Evaluating Epidemiologic and Genetic Risk Factors for Uterine Fibroid Characteristics

By

Michael Joseph Bray

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Human Genetics

May 11, 2018

Nashville, Tennessee

Approved:

Digna R. Velez Edwards, Ph.D.
Nancy J. Cox, Ph.D.
Todd L. Edwards, Ph.D.
Bingshan Li, Ph.D.
David C. Samuels, Ph.D.
Melissa F. Wellons, M.D.

I dedicate this thesis to my parents, Bill and Susan, to my sisters, Staci, Keely, and Mollie, and to my husband, Siripong. I love my family from the bottom of my heart with all of my heart.

ACKNOWLEDGMENTS

These studies were funded by the National Institutes of Health (NIH) grants (R01HD074711, R03HD078567, and R01HD093671) to Digna R. Velez Edwards and by the Human Genetic Training Grant (5T32GM080178) and the VICTR Training Grant (6TL1TR000447) to Michael J. Bray. The VICTR grant (VR20169) to Michael J. Bray helped fund the abstraction of risk factors for fibroids from the Synthetic Derivative.

The thesis described was supported by CTSA award No. UL1TR000445 from the National Center for Advancing Translational Sciences. Its contents are solely the responsibility of the authors and do not necessarily represent official views of the National Center for Advancing Translational Sciences or the National Institutes of Health.

The Coronary Artery Risk Development in Young Adults Study (CARDIA) is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the University of Alabama at Birmingham (HHSN268201300025C & HHSN268201300026C), Northwestern University (HHSN268201300027C), University of Minnesota (HHSN268201300028C), Kaiser Foundation Research Institute (HHSN268201300029C), and Johns Hopkins University School of Medicine (HHSN268200900041C). CARDIA is also partially supported by the Intramural Research Program of the National Institute on Aging (NIA) and an intra-agency agreement between NIA and NHLBI (AG0005). This manuscript has been reviewed by CARDIA for scientific content. The CARDIA Women's Study was supported by the NHLBI (R01-HL-065611). Genotyping was funded as part of the NHLBI Candidate-gene Association Resource (N01-HC-65226) and the NHGRI Gene Environment Association Studies (GENEVA) (U01-HG004729, U01-HG04424, and U01-HG004446).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AA	-	African American
ACB	-	African Caribbeans in Barbados
AFR	-	African
AMR	-	Ad Mixed American
aOR	-	adjusted odds ratio
ARMDS	-	American Registry of Diagnostic Medical Sonographers
ASW	-	Americans of African Ancestry in SW USA
BET1L	-	blocked early in transport 1 homolog
BioVU	-	biorepository at Vanderbilt University
BMI	-	body mass index
BP3	-	protein binding 3
CPT	-	current procedure terminology
CARDIA	-	Coronary Artery Risk Development in Young Adults
CARe	-	Candidate Gene Association Resource
CI	-	confidence interval
cM	-	centimorgan
cm	-	centimeter
cm^3	-	centimeter cubed
CT	-	computed tomography
CWS	-	CARDIA Women's Study
CYTH4	-	cytohesin 4
DZ	-	dizygotic
EA	-	European American
EAS	-	East Asian
EHR	-	electronic health record
EUR	-	European
G6PD	-	glucose-6-phosphate dehydrogenase
GCTA	-	Genome-wide Complex Trait Analysis
GnRH	-	gonadotropin-releasing hormone
GRM	-	genetic relationship matrix
GTEx	-	Genotype-Tissue Expression
GWAS	-	genome-wide association study
HMGA	-	high mobility group A
HMGA1	-	high mobility group AT-hook 1
HMGA2	-	high mobility group AT-hook 2
IBD	-	identity by descent
ICD-9	-	international classification of diseases, 9^{th} revision
IPMK	-	inositol polyphosphate multikinase
IRR	-	incidence rate ratio
LAMP-ANC	-	local ancestry in admixed populations - ancestral
LOH	-	Loss of heterozygosity
LZTS2	-	leucine zipper tumor suppressor 2
MAF	-	minor allele frequency
Mb	-	mega bases

MED12	-	mediator complex subunit 12
MI	-	multiple imputation
MLM	-	mixed linear model
MRgFUS	-	magnetic resonance-guided focused ultrasound
MZ	-	monozygotic
OR	-	odds ratio
PC	-	Principal component
PC1	-	Principal component 1
PC2	-	Principal component 2
PLXNA4	-	plexin A4
PSMD6	-	proteasome 26S subunit, non-ATPase 6
QQ	-	Quantile-quantile
REML	-	restricted maximum likelihood approach
RPKM	-	reads per kilobase of transcript per million mapped reads
SAS	-	South Asian
SAP130	-	sin3A association protein 130
SNP	-	single nucleotide polymorphism
SNW1	-	SNW domain containing 1
SW	-	southwest
TNRC6B	-	trinucleotide repeat containing 6b
Tsc2	-	tuberous sclerosis complex 2
UK	-	United Kingdom
UAE	-	uterine artery embolization
VUMC	-	Vanderbilt University Medical Center
WHO	-	World Health Organization

OVERVIEW

Uterine leiomyoma (fibroids), benign tumors of the uterus, are the most common female pelvic tumor (1). The majority of U.S. women have at least one uterine fibroid by the age of menopause (2). Symptoms of fibroids include abnormal bleeding during menstruation, increased urinary frequency, and pelvic pressure (3, 4). Fibroids are a major cause of hysterectomies (5), primarily to manage symptoms and costs in the U.S. are up to $34.4 billion (2010 dollars) annually in healthcare, treatment, and work loss costs (6).

