. Morii, K. Wah, Y. Figueroa, L. A. Criswell, R. L. Hanson, W. C. Knowler, G. Silva, J. W. Belmont and M. F. Seldin (2004). Mexican American ancestry-informative markers: examination of population structure and marker characteristics in European Americans, Mexican Americans, Amerindians and Asians. Hum Genet 114(3): 263-71.

Conde L, Vaquerizas JM, Santoyo J, Al-Shahrour F, Ruiz-Llorente S, Robledo M, Dopazo J 2004 PupaSNP Finder: a web tool for finding SNPs with putative effect at transcriptional level. Nucleic Acids Res 32(Web Server issue):W242-8.

Conneely, K. N. and M. Boehnke (2007). So Many Correlated Tests, So Little Time! Rapid Adjustment of P Values for Multiple Correlated Tests. Am J Hum Genet 81(6).

Consortium, T. I. H. (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature 449(7164): 851-61.

Cooper C, Atkinson EJ, Jacobsen EJ, O'Fallon WM, Melton LJ. (1993) Population-based study of survival after osteoporotic factures. Am J Epidemiol 137:1001-1005.

Cornuz, J., D. Feskanich, W. C. Willett and G. A. Colditz (1999). Smoking, smoking cessation, and risk of hip fracture in women. Am J Med 106(3): 311-4.

Deng HW, Chen WM, Recker S, Stegman MR, Li JL, Davies KM, Zhou Y, Deng H, Heaney R, Recker RR. (2000) Genetic determinantion of Colles' fracture and differential bone mass in women with and without Colles' fracture. J Bone Miner Res 15:1243-1252.

Dennison, E., Z. Cole and C. Cooper (2005). Diagnosis and epidemiology of osteoporosis. Curr Opin Rheumatol 17(4): 456-61.

Dequeker, J., J. Nijs, A. Verstraeten, P. Geusens and G. Gevers (1987). Genetic determinants of bone mineral content at the spine and radius: a twin study. Bone 8(4): 207-9.

Devlin, B. and K. Roeder (1999). Genomic control for association studies. Biometrics 55(4): 997-1004.

Devoto, M., L. D. Spotila, D. L. Stabley, G. N. Wharton, H. Rydbeck, J. Korkko, R. Kosich, D. Prockop, A. Tenenhouse and K. Sol-Church (2005). Univariate and bivariate variance component linkage analysis of a whole-genome scan for loci contributing to bone mineral density. Eur J Hum Genet 13(6): 781-8.

Duan, Y., V. De Luca and E. Seeman (1999). Parathyroid hormone deficiency and excess: similar effects on trabecular bone but differing effects on cortical bone. J Clin Endocrinol Metab 84(2): 718-22.

Duan, Y., C. H. Turner, B. T. Kim and E. Seeman (2001). Sexual dimorphism in vertebral fragility is more the result of gender differences in age-related bone gain than bone loss. J Bone Miner Res 16(12): 2267-75.

Duncan, E. L., M. A. Brown, J. Sinsheimer, J. Bell, A. J. Carr, B. P. Wordsworth and J. A. Wass (1999). Suggestive linkage of the parathyroid receptor type 1 to osteoporosis. J Bone Miner Res 14(12): 1993-9.

Easton, D. F., K. A. Pooley, A. M. Dunning, P. D. Pharoah, D. Thompson, D. G. Ballinger, J. P. Struewing, J. Morrison, H. Field, R. Luben, N. Wareham, S. Ahmed, C. S. Healey, R. Bowman, K. B. Meyer, C. A. Haiman, L. K. Kolonel, B. E. Henderson, L. Le Marchand, P. Brennan, S. Sangrajrang, V. Gaborieau, F. Odefrey, C. Y. Shen, P. E. Wu, H. C. Wang, D. Eccles, D. G. Evans, J. Peto, O. Fletcher, N. Johnson, S. Seal, M. R. Stratton, N. Rahman, G. Chenevix-Trench, S. E. Bojesen, B. G. Nordestgaard, C. K. Axelsson, M. Garcia-Closas, L. Brinton, S. Chanock, J. Lissowska, B. Peplonska, H. Nevanlinna, R. Fagerholm, H. Eerola, D. Kang, K. Y. Yoo, D. Y. Noh, S. H. Ahn, D. J. Hunter, S. E. Hankinson, D. G. Cox, P. Hall, S. Wedren, J. Liu, Y. L. Low, N. Bogdanova, P. Schurmann, T. Dork, R. A. Tollenaar, C. E. Jacobi, P. Devilee, J. G. Klijn, A. J. Sigurdson, M. M. Doody, B. H. Alexander, J. Zhang, A. Cox, I.