Fibroids are highly heterogeneous, with some women developing a single small fibroid while other women develop multiple and/or large fibroids. African American (AA) women for example have an approximately two- to three-fold higher risk of fibroids when compared to European American (EA) women (7) and also have more numerous and larger fibroids (2). Additionally, AAs are two times more likely than EAs to receive surgical treatments for fibroids such as a hysterectomy (8). Unfortunately, most research on fibroids to date has not evaluated risk factors for specific fibroid characteristics. The purpose of this thesis is to provide a deeper understanding on both epidemiology and genetic risk factors of fibroid characteristics, fibroid number (single vs. multiple), volume of largest fibroid, and largest dimension of all fibroid measurements.

Chapter I provides a broad overview of the epidemiology on uterine fibroids and fibroid characteristics. This encompasses the morphology, incidence, and progression of fibroids. Detailed descriptions of symptoms associated with fibroids and subsequent treatments for

fibroids are given. Numerous risk factors for fibroids and fibroid characteristics are described including age (2), menarche (9), body mass index (BMI) (kg/m^2) (10), race (11-14), parity (15, 16), and type 2 diabetes (17, 18). Finally, fibroids are heritable and have a genetic component (19, 20). Current advances, understandings, and limitations of fibroid genetic research are listed.

Chapter II focuses on identifying physical and clinical features of women that are associated with fibroid size and number. We accomplished this by using a validated phenotyping algorithm (21) to identify 2,302 fibroid cases from the Synthetic Derivative (22). The Synthetic Derivative consists of de-identified electronic health records (EHRs) of patients who visit the Vanderbilt University Medical Center (VUMC) network (22). After abstracting fibroid characteristics and candidate risk factors using a combination of manual review and phenotyping algorithms, we performed regression analyses using exposures age, BMI, race (black or white), type 2 diabetes, and number of living children (proxy to parity). We assessed for potential effect measure modification by race and both age and BMI using a likelihood ratio test.

Chapter III estimates the heritability of fibroids using genome-wide association study (GWAS) data. In this chapter, we estimate and characterize the heritability of image-confirmed fibroid cases and controls from EA women using GWAS data. Our study consisted of women from BioVU. We estimated the genetic variance explained by all single nucleotide polymorphisms (SNPs) using Genome-wide Complex Trait Analysis (GCTA) (23) for fibroid risk and size on imputed data. The final models were adjusted for age, BMI, and five PCs. We also characterized the heritability of fibroids by partitioning the genetic variance into

chromosomes and into genic and intergenic regions. Lastly, we characterized fibroid heritability after excluding genomic regions previously implicated with fibroid risk.

Chapter IV focuses on the effects that African ancestry has on fibroid number and size (24). Using 609 AAs with image- or surgery confirmed fibroids from BioVU and the Coronary Artery Risk Development in Young Adults (CARDIA) study, we estimated local and global African and European ancestry from GWAS data. Next, we performed linear and logistic regression on global and local ancestry for each fibroid characteristics for BioVU and CARDIA. Meta-analyses comparing global or local ancestry across CARDIA and BioVU AAs were then performed for each outcome. Suggestive peaks were further evaluated for single SNP associations in separate analyses by characteristics. Lastly, we performed local ancestry analysis adjusting for the most significant single SNP association within the admixture mapping peak regions in order to identify any genetic/imputed SNPs that explain the ancestry peaks.

Chapter V summarizes Chapters II-IV, highlighting main strengths and limitations of our results. We review how our findings contribute to epidemiologic and genetic risk factors for fibroid characteristics among AA and EA women. In addition, our research gives insight for future directions for genetic studies on fibroid heritability. Lastly, we discuss future directions involving characteristics and heritability of fibroids.

CHAPTER 1

INTRODUCTION TO UTERINE FIBROIDS AND FIBROID CHARACTERISTICS

Fibroid Morphology

Uterine fibroids, or leiomyomas, are the most common female pelvic tumor (1), are located in the myometrium of the uterus (25, 26), and very rarely become malignant (26). Fibroids are composed of smooth muscle fibers and fibrous connective tissue and are roundish in shape (26). Tumors can range in size from a few millimeters to many centimeters and can form in any location within the uterus (26). Fibroid subtypes are denoted by location within the uterus: submucosal fibroids are located along the inside wall, intramural fibroids are located within the wall, subserosal are located on the outside wall, and pedunculated fibroids are attached to the uterine wall via a stalk (25). Fibroids are monoclonal in origin, likely rising from a single cell (27). In a study by Linder and Gartler (1965), the authors ascertained the expression of X-linked glucose-6-phosphate dehydrogenase (*G6PD*) of 86 normal myometrium tissue samples and 27 fibroids from five women who were heterozygous for the *G6PD* loci (A and B). The authors found that nearly all the normal myometrial samples expressed both A and B while all the fibroid tissues expressed exclusively *G6PD* variant A or B (27). In addition to fibroids being monoclonal in origin, fibroids are a heterogeneous disease where some women will develop a small/single fibroid, and other women will large/numerous fibroids.

Fibroid Epidemiology

Fibroid Incidence and Prevalence

Prevalence estimates from pathology reports place fibroid prevalence between 20 to 77% (1). In 1933, Graves published the first prevalence estimate for fibroid presence by examining autopsies of women over 35 years old and observed that 20% of women had fibroids (1). In 1990, Cramer and Patel observed that 77% women have fibroids after examining 100 consecutive hysterectomy specimens from serial sectioning at two mm intervals (1, 28). More recently in 2003, Baird et al. assessed the cumulative incidence of fibroids via ultrasounds using 1,364 black and white US women aged 35 to 49 years who were part of an urban health plan (2). Baird et al. found that the cumulative incidence of fibroids was above 80% for black women and approximately 70% for white women by age 49 (2). In addition, the authors estimated the prevalence of clinically relevant fibroids if a women had at least one submucosal fibroid, a large uterus (≥10 centimeters [cm] in length), or at least one large fibroid (≥4 cm in largest dimension) (2). The authors found that approximately 50% of black women and 35% of white women had a clinically relevant fibroid by age 49 (2). Interestingly, the authors noted that 51% of women without a previous diagnosis of fibroids observed a fibroid during ultrasound assessment (2), noting the need for imaging studies on fibroid risk. Lastly, the authors also found that black women were more likely to have multiple fibroids and larger newly detected fibroids than white women (2).

Public Health Impact of Fibroids

Approximately 25% of fibroids are symptomatic with symptoms depending on location, number, and size of the leiomyomas (25, 29). Symptoms of fibroids include the following:

- menorrhagia - prolonged and heavy bleeding during menstruation
- anemia - lack of red blood cells
- pain
- pelvic pressure
- increased urinary frequency
- constipation
- pregnancy complication
- premature labor
- infertility (25)

Symptoms are connected with fibroid location, number, and size (25, 29). For example, submucosal fibroids are accompanied with menorrhagia while large fibroids are accompanied with pelvic pressure of the lower abdomen (25). In addition, large fibroids have previously been associated with preterm premature rupture of membranes, while multiple fibroids have previously been associated with cesarean delivery and preterm birth (30).

Fibroids costs the US approximately 5.9 to 34.4 billion dollars annually for treatment, healthcare, and lost work costs (6), largely to manage symptoms. Surgical treatment options include myomectomies and hysterectomies (25, 31, 32). Myomectomies include removing the fibroid(s) while preserving the uterus, and hysterectomies involve removal of the fibroids with

the uterus, thus removing the chance of fibroid recurrence. In addition, fibroids are most common indication of hysterectomy in the US (5). Less invasive procedures include uterine artery embolization (UAE) and magnetic resonance-guided focused ultrasound (MRgFUS) (14, 31, 32). The UAE procedure includes injecting occluding agents into arteries of the uterus (32) to reduce or stop blood flow to a fibroid, thus reducing its size. The MRgFUS procedure uses focused ultrasounds to selectively cause thermal damage and necrosis of fibroid tissue (33). Lastly, fibroids, being composed of uterine tissue, have a hormonal component (14), and there are short-tern hormonal treatment options. Among the most used drugs, gonadotropin-releasing hormone (GnRH) agonists cause a temporary menopause among premenopausal women, thus reducing the size of fibroids and causing amenorrhoea (25). The GnRH agonists, however, are only approved for temporary use because of long-term side effects, and upon GnRH agonist discontinuation, bleeding symptoms and fibroid size returns to pretreatment levels (25).

Fibroid Risk Factors

Selected studies on exposures of fibroid risk are shown in Table 1.1.

Age

Cumulative incidence and clinically relevant prevalence of fibroids increases with age during reproductive years (2). Since fibroids have a hormonal component responding to estrogen and progesterone, fibroids have not been documented in prepubescent girls (25) likely forming after menarche. Prevalence of fibroids needing surgery decreases after menopause (34). In 1990, Cramer and Patel (1990) performed serial sectioning of uterine tissue from 100 consecutive

hysterectomies and found that postmenopausal women were found to have smaller and fewer fibroids compared to premenopausal women (28).

Menarche

Age at menarche has been inversely associated with fibroid risk (35). Having an early age at menarche could lead to an increase in menstrual cycles (35). The increase in uterine cell divisions from menstruating could lead to a higher mutation load in genes controlling myometrial proliferation (35) leading to an increase in fibroid risk. Using a cohort of 95,061 women from the *Nurses' Health Study II cohort*, Marshall et al. found that compared to women who were 12 years at menarche, those who were ≤10 years old at menarche had a higher risk of fibroids (35). In addition, the authors found that women who were ≥16 years at menarche had a decreased fibroid risk (35). In another study by Velez Edwards et al. (2013), the authors performed association analyses between age at menarche and fibroid risk, number, size, and location using 540 fibroid cases and 4,483 control from the *Right From the Start study*, a community-based pregnancy cohort (9). Velez Edwards et al. observed that 1-year increase in age at menarche was inversely associated with fibroid risk (9). Furthermore, the authors found that early age at menarche was associated with increased fibroid number and size (9).

BMI

Increasing BMI and obesity have been associated with fibroid risk (34). Obesity decreases the metabolism of estradiol to inactive metabolites, which could produce a

hyperestrogenic state (34, 36). This increase in estrogen could lead to an increased fibroid risk or to more numerous and larger fibroids since fibroids have a hormonal component responding to estrogen (25). Using women from the Oxford Family Planning Association study, Ross et al. (1986) observed that women had a 21% increase risk for fibroids for every 10 kg increase using 1,070 women (535 cases) (37). Similar trends were observed when using BMI instead of weight (37). In another study, Marshal et al. (1998) performed association analyses between BMI and fibroid risk using 94,095 premenopausal from the *Nurses' Health Study II* (38). The authors observed that adult BMI is associated with increased fibroid risk (38). The associations between BMI and fibroid characteristics are less known. In one study, Dandolu et al. (2010) compared BMI with fibroid risk and largest dimension using 873 women (533 cases) who had hysterectomies at the Temple University Hospital between 1995-2002 (10). Fibroid size was dichotomized to larger or smaller than 2 cm (10). The authors found that while BMI positively correlated with fibroid risk ($p < 0.0001$), BMI was not correlated with fibroid largest dimension ($p = 0.11$) (10).