W. Brock, G. MacPherson, M. W. Reed, F. J. Couch, E. L. Goode, J. E. Olson, H. Meijers-Heijboer, A. van den Ouweland, A. Uitterlinden, F. Rivadeneira, R. L. Milne, G. Ribas, A. Gonzalez-Neira, J. Benitez, J. L. Hopper, M. McCredie, M. Southey, G. G. Giles, C. Schroen, C. Justenhoven, H. Brauch, U. Hamann, Y. D. Ko, A. B. Spurdle, J. Beesley, X. Chen, A. Mannermaa, V. M. Kosma, V. Kataja, J. Hartikainen, N. E. Day, D. R. Cox and B. A. Ponder (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447(7148): 1087-93.

Econs, M. J., D. L. Koller, S. L. Hui, T. Fishburn, P. M. Conneally, C. C. Johnston, Jr., M. Peacock and T. M. Foroud (2004). Confirmation of linkage to chromosome 1q for peak vertebral bone mineral density in premenopausal white women. Am J Hum Genet 74(2): 223-8.

Evans, E. M., K. M. Ross, K. L. Heinrichs, E. McAuley and K. S. Rosengren (2005). Ultrasound of the calcaneus and bone mineral density differs in older black and white women but is not impacted by current physical activity. Osteoporos Int 16(12): 1755-60.

Feskanich, D., W. C. Willett and G. A. Colditz (2003). Calcium, vitamin D, milk consumption, and hip fractures: a prospective study among postmenopausal women. Am J Clin Nutr 77(2): 504-11.

Fisher, S. A. and C. M. Lewis (2008). Power of Genetic Association Studies in the Presence of Linkage Disequilibrium and Allelic Heterogeneity. Hum Hered 66(4): 210-222.

Freimer, N. B. and C. Sabatti (2007). Human genetics: variants in common diseases. Nature 445(7130): 828-30.

Gabriel, S. B., S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S. N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E. S. Lander, M. J. Daly and D. Altshuler (2002). The structure of haplotype blocks in the human genome. Science 296(5576): 2225-9.

Gennari, L., L. Becherini, A. Falchetti, L. Masi, F. Massart and M. L. Brandi (2002). Genetics of osteoporosis: role of steroid hormone receptor gene polymorphisms. J Steroid Biochem Mol Biol 81(1): 1-24.

Gennari, L., L. Becherini, R. Mansani, L. Masi, A. Falchetti, A. Morelli, E. Colli, S. Gonnelli, C. Cepollaro and M. L. Brandi (1999). FokI polymorphism at translation initiation site of the vitamin D receptor gene predicts bone mineral density and vertebral fractures in postmenopausal Italian women. J Bone Miner Res 14(8): 1379-86.

Gennari, L., L. Becherini, L. Masi, R. Mansani, S. Gonnelli, C. Cepollaro, S. Martini, A. Montagnani, G. Lentini, A. M. Becorpi and M. L. Brandi (1998). Vitamin D and estrogen receptor allelic variants in Italian postmenopausal women: evidence of multiple gene contribution to bone mineral density. J Clin Endocrinol Metab 83(3): 939-44.

Grados, F., M. Brazier, S. Kamel, M. Mathieu, N. Hurtebize, M. Maamer, M. Garabedian, J. L. Sebert and P. Fardellone (2003). Prediction of bone mass density variation by bone remodeling markers in postmenopausal women with vitamin D insufficiency treated with calcium and vitamin D supplementation. J Clin Endocrinol Metab 88(11): 5175-9.

Guglielmi, G. and T. F. Lang (2002). Quantitative computed tomography. Semin Musculoskelet Radiol 6(3): 219-27.

Hagino, H. (2004). [Dual-energy X-ray absorptiometry in clinical practice and its trouble spots in bone mineral measurement]. Clin Calcium 14(12): 33-8.

Hannan, M. T., K. L. Tucker, B. Dawson-Hughes, L. A. Cupples, D. T. Felson and D. P. Kiel (2000). Effect of dietary protein on bone loss in elderly men and women: the Framingham Osteoporosis Study. J Bone Miner Res 15(12): 2504-12.

Hastie, T., R. Tibshirani and J. H. Friedman (2001). The elements of statistical learning : data mining, inference, and prediction. New York, Springer.