Race

Black race has been associated with fibroids in many studies (2, 7, 25). In 1997, Marshall et al. estimated the incidence rate of fibroids using 95,061 black and white women from the *Nurses' Health Study II* (7). The authors observed that the incidence of fibroids was two- to three-fold higher among black women compared to white women (7). In 2003, Baird et al. assessed the cumulative incidence and clinically relevant prevalence of fibroids by ultrasounds using 1,364 black and white women from ages 35 to 49 years old who were part of an urban

health plan (2). Baird et al. observed that the cumulative incidence of fibroids was higher at every examined age for black women compared to white women (2). In addition, the authors estimated the prevalence of clinically relevant fibroids if a women had at least one submucosal fibroid, a large uterus (≥10 cm in length), or at least one large fibroid (≥4 cm in largest dimension) (2). The authors again observed that black women had a higher prevalence of clinically relevant fibroids at every age compared to white women (2). Furthermore, authors also found that black women were more likely to have multiple fibroids and larger newly detected fibroids than white women (2).

In another study by Moorman et al., the authors examined differences in fibroid characteristics and their risk factors between black and white women (N=360) undergoing hysterectomy from the Prospective Research on Ovarian Function study. All fibroid characteristic information was abstracted from pathology and operative notes (39). The authors found that black women on average had larger and more numerous fibroids than white women (39). Likely culminating from the higher incidence of fibroids and from the larger and more numerous fibroids among black women, Wilcox et al. (1994) observed that black women had a two-fold higher risk of having a hysterectomy because of fibroids. The authors used data obtained from the *National Hospital Discharge Survey* (8).

Parity

Parity, or the number of times a women gives birth, has been inversely associated with fibroids (34), where nulliparity has been associated with increased risk (40). Interestingly,

pregnancies that stop before viability have not been implicated with decreasing fibroid risk (40). In 1988, Parazzini et al. recruited 275 fibroid cases and 722 controls from women in the greater Milan area and observed that parous women were 40% less likely to have fibroids compared to nulliparous women (41). In another study, Ross et al. (1986) observed that fibroid risk decreased with each term pregnancy using 1,070 women (535 cases) from the *Oxford Family Planning Association* study (37). The authors observed that women with five term pregnancies had an approximately four-fold reduction in fibroid risk compared to women with zero term pregnancies (37). The relationship between parity and fibroid characteristics such as fibroid number and size remain less clear.

Type 2 Diabetes

There have been a few studies evaluating the association between diabetes and fibroid risk where having diabetes is associated with a decreased fibroid risk (17, 18, 42). In 2007, Wise et al. examined the association between diabetes with and without medication use and uterine fibroids using 23,571 AA women from the *Black Women's Health Study*, a prospective cohort (17). Wise et al. observed that diabetes was inversely associated with fibroids for women using medication (incidence rate ratio [IRR]: 0.77; 95% confidence interval [CI]: 0.60, 0.98) but was not associated with fibroids for women not using medication (IRR: 0.91; 95% CI: 0.64, 1.28) (17). In another study in 2017, Velez Edwards et al. performed association analyses between diabetes with and without medication use and fibroids using black and white women from the Synthetic Derivative, a database of de-identified EHRs of patients from the VUMC (18). Velez Edwards et al. observed that diabetes was inversely associated among white women (adjusted

odds ratio [aOR]: 0.50; 95% CI: 0.35, 0.72) and not among black women (aOR: 0.76; 95% CI: 0.50, 1.17) (18). When stratifying women by type of diabetes medication usage, Velez Edwards observed that diabetes was inversely associated with fibroids for white women on insulin only (aOR: 0.42; 95% CI: 0.26, 0.68) compared to other drugs Metformin and Thiazolidinedione (18). There was a trend towards an inverse association between diabetes and fibroids for black women on insulin, but this association was not significant (aOR: 0.60; 95% CI: 0.36, 1.01) (18). While there is little known about the association between diabetes and fibroid size or number, there has been one study examining associations between hormones (insulin-like growth factor-I [IGF-I] and insulin) and proteins (protein binding 3 [BP3]) related to diabetes and fibroid size (42). In a cross-sectional study by Baird et al. (2009), the authors used 988 women (630 cases) who had ultrasounds from the *National Institute of Environmental Health Sciences Uterine Fibroid Study* and found that having a small (<2 cm) and medium (≥2 to <4 cm) sized fibroid was inversely associated with increased levels of insulin-like growth factor-I in white women and having a large fibroid (≥4 cm) was inversely associated with insulin levels, especially in black women (42). Baird et al. found no association between BP3 and fibroid size (42).

Genetic Contributions to Uterine Fibroids

Heritability Studies

Fibroids are heritable with two twin studies estimating fibroid heritability to be between 26 to 69% among women of European ancestry (19, 20). In the first twin study, Snieder et al. (1998) used data on 98 monozygotic (MZ) and 125 dizygotic (DZ) twin pairs from the United Kingdom (UK) where at least one twin from each pair underwent a hysterectomy (19). Snieder et

al. found the heritability of fibroids as an indication of hysterectomy to be 69% (19). In the second twin study, Luoto et al. (2000) examined heritability of fibroid number using a smaller sample size of 17 MZ and 16 DZ twin pairs from ages 40 and 47 from the *Finnish Twin Cohort* who underwent an ultrasound (20). Luoto et al. observed that the mean fibroid number was 1.7 for fibroid cases and 1.1 for all women (including fibroid controls) and estimated the heritability of fibroid number to be 26% (20).

There has been one previous study estimating the heritability of fibroids using common genetic variants derived from GWAS data (43). The study by Ge et al. (2017) had a sample size of 57,151 women from the UK Biobank from ages 40 to 69 with a fibroid prevalence of 2.77% that was self-reported (43). Ge et al. estimated that fibroids are 8.71% heritable (43). In a study by Baird et al. (2003), the authors observed that 51% of premenopausal women without a previous fibroid diagnosis from the ages of 25 to 49 were diagnosed with a fibroid after having an ultrasound (2). Furthermore, Baird et al. found that the cumulative incidence of fibroids approached 70% for white women in the US (2). The large discrepancy between prevalence estimates between the two studies could signal potential misclassification for fibroid controls resulting in a lower heritability estimate in the Ge et al. study.