Hinds, D. A., L. L. Stuve, G. B. Nilsen, E. Halperin, E. Eskin, D. G. Ballinger, K. A. Frazer and D. R. Cox (2005). Whole-genome patterns of common DNA variation in three human populations. Science 307(5712): 1072-9.

Ho, A. Y. and A. W. Kung (2005). Determinants of peak bone mineral density and bone area in young women. J Bone Miner Metab 23(6): 470-5.

Hosmer, W. D., H. K. Genant and W. S. Browner (2002). Fractures before menopause: a red flag for physicians. Osteoporos Int 13(4): 337-41.

Huang, Q. Y., F. H. Xu, H. Shen, L. J. Zhao, H. Y. Deng, Y. J. Liu, V. Dvomyk, T. Conway, K. M. Davies, J. L. Li, Y. Z. Liu, R. R. Recker and H. W. Deng (2004). A second-stage genome scan for QTLs influencing BMD variation. Calcif Tissue Int 75(2): 138-43.

Hui, S. L., D. L. Koller, T. M. Foroud, M. J. Econs, C. C. Johnston and M. Peacock (2006). Heritability of changes in bone size and bone mass with age in premenopausal white sisters. J Bone Miner Res 21(7): 1121-5.

Hustmyer, F. G., G. Liu, C. C. Johnston, J. Christian and M. Peacock (1999). Polymorphism at an Sp1 binding site of COL1A1 and bone mineral density in premenopausal female twins and elderly fracture patients. Osteoporos Int 9(4): 346-50.

Ioannidis, J. P., I. Stavrou, T. A. Trikalinos, C. Zois, M. L. Brandi, L. Gennari, O. Albagha, S. H. Ralston and A. Tsatsoulis (2002). Association of polymorphisms of the estrogen receptor alpha gene with bone mineral density and fracture risk in women: a meta-analysis. J Bone Miner Res 17(11): 2048-60.

Johnson, M. L., G. Gong, W. Kimberling, S. M. Recker, D. B. Kimmel and R. B. Recker (1997). Linkage of a gene causing high bone mass to human chromosome 11 (11q12-13). Am J Hum Genet 60(6): 1326-32.

Jordan, K. M. and C. Cooper (2002). Epidemiology of osteoporosis. Best Pract Res Clin Rheumatol 16(5): 795-806.

Judex, S., R. Garman, M. Squire, L. R. Donahue and C. Rubin (2004). Genetically based influences on the site-specific regulation of trabecular and cortical bone morphology. J Bone Miner Res 19(4): 600-6.

Kalender, W. A. (1992). Effective dose values in bone mineral measurements by photon absorptiometry and computed tomography. Osteoporos Int 2(2): 82-7.

Kammerer, C. M., J. L. Schneider, S. A. Cole, J. E. Hixson, P. B. Samollow, J. R. O'Connell, R. Perez, T. D. Dyer, L. Almasy, J. Blangero, R. L. Bauer and B. D. Mitchell (2003). Quantitative trait loci on chromosomes 2p, 4p, and 13q influence bone mineral density of the forearm and hip in Mexican Americans. J Bone Miner Res 18(12): 2245-52.

Karim, N., D. MacDonald, A. L. Dolan, I. Fogelman, A. S. Wierzbicki and G. Hampson (2008). The relationship between gonadotrophins, gonadal hormones and bone mass in men. Clin Endocrinol (Oxf) 68(1): 94-101.

Kelly, P. J., T. Nguyen, J. Hopper, N. Pocock, P. Sambrook and J. Eisman (1993). Changes in axial bone density with age: a twin study. J Bone Miner Res 8(1): 11-7.

Khodadadyan-Klostermann, C., M. von Seebach, W. R. Taylor, G. N. Duda and N. P. Haas (2004). Distribution of bone mineral density with age and gender in the proximal tibia. Clin Biomech (Bristol, Avon) 19(4): 370-6.

Khosla, S. (2004). Role of hormonal changes in the pathogenesis of osteoporosis in men. Calcif Tissue Int 75(2): 110-3.

Klei, L., D. Luca, B. Devlin and K. Roeder (2008). Pleiotropy and principal components of heritability combine to increase power for association analysis. Genet Epidemiol 32(1): 9-19.

Koay, M. A., P. Y. Woon, Y. Zhang, L. J. Miles, E. L. Duncan, S. H. Ralston, J. E. Compston, C. Cooper, R. Keen, B. L. Langdahl, A. MacLelland, J. O'Riordan, H. A. Pols, D. M. Reid, A. G. Uitterlinden, J. A. Wass and M. A. Brown (2004). Influence of LRP5 polymorphisms on normal variation in BMD. J Bone Miner Res 19(10): 1619-27.