Somatic Mutations

Even though fibroids have an unknown etiology (40), there are numerous recurrent mutations that occur in fibroid tissue (11, 26, 44, 45). While up to 60% of fibroids are karyotypically normal (45), it is estimated that approximately 40-50% of fibroids have

cytogenetic abnormalities (11, 26, 44, 45), and these cytogenetic abnormalities are tumor-specific and nonrandom (11). The most common chromosomal translocation includes t(12;14) (q14-q15; q23-24) being present in approximately 20% fibroid tumors with chromosomal rearrangements (44, 45). Other common chromosomal alterations include del(7)(q22q32) (17%) (11, 45) and 6p21 (11, 44-47) (~5%) (11). Less common chromosomal aberrations include chromosomes X, 1, 3, 5, 10, 12, and 13 (45). Fibroids that are karyotypically abnormal tend to be larger than karyotypically normal fibroids (44).

The break points within 12q15 from t(12;14) (q14-q15; q23-24) mostly occur upstream of gene promoter high mobility group AT-hook 2 (*HMGA2*), leading to increased *HMGA2* expression (44). The other rearrangement including 6p21 encompasses high mobility group AT-hook 1 (*HMGA1*) (11, 45, 47). Interestingly, the high mobility group A (HMGA) subfamily moderates the activity of receptors for progesterone and estrogen (47). The genes *HMGA1* and *HMGA2* are aberrantly overexpressed in fibroid tissue likely resulting from their corresponding chromosomal rearrangements in either 12q15 or 6p21 (47). In contrast to *HMGA1* expression, *HMGA2* is not expressed in normal myometrium tissue (45). Likely related to *HMGA1* sometimes being expressed in normal myometrial tissue, *HMGA1* expression has been found in fibroid tumors that are karyotypically normal (45). Cytogenic alterations involving *HMGA1* and *HMGA2* are associated with other benign mesenchymal tumors including endometrial polyps, pulmonary chondroid hamartomas, and lipomas (47). Expression of *HMGA1* has also been associated with various carcinomas including mammary, prostate, and colorectal carcinomas (47). Fibroids with chromosomal rearrangements affecting *HMGA2* are associated with increased fibroid size (44).

In addition to chromosomal aberrations, there are recurrent mutations that occur in specific genes (44). The most commonly mutated gene in fibroids is mediator complex subunit 12 (*MED12*) (44) in Xq13.1, which is part of the Mediator complex that helps regulate gene expression levels (48). The mutation rate for *MED12* is between 50-85% for all fibroids with the majority of mutations occurring in exons 1 or 2 (44, 49). It is thought that mutations in *MED12* might lead to fibroid formation (49). Interestingly, fibroids with *MED12* mutations are associated with smaller fibroids (44).

Animal Models

There are animal models for uterine fibroids (14). A lesser used animal model includes the guinea pig (14). The prevalence of fibroids among aged guinea pigs is approximately 8%, but ovariectomy combined with estrogen supplementation can increase the frequency of fibroid development (14). The best-characterized animal model that is most similar to human fibroid development, however, is the Eker rat (14). The Eker rat is unique in that one copy of tuberous sclerosis complex 2 (*Tsc2*) gene is deactivated by a retroviral insertion and spontaneously forms fibroids at a frequency of approximately 65% (47). Loss of heterozygosity (LOH) of the *Tsc2* gene occurs at a frequency of ~50% in fibroid tissue, while loss of *Tsc2* gene expression occurs for nearly all fibroids (47). Similarly to human fibroids and fibroid risk, Eker rat fibroid tumors are benign and have a hormonal component (47). The Eker rat fibroids do not form if an ovariectomy is performed by 4 months of age (47). Pregnancy is protective against Eker rat fibroid formation (47). In addition, the Eker rat *hmga2* gene, which is the homolog to human *HMGA2*, is aberrantly expressed fibroid tissue (47). While human fibroid cell lines lose their

estrogen responsiveness causing it difficult to study hormone drug treatments in vitro, Eker rat cell lines maintain responsiveness to estrogen (47). This retention of sensitivity to hormones such as estrogen (47) makes the Eker rat a valuable animal model for research on fibroids.

Genome-wide Association Studies

There have been two prior GWAS on fibroid risk with the most significant index SNPs shown in Table 1.2 (50, 51). The first GWAS by Cha et al. (2011) found three loci, 10q24.33 (rs7913069; odds ratio [OR] = 1.47; $p = 8.65x10^{-14}$), 22q13.1 (rs12484776; OR = 1.23; $p = 2.79x10^{-12}$), and 11p15.5 (rs2280543; OR = 1.39; $p = 3.82x10^{-12}$) associating with fibroid risk using women from the BioBank Japan Project (50). These genetic findings have been replicated (52). For example, Edwards et al. (2013) replicated associations in rs2280543 in blocked early in transport 1 homolog (*BET1L*) and in rs12484776 in trinucleotide repeat containing 6b (*TNRC6B*) in fibroid risk from a meta-analysis of candidate SNPs using two cohorts composed of women of European ancestry (52). These genetic findings have also been implicated with fibroid size. In an additional study, Edwards et al. (2013) found that the SNP, rs12484776, in 22q13.1 within the gene *TNRC6B* associated with increasing fibroid volume from a cohort of women of European ancestry from the *Right from the Start* cohort (53). In the second GWAS, Hellwege et al. (2017) performed meta-analyses from GWAS on fibroid risk using datasets of AA women from EHR populations. The authors observed a genome-wide significant association between rs739187 in cytohesin 4 (*CYTH4*) in 22q13.1 and fibroid risk (51). There have been no prior GWAS on fibroid characteristics such as fibroid size or number.