Koller, D. L., L. A. Rodriguez, J. C. Christian, C. W. Slemenda, M. J. Econs, S. L. Hui, P. Morin, P. M. Conneally, G. Joslyn, M. E. Curran, M. Peacock, C. C. Johnston and T. Foroud (1998). Linkage of a QTL contributing to normal variation in bone mineral density to chromosome 11q12-13. J Bone Miner Res 13(12): 1903-8.

Livingston, R. J., A. von Niederhausern, A. G. Jegga, D. C. Crawford, C. S. Carlson, M. J. Rieder, S. Gowrisankar, B. J. Aronow, R. B. Weiss and D. A. Nickerson (2004). Pattern of sequence variation across 213 environmental response genes. Genome Res 14(10A): 1821-31.

Lohman G, Roche F et al. (1996) Dual-Energy X-ray Absorptiometry. Human Kenetics 63-78.

Looker, A. C., H. W. Wahner, W. L. Dunn, M. S. Calvo, T. B. Harris, S. P. Heyse, C. C. Johnston, Jr. and R. Lindsay (1998). Updated data on proximal femur bone mineral levels of US adults. Osteoporos Int 8(5): 468-89.

Masi, L., R. Cimaz, G. Simonini, G. Bindi, S. Stagi, A. Gozzini, C. Malentacchi, M. L. Brandi and F. Falcini (2002). Association of low bone mass with vitamin d receptor gene and calcitonin receptor gene polymorphisms in juvenile idiopathic arthritis. J Rheumatol 29(10): 2225-31.

Melton LJ 3d (1988). Epidemiology of fractures. In: Riggs BL, Melton LJ 3d, eds. Osteoporosis: Etiology, Diagnosis, and Management. New York: Raven;:133-54.

Melton, L. J., 3rd, S. Khosla, E. J. Atkinson, M. K. Oconnor, W. M. Ofallon and B. L. Riggs (2000). Cross-sectional versus longitudinal evaluation of bone loss in men and women. Osteoporos Int 11(7): 592-9.

Melton, L. J., 3rd, E. S. Orwoll and R. D. Wasnich (2001). Does bone density predict fractures comparably in men and women? Osteoporos Int 12(9): 707-9.

Michaelsson, K., H. Melhus, H. Ferm, A. Ahlbom and N. L. Pedersen (2005). Genetic liability to fractures in the elderly. Arch Intern Med 165(16): 1825-30.

Mitchell, B. D., C. M. Kammerer, J. L. Schneider, R. Perez and R. L. Bauer (2003). Genetic and environmental determinants of bone mineral density in Mexican Americans: results from the San Antonio Family Osteoporosis Study. Bone 33(5): 839-46.

Montpetit, A., M. Nelis, P. Laflamme, R. Magi, X. Ke, M. Remm, L. Cardon, T. J. Hudson and A. Metspalu (2006). An evaluation of the performance of tag SNPs derived from HapMap in a Caucasian population. PLoS Genet 2(3): e27.

Mueller, J. C., E. Lohmussaar, R. Magi, M. Remm, T. Bettecken, P. Lichtner, S. Biskup, T. Illig, A. Pfeufer, J. Luedemann, S. Schreiber, P. Pramstaller, I. Pichler, G. Romeo, A. Gaddi, A. Testa, H. E. Wichmann, A. Metspalu and T. Meitinger (2005). Linkage disequilibrium patterns and tag SNP transferability among European populations. Am J Hum Genet 76(3): 387-98.

Ng, M. Y., P. C. Sham, A. D. Paterson, V. Chan and A. W. Kung (2006). Effect of environmental factors and gender on the heritability of bone mineral density and bone size. Ann Hum Genet 70(Pt 4): 428-38.

Nyholt, D. R. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet 74(4): 765-9.

Orwoll, E., J. B. Blank, E. Barrett-Connor, J. Cauley, S. Cummings, K. Ensrud, C. Lewis, P. M. Cawthon, R. Marcus, L. M. Marshall, J. McGowan, K. Phipps, S. Sherman, M. L. Stefanick and K. Stone (2005). Design and baseline characteristics of the osteoporotic fractures in men (MrOS) study--a large observational study of the determinants of fracture in older men. Contemp Clin Trials 26(5): 569-85.

Ota, N., S. C. Hunt, T. Nakajima, T. Suzuki, T. Hosoi, H. Orimo, Y. Shirai and M. Emi (1999). Linkage of interleukin 6 locus to human osteopenia by sibling pair analysis. Hum Genet 105(3): 253-7.

Ota, N., S. C. Hunt, T. Nakajima, T. Suzuki, T. Hosoi, H. Orimo, Y. Shirai and M. Emi (2000). Linkage of human tumor necrosis factor-alpha to human osteoporosis by sib pair analysis. Genes Immun 1(4): 260-4.

Ota, N., T. Nakajima, Y. Ezura, H. Iwasaki, T. Suzuki, T. Hosoi, H. Orimo, S. Inoue, H. Ito and M. Emi (2002). Association of a single nucleotide variant in the human tumour necrosis factor alpha promoter region with decreased bone mineral density. Ann Hum Biol 29(5): 550-8.

Ota, N., T. Nakajima, I. Nakazawa, T. Suzuki, T. Hosoi, H. Orimo, S. Inoue, Y. Shirai and M. Emi (2001). A nucleotide variant in the promoter region of the interleukin-6 gene associated with decreased bone mineral density. J Hum Genet 46(5): 267-72.

Patterson, N., A. L. Price and D. Reich (2006). Population structure and eigenanalysis. PLoS Genet 2(12): e190.

Pocock, N. A., J. A. Eisman, J. L. Hopper, M. G. Yeates, P. N. Sambrook and S. Eberl (1987). Genetic determinants of bone mass in adults. A twin study. J Clin Invest 80(3): 706-10.

Pothiwala, P., E. M. Evans and K. M. Chapman-Novakofski (2006). Ethnic variation in risk for osteoporosis among women: a review of biological and behavioral factors. J Womens Health (Larchmt) 15(6): 709-19.

Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick and D. Reich (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38(8): 904-9.

Pritchard, J. K., M. Stephens and P. Donnelly (2000). Inference of population structure using multilocus genotype data. Genetics 155(2): 945-59.

Raisz, L. G. (1997). The osteoporosis revolution. Ann Intern Med 126(6): 458-62.

Richards, J. B., L. Joseph, K. Schwartzman, N. Kreiger, A. Tenenhouse and D. Goltzman (2006). The effect of cyclooxygenase-2 inhibitors on bone mineral density: results from the Canadian Multicentre Osteoporosis Study. Osteoporos Int 17(9): 1410-9.

Richards, J. B., F. Rivadeneira, M. Inouye, T. M. Pastinen, N. Soranzo, S. G. Wilson, T. Andrew, M. Falchi, R. Gwilliam, K. R. Ahmadi, A. M. Valdes, P. Arp, P. Whittaker, D. J. Verlaan, M. Jhamai, V. Kumanduri, M. Moorhouse, J. B. van Meurs, A. Hofman, H. A. Pols, D. Hart, G. Zhai, B. S. Kato, B. H. Mullin, F. Zhang, P. Deloukas, A. G. Uitterlinden and T. D. Spector (2008). Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. Lancet 371(9623): 1505-12.

Riggs, B. L., S. Khosla and L. J. Melton, 3rd (2002). Sex steroids and the construction and conservation of the adult skeleton. Endocr Rev 23(3): 279-302.

Riggs, B. L., L. J. Melton Iii, 3rd, R. A. Robb, J. J. Camp, E. J. Atkinson, J. M. Peterson, P. A. Rouleau, C. H. McCollough, M. L. Bouxsein and S. Khosla (2004). Population-based study of age and sex differences in bone volumetric density, size, geometry, and structure at different skeletal sites. J Bone Miner Res 19(12): 1945-54.

Rinaldo, A., S. A. Bacanu, B. Devlin, V. Sonpar, L. Wasserman and K. Roeder (2005). Characterization of multilocus linkage disequilibrium. Genet Epidemiol 28(3): 193-206.

Risch, N. and K. Merikangas (1996). The future of genetic studies of complex human diseases. Science 273(5281): 1516-7.

Ritchie, M. D., L. W. Hahn, N. Roodi, L. R. Bailey, W. D. Dupont, F. F. Parl and J. H. Moore (2001). Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 69(1): 138-47.

Roeder K, Bacanu SA, Sonpar V, Zhang X, Devlin B 2005 Analysis of single-locus tests to detect gene/disease associations. Genet Epidemiol 28(3):207-19.