Table 1.1. Selected Fibroid Risk Factors for each exposure.

Study	Exposure	N	aOR (95% CI)
Baird et al. (2003) (2)	Race[a]		
	White	-	1.00 (referent)
	Black	-	2.7 (2.3, 3.2)
Velez Edwards et al. (2017) (18)	Type 2 Diabetes[b]	2,353	
	No		1.00 (referent)
	Yes		0.61 (0.47, 0.80)
Study	**Exposure**	**N***	**aRR (95% CI)**
Marshall et al. (1998) (35)	Age at Menarche[c]		
	≤10	313	1.24 (1.08, 1.41)
	11	584	1.11 (1.00, 1.23)
	12	930	1.00 (referent)
	13	748	0.91 (0.82, 1.00)
	14-15	367	0.87 (0.77, 0.98)
	≥16	64	0.68 (0.53, 0.88)
	Parity Status[c]		
	Nulliparous	949	1.00 (referent)
	Parous	2,026	0.67 (0.61, 0.74)
Marshall et a. (1988) (38)	BMI[d]		
	<20.0	309	0.90 (0.79, 1.03)
	20.0-21.9	655	1.00 (referent)
	22.0-23.9	584	1.08 (0.97, 1.21)
	24.0-25.9	458	1.16 (1.03, 1.31)
	26.0-27.9	262	1.21 (1.05, 1.40)
	28.0-29.9	209	1.36 (1.16, 1.59)
	≥30.0	481	1.23 (1.09, 1.39)
Study	**Exposure**	**N**	**Matched Relative Risk**
Ross et al. (1986) (37)	BMI		
	<19	59	1.00
	19-20.9	241	1.85
	21-22.9	347	2.17
	23-24.9	242	2.51

	25-26.9	88	2.88
	≥27	93	2.47
	Parity		
	0	89	1.00
	1	143	0.87
	2	515	0.47
	3	225	0.43
	4	67	0.39
	5	31	0.24
Study	**Exposure**	**N**	**Rate/1000 women years**
Ross et al. (1986) (37)	Age		
	25-29	7	0.31
	30-34	48	0.96
	35-39	169	2.67
	40-44	201	4.63
	45-49	103	6.20
	≥50	10	4.24

N-number; BP-base pairs; OR-odds ratio; CI-confidence interval; aOR-adjusted odds ratio; RR-relative risk; aRR-adjusted relative risk; BMI-body mass index.
[a] Model adjusted for BMI and parity.
[b] Model adjusted for age, BMI, and race.
[c] Model adjusted for age, race, marital status, age at menarche, age at first birth, years since last birth, age at first oral contraceptive use, and history of infertility.
[d] Model adjusted for age, race, marital status, age at menarche, body mass index, age at first birth, years since last birth, age at first oral contraceptive use, and history of infertility.
*Number of cases only.

Table 1.2. Index SNPs from GWAS of Fibroid Risk.

Study	Chr Region	BP	SNP	A2/A1	OR (95% CI)	P-value
Cha et al. (2011) (50)	10q24.33	105,715,399	rs7913069	A/G	1.47 (1.23, 1.75)	$8.65x10^{-14}$
	22q13.1	40,652,873	rs12484776	G/A	1.23 (1.11, 1.37)	$2.79x10^{-12}$
	11p15.5	203,788	rs2280543	G/A	1.39 (1.17, 1.64)	$3.82x10^{-12}$
Hellwege et al. (2017) (51)	22q13.1	37,728,254	rs739187	T/C	1.23 (1.16, 1.30)	$7.83x10^{-09}$
	3p26.1	5,348,595	rs55768811	A/T	0.76 (0.66, 0.87)	$2.62x10^{-07}$
	8q21.11	75,119,342	rs6472827	T/C	0.76 (0.65, 0.86)	$3.89x10^{-07}$
	8p23.2	4,441,780	rs11987640	C/G	0.81 (0.73, 0.89)	$4.85x10^{-07}$

Chr-chromosome; BP-base pairs; SNP-single nucleotide polymorphism; A2-effect allele; A1-reference allele; OR-odds ratio; CI-confidence interval. BP positions are based on the reference genome build GRCh37.p13.

CHAPTER 2

EVALUATING RISK FACTORS FOR DIFFERENCES IN FIBROID SIZE AND NUMBER USING A LARGE ELECTRONIC HEALTH RECORD POPULATION

INTRODUCTION

As noted in Chapter 1, fibroids are a heterogeneous disease with many health implications. Some women develop a single, small fibroid while others develop large and/or multiple fibroids. Multiple fibroids have been found to be associated with preterm birth and cesarean delivery, while large fibroids have been found to be associated with preterm premature rupture of membranes (30).

Increasing BMI (kg/m^2) (10), nulliparity (15, 16), and black race (11-14) have been previously associated with fibroid risk. Additionally, type 2 diabetes has been shown to be protective for fibroids (17, 18). Although there are well documented studies examining fibroid risk factors (10-15), there is limited understanding of factors associated with fibroid size and number. One potential reason for the lack of studies on fibroid characteristics is because fibroid size and number can only be assessed by imaging or surgical procedures (such as ultrasounds or hysterectomies).

Few studies examine the risk factors for fibroid characteristics. In one study by Velez Edwards et al., the authors associated decreasing age at menarche with increasing fibroid number and size using the *Right From the Start* (2001-2010) cohort which had performed ultrasounds on 5,023 women (540 cases) (9). In a cross-sectional study by Baird et al., the authors included 988 women (630 cases) who had ultrasounds from the *National Institute of Environmental Health*

Sciences Uterine Fibroid Study and found that decreased levels of insulin-like growth factor-I was associated with having a small (<2 cm) and medium (≥2 to <4 cm) sized fibroid in white women and decreasing insulin levels was associated with having a large fibroid (≥4 cm), especially in black women (42). In another cross-sectional study by Moorman et al., the authors examined differences in fibroid characteristics and their risk factors between black and white women (N=360) undergoing hysterectomy from the *Prospective Research on Ovarian Function* study where all fibroid characteristic information was abstracted from pathology and operative notes (39). The authors found that black women on average had larger and more numerous fibroids than white women (39). The authors also found that nulligravid black and white women were more likely to have larger fibroids and only nulligravid white women were more likely to have multiple fibroids than women with one or more pregnancies (39).