Rubin, C. J., H. Brandstrom, D. Wright, S. Kerje, U. Gunnarsson, K. Schutz, R. Fredriksson, P. Jensen, L. Andersson, C. Ohlsson, H. Mallmin, S. Larsson and A. Kindmark (2007). Quantitative trait loci for BMD and bone strength in an intercross between domestic and wildtype chickens. J Bone Miner Res 22(3): 375-84.

Salari, K., S. Choudhry, H. Tang, M. Naqvi, D. Lind, P. C. Avila, N. E. Coyle, N. Ung, S. Nazario, J. Casal, A. Torres-Palacios, S. Clark, A. Phong, I. Gomez, H. Matallana, E. J. Perez-Stable, M. D. Shriver, P. Y. Kwok, D. Sheppard, W. Rodriguez-Cintron, N. J. Risch, E. G. Burchard and E. Ziv (2005). Genetic admixture and asthma-related phenotypes in Mexican American and Puerto Rican asthmatics. Genet Epidemiol 29(1): 76-86.

Sawyer, S. L., N. Mukherjee, A. J. Pakstis, L. Feuk, J. R. Kidd, A. J. Brookes and K. K. Kidd (2005). Linkage disequilibrium patterns vary substantially among populations. Eur J Hum Genet 13(5): 677-86.

Saxena, R., B. F. Voight, V. Lyssenko, N. P. Burtt, P. I. de Bakker, H. Chen, J. J. Roix, S. Kathiresan, J. N. Hirschhorn, M. J. Daly, T. E. Hughes, L. Groop, D. Altshuler, P. Almgren, J. C. Florez, J. Meyer, K. Ardlie, K. Bengtsson Bostrom, B. Isomaa, G. Lettre, U. Lindblad, H. N. Lyon, O. Melander, C. Newton-Cheh, P. Nilsson, M. Orho-Melander, L. Rastam, E. K. Speliotes, M. R. Taskinen, T. Tuomi, C. Guiducci, A. Berglund, J. Carlson, L. Gianniny, R. Hackett, L. Hall, J. Holmkvist, E. Laurila, M. Sjogren, M. Sterner, A. Surti, M. Svensson, R. Tewhey, B. Blumenstiel, M. Parkin, M. Defelice, R. Barry, W. Brodeur, J. Camarata, N. Chia, M. Fava, J. Gibbons, B. Handsaker, C. Healy, K. Nguyen, C. Gates, C. Sougnez, D. Gage, M. Nizzari, S. B. Gabriel, G. W. Chirn, Q. Ma, H. Parikh, D. Richardson, D. Ricke and S. Purcell (2007). Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316(5829): 1331-6.

Searle SR (1987) Linear models for unbalanced data. Wiley and Sons, Inc., New York, NY.

Shen, H., Y. Y. Zhang, J. R. Long, F. H. Xu, Y. Z. Liu, P. Xiao, L. J. Zhao, D. H. Xiong, Y. J. Liu, V. Dvornyk, S. Rocha-Sanchez, P. Y. Liu, J. L. Li, T. Conway, K. M. Davies, R. R. Recker and H. W. Deng (2004). A genome-wide linkage scan for bone mineral density in an extended sample: evidence for linkage on 11q23 and Xq27. J Med Genet 41(10): 743-51.

Shriver, M. D., R. Mei, E. J. Parra, V. Sonpar, I. Halder, S. A. Tishkoff, T. G. Schurr, S. I. Zhadanov, L. P. Osipova, T. D. Brutsaert, J. Friedlaender, L. B. Jorde, W. S. Watkins, M. J. Bamshad, G. Gutierrez, H. Loi, H. Matsuzaki, R. A. Kittles, G. Argyropoulos, J. R. Fernandez, J. M. Akey and K. W. Jones (2005). Large-scale SNP analysis reveals clustered and continuous patterns of human genetic variation. Hum Genomics 2(2): 81-9.

Sidak, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions. Journal of th.e American Statistical Association 62: 626-633.

Silverberg, S. J., F. Gartenberg, T. P. Jacobs, E. Shane, E. Siris, R. B. Staron, D. J. McMahon and J. P. Bilezikian (1995). Increased bone mineral density after parathyroidectomy in primary hyperparathyroidism. J Clin Endocrinol Metab 80(3): 729-34.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. Biometrika 73(3): 751-754.

Smigielski, E. M., K. Sirotkin, M. Ward and S. T. Sherry (2000). dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Res 28(1): 352-5.

Smith, M. W., J. A. Lautenberger, H. D. Shin, J. P. Chretien, S. Shrestha, D. A. Gilbert and S. J. O'Brien (2001). Markers for mapping by admixture linkage disequilibrium in African American and Hispanic populations. Am J Hum Genet 69(5): 1080-94.

Snel, B., G. Lehmann, P. Bork and M. A. Huynen (2000). STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res 28(18): 3442-4.

Spielman, R. S., R. E. McGinnis and W. J. Ewens (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52(3): 506-16.

Strotmeyer, E. S., J. A. Cauley, A. V. Schwartz, M. C. Nevitt, H. E. Resnick, J. M. Zmuda, D. C. Bauer, F. A. Tylavsky, N. de Rekeneire, T. B. Harris and A. B. Newman (2004). Diabetes is associated independently of body composition with BMD and bone volume in older white and black men and women: The Health, Aging, and Body Composition Study. J Bone Miner Res **19**(7): 1084-91.

Styrkarsdottir, U., B. V. Halldorsson, S. Gretarsdottir, D. F. Gudbjartsson, G. B. Walters, T. Ingvarsson, T. Jonsdottir, J. Saemundsdottir, J. R. Center, T. V. Nguyen, Y. Bagger, J. R. Gulcher, J. A. Eisman, C. Christiansen, G. Sigurdsson, A. Kong, U. Thorsteinsdottir and K. Stefansson (2008). Multiple genetic loci for bone mineral density and fractures. N Engl J Med 358(22): 2355-65.

Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander and J. P. Mesirov (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102(43): 15545-50.

Szulc, P., P. Garnero, B. Claustrat, F. Marchand, F. Duboeuf and P. D. Delmas (2002). Increased bone resorption in moderate smokers with low body weight: the Minos study. J Clin Endocrinol Metab 87(2): 666-74.

Taaffe, D. R., T. F. Lang, T. Fuerst, J. A. Cauley, M. C. Nevitt and T. B. Harris (2003). Sex- and race-related differences in cross-sectional geometry and bone density of the femoral mid-shaft in older adults. Ann Hum Biol 30(3): 329-46.

Takacs, I., D. L. Koller, M. Peacock, J. C. Christian, W. E. Evans, S. L. Hui, P. M. Conneally, C. C. Johnston, Jr., T. Foroud and M. J. Econs (2000). Sib pair linkage and association studies between bone mineral density and the interleukin-6 gene locus. Bone 27(1): 169-73.

Takacs, I., D. L. Koller, M. Peacock, J. C. Christian, S. L. Hui, P. M. Conneally, C. C. Johnston, Jr., T. Foroud and M. J. Econs (1999). Sibling pair linkage and association studies between bone mineral density and the insulin-like growth factor I gene locus. J Clin Endocrinol Metab 84(12): 4467-71.

Tantoso, E., Y. Yang and K. B. Li (2006). How well do HapMap SNPs capture the untyped SNPs? BMC Genomics 7: 238.

Tuck, S. P., M. S. Pearce, D. J. Rawlings, F. N. Birrell, L. Parker and R. M. Francis (2005). Differences in bone mineral density and geometry in men and women: the Newcastle Thousand Families Study at 50 years old. Br J Radiol 78(930): 493-8.

Tucker, K. L., K. Morita, N. Qiao, M. T. Hannan, L. A. Cupples and D. P. Kiel (2006). Colas, but not other carbonated beverages, are associated with low bone mineral density in older women: The Framingham Osteoporosis Study. Am J Clin Nutr 84(4): 936-42.

van den Beld, A. W., F. H. de Jong, D. E. Grobbee, H. A. Pols and S. W. Lamberts (2000). Measures of bioavailable serum testosterone and estradiol and their relationships with muscle strength, bone density, and body composition in elderly men. J Clin Endocrinol Metab 85(9): 3276-82.

Visser, M., B. H. Goodpaster, S. B. Kritchevsky, A. B. Newman, M. Nevitt, S. M. Rubin, E. M. Simonsick and T. B. Harris (2005). Muscle mass, muscle strength, and muscle fat infiltration as predictors of incident mobility limitations in well-functioning older persons. J Gerontol A Biol Sci Med Sci 60(3): 324-33.

Wang, K., M. Li and M. Bucan (2007). Pathway-Based Approaches for Analysis of Genomewide Association Studies. Am J Hum Genet 81(6).

Wang, L., Y. J. Liu, P. Xiao, H. Shen, H. Y. Deng, C. J. Papasian, B. M. Drees, J. J. Hamilton, R. R. Recker and H. W. Deng (2007). Chromosome 2q32 may harbor a QTL affecting BMD variation at different skeletal sites. J Bone Miner Res 22(11): 1672-8.