Physical and clinical patterns of a woman may put her at risk of developing fibroids with specific characteristics, such as single versus multiple fibroids or a small versus large fibroid. Knowing which of these physical and clinical features are most associated with fibroid characteristics might allow for potential targeted fibroid treatments. Our objective was to identify physical and clinical features associated with fibroid size or number. We examined the association of fibroid characteristics (fibroid volume, largest fibroid dimension, and fibroid number) with candidate risk factors using a clinical cohort of 2,302 women identified from EHRs.

MATERIAL AND METHODS

Study Population

The Synthetic Derivative

We conducted our analyses using subjects from the Synthetic Derivative, a clinical dataset at VUMC (22) that consists of de-identified demographic and clinical information from patient EHRs.

We used a previously validated phenotyping algorithm with a positive predictive value of 96% to identify fibroid cases (21). The algorithm included only black and white women who were 18 years or older, had at least one International Classification of Diseases, 9^{th} Revision (ICD-9) or current procedure terminology (CPT) codes for pelvic imaging, and at least one ICD-9 or CPT code indicating a fibroid diagnosis. Fibroid status for 2,302 cases was manually validated by examining image or surgical reports from patient EHRs. We manually extracted dimensions for each reported fibroid, number of fibroids, and relevant demographic information from pelvic imaging reports including ultrasound, magnetic resonance imaging, and computed tomography (CT) scans or surgical reports from myomectomies and hysterectomies. Precedence for recording patient information was given to the first image report mentioning fibroids. If a patient's EHR lacked an image report citing fibroids, we entered patient data from the first surgical report describing their fibroids.

Abstracted fibroid characteristics included fibroid number (single vs. multiple), volume of largest fibroid (centimeter cubed [cm^3]), and largest dimension (cm) of all reported fibroids. The following formula for volume of an ellipsoid was used to calculate fibroid volume: (Length x Width x Height x 0.523). The largest fibroid dimension and volume measurements were $\log_{10}$ transformed to accommodate the assumption of normally distributed residuals for linear

regression. For individuals with only two recorded dimensions of their largest fibroid (35.6% of cases), we imputed the last measurement by averaging the initial two to calculate fibroid volume, assuming the first two measurements were predictive of the missing third measurements. For individuals with data for number of fibroids, some EHRs (18.6%) noted the presence of many fibroids but gave no specific number. Because of this limitation, we coded fibroid number as one versus multiple fibroids.

Abstracted fibroid risk factors included age at diagnosis, BMI (continuous), self-reported or clinically identified race, type 2 diabetes status, and number of living children. We used previously published programming algorithms to abstract type 2 diabetes from EHRs (54). Number of living children was chosen as a proxy for parity. This study has been evaluated and approved by the Vanderbilt University Medical Center Institutional Review Board.

Statistical Analyses

We performed univariate and multivariate regression analyses on the outcomes volume of largest fibroid, largest dimension of all fibroids, or number of fibroids (single vs. multiple). Exposures included age at diagnosis, BMI, self- or third-party-identified race, type 2 diabetes status, and number of living children (a proxy for parity) as covariates. All multivariate regression analyses included age, BMI, and race as covariates. Age and BMI were not normally distributed and had nonlinear effect sizes; therefore analyses were performed grouping age into quintiles and BMI into World Health Organization (WHO) categories (55). There were too few underweight individuals to be included in statistical analyses (underweight N=27), so we combined individuals from the underweight and normal weight categories. We assessed potential effect measure modification by race on age quintiles and BMI categories using a likelihood ratio test comparing the full model with interaction term to the same model lacking the interaction

term, and a likelihood ratio test p-value of 0.10 or less was considered significant. Because we did not observe effect measure modification within our dataset, we did not stratify the analyses by race.

In secondary analyses multiple imputation (MI) was performed for missing data (number of living children) for each outcome to determine if missingness in this exposure significantly affected our analyses. All exposures in the multivariate regression model and the outcome were included in each MI data model. Each dataset was imputed ten times. To ensure consistency in MI analyses, the random-number generation functions were seeded with the number 12,345. MI analyses were largely consistent with effect sizes estimated from non-imputed data; therefore, we present and discuss results using non-imputed missing data.

We used Stata/SE 13 (College Station, Texas) for all statistical analyses.

RESULTS

Demographic Data

There were 2,302 individuals with fibroids within our study population (**Table 2.1**). The mean age was 45.5 years ± 12, and the majority of women were either overweight (29%) or obese (44%) with a mean BMI of all women being 30.4 kg/m^2. The median volume of largest fibroid and largest fibroid dimension were 12.6 cm^3 and 2.9 cm, respectively. Additionally, about half of the women had multiple fibroids (52%).

Fibroid Number

Increasing age when compared to the referent groups (first quintile) was significantly associated with multiple fibroids (**Table 2.2**). The strongest association was observed when comparing ages 43 to 47 to the referent group (ages 18 to 36) (adjusted odds ratio [aOR]: 3.37;

95% confidence interval [CI]: 2.55, 4.46). Additionally, black race was significantly associated with having multiple fibroids (aOR: 1.83; 95% CI: 1.49, 2.24). BMI and type 2 diabetes were not associated with fibroid number. Finally, increasing number of living children was associated with having multiple fibroids (aOR: 0.88; 95% CI: 0.78, 0.99).