Weedon, M. N., G. Lettre, R. M. Freathy, C. M. Lindgren, B. F. Voight, J. R. Perry, K. S. Elliott, R. Hackett, C. Guiducci, B. Shields, E. Zeggini, H. Lango, V. Lyssenko, N. J. Timpson, N. P. Burtt, N. W. Rayner, R. Saxena, K. Ardlie, J. H. Tobias, A. R. Ness, S. M. Ring, C. N. Palmer, A. D. Morris, L. Peltonen, V. Salomaa, G. Davey Smith, L. C. Groop, A. T. Hattersley, M. I. McCarthy, J. N. Hirschhorn and T. M. Frayling (2007). A common variant of HMGA2 is associated with adult and childhood height in the general population. Nat Genet 39(10): 1245-50.

Weiss, L. A., E. Barrett-Connor, D. von Muhlen and P. Clark (2006). Leptin predicts BMD and bone resorption in older women but not older men: the Rancho Bernardo study. J Bone Miner Res 21(5): 758-64.

Wilkins, C. H. and S. J. Birge (2005). Prevention of osteoporotic fractures in the elderly. Am J Med 118(11): 1190-5.

Willaert, A., I. Van Pottelbergh, H. Zmierczak, S. Goemaere, J. M. Kaufman, A. De Paepe and P. Coucke (2008). A genome-wide linkage scan for low spinal bone mineral density in a single extended family confirms linkage to 1p36.3. Eur J Hum Genet 16(8): 970-6.

Yang, Q., J. Cui, I. Chazaro, L. A. Cupples and S. Demissie (2005). Power and type I error rate of false discovery rate approaches in genome-wide association studies. BMC Genet 6 Suppl 1: S134.

Yongchao Ge, S. D., Terence P. Speed (2003). Resampling-based Multiple Testing for Microarray Data Analysis. Sociedad de Estadistica e Investigacian Operativa Test 12(1): 1-77.

Yu, K., Z. Wang, Q. Li, S. Wacholder, D. J. Hunter, R. N. Hoover, S. Chanock and G. Thomas (2008). Population substructure and control selection in genome-wide association studies. PLoS ONE 3(7): e2551.

Zaykin, D. V. and L. A. Zhivotovsky (2005). Ranks of genuine associations in whole-genome scans. Genetics 171(2): 813-23.

Zeger, S. L. and K. Y. Liang (1986). Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42(1): 121-30.

Zeggini, E., M. N. Weedon, C. M. Lindgren, T. M. Frayling, K. S. Elliott, H. Lango, N. J. Timpson, J. R. Perry, N. W. Rayner, R. M. Freathy, J. C. Barrett, B. Shields, A. P. Morris, S. Ellard, C. J. Groves, L. W. Harries, J. L. Marchini, K. R. Owen, B. Knight, L. R. Cardon, M. Walker, G. A. Hitman, A. D. Morris, A. S. Doney, M. I. McCarthy and A. T. Hattersley (2007). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316(5829): 1336-41.

Zhang, F., P. Xiao, F. Yang, H. Shen, D. H. Xiong, H. Y. Deng, C. J. Papasian, B. M. Drees, J. J. Hamilton, R. R. Recker and H. W. Deng (2008). A whole genome linkage scan for QTLs underlying peak bone mineral density. Osteoporos Int 19(3): 303-10.

Zhang, Y., S. S. Padalecki, A. R. Chaudhuri, E. De Waal, B. A. Goins, B. Grubbs, Y. Ikeno, A. Richardson, G. R. Mundy and B. Herman (2007). Caspase-2 deficiency enhances aging-related traits in mice. Mech Ageing Dev 128(2): 213-21.

Zhao T, Chang LW, McLeod HL, Stormo GD 2004 PromoLign: a database for upstream region analysis and SNPs. Hum Mutat 23(6):534-9.

Zmuda, J. M., J. A. Cauley, M. E. Danielson, T. M. Theobald and R. E. Ferrell (1999). Vitamin D receptor translation initiation codon polymorphism and markers of osteoporotic risk in older African-American women. Osteoporos Int 9(3): 214-9.

Zmuda, J. M., J. E. Eichner, R. E. Ferrell, D. C. Bauer, L. H. Kuller and J. A. Cauley (1998). Genetic variation in alpha 2HS-glycoprotein is related to calcaneal broadband ultrasound attenuation in older women. Calcif Tissue Int 63(1): 5-8.