Fibroid Size (Volume and Largest Dimension)

Increasing age significantly was associated with larger fibroid volume when comparing age groups 43 to 47 (adjusted beta: 1.47; 95% CI: 1.05, 2.07) and 48 to 54 (adjusted beta: 1.57; 95% CI: 1.12, 2.22) to the referent group (ages 18 to 36) (**Table 2.3**). Increasing age was also significantly associated with larger fibroid dimensions when comparing age groups 43 to 47 (adjusted beta: 1.16; 95% CI: 1.04, 1.30) and 48 to 54 (adjusted beta: 1.19; 95% CI: 1.07, 1.32) to the referent group (ages 18 to 36) (**Table 2.4**). Black race was significantly associated with increasing fibroid volume (adjusted beta: 1.77; 95% CI: 1.38, 2.27) and largest dimension (adjusted beta: 1.28; 95% CI: 1.18, 1.38). BMI and type 2 diabetes were not associated with fibroid size. Number of living children was not associated with fibroid volume or largest dimension.

DISCUSSION

To our knowledge we conducted the largest study evaluating the relationship between fibroid characteristics and clinical and physical features. In this study, we observed no effect modification by race in analyses of age or BMI and fibroid characteristics. We observed the strongest associations between age and multiple fibroids when comparing ages 43 to 47 to the referent group, and the strongest associations between age and increasing fibroid volume and largest fibroid dimension was later between ages 48 to 54. We observed that black race is

strongly associated with larger fibroid sizes and increasing number of fibroids. We observed no associations between fibroid characteristics and BMI or type 2 diabetes. Finally, we observed that having more children was associated with single fibroids. These results are consistent with phenotypic heterogeneity in the fibroid phenotype; in that not all women have fibroids because of the same set of risk factors. For example, there are different risk factors for a woman getting a single fibroid than there are for developing multiple large fibroids.

Risk for having multiple fibroids peaked between ages 43 and 47, while risk having for larger fibroids peaked later between ages 48 to 54. This is overall consistent with fibroid risk that also increases until menopause (2). In a study by Baird et al. (2003), increasing clinically relevant fibroid prevalence for both white and black women was associated with increasing age between 35 and 49 (2). We observed that risk for having multiple fibroids increased and peaked when women reached premenopause/early menopause age and then that risk declined for subsequent age groups, across all age analyses (fibroid number, volume, and largest dimension). Interestingly, Cramer and Patel (1990) estimated the prevalence estimate of fibroids by performing serial sectioning of uterine tissue from 100 consecutive hysterectomies (28). The authors found that while there was little difference in fibroid risk with regards to menopausal status (28). The authors also found that postmenopausal women were found to have smaller and fewer fibroids compared to premenopausal women. These observations support a hormonal influence on having larger and more numerous fibroids. In addition, fibroids have not been documented in prepubescent girls (25), since fibroids have a hormonal component. It could be possible that fibroids form and progress only after many years of exposure to hormones including estrogen and progesterone. After events like menopause where estrogen and

progesterone levels are reduced, fibroids progression/development is stunted. This information is supported by postmenopausal women having less surgery for fibroid-related reasons (34).

Increasing BMI was not associated with either fibroid number or size. In a previous study, Dandolu et al. (2010) found that increasing BMI was associated with fibroid risk but not with fibroid largest dimension (10). The authors dichotomized fibroid size as greater than or less than two centimeters and included 533 women (10). The lack of an association between BMI and fibroid characteristics could mean that BMI influences fibroid risk only and not progression.

Type 2 diabetes was not significantly associated with fibroid number or size in our study. It should be noted, however, that type 2 diabetes tended to be associated with single and smaller fibroid volume. Diabetes has been inversely associated with fibroid risk. In 2017, Velez Edwards et al. observed a protective effect of diabetes on fibroid risk (aOR: 0.61; 95% CI: 0.47, 0.80) (18). Studies between diabetes and fibroid characteristics are limited. In the Baird et al. study (2009), the authors performed association analyses between diabetes metabolites and fibroid size in white and black women. The authors observed an inverse association between insulin-like growth factor-I and fibroid size in both white and black women (42). The authors only had approximately 92 individuals with diabetes (including gestational diabetes) and could not perform association analyses between diabetes and fibroid size (42). There were only 317 women that were diabetic in our dataset. The lack of an association between type 2 diabetes and fibroid number and size could mean that our sample size was too small. Future studies with a larger diabetic population should examine the association between type 2 diabetes and fibroid characteristics.

We found that having more children was associated with having fewer fibroids suggesting that pregnancy may reduce the number of fibroids present. We also observed

nonsignificant trends between increasing number of living children and smaller fibroid sizes. This is consistent with literature showing that women with a higher parity are at a reduced risk for fibroids (relative risk: 0.5; 95% CI: 0.4-0.6) (16). The protective effects of carrying a child to term could not only decrease the risk of having fibroids but also the number of fibroids only (not fibroid size). More studies are needed to assess the association between number of living children/parity and fibroid characteristics.

Parity was not consistently documented in the EHRs, and we used number of living children as a proxy. Additionally, prior studies examining the relationship between fibroid characteristics and fibroid risk factors compared each characteristic to controls. With that approach it is difficult to determine if the observed effect is due to a woman having a fibroid or to having specific fibroid characteristics.

Using a dataset of EHRs consisting of women with image- or surgery-confirmed fibroids, we found that black race was associated with multiple and larger fibroids. We also observed that increasing age was nonlinearly associated with multiple and larger fibroids with risk for multiple and larger fibroids being different, respectively, for different age groups. Lastly, we observed that multiple fibroids and not size were associated with increasing number of living children. Our findings are consistent with the phenotypic heterogeneity of fibroids and suggest that there may be different underlying etiology involved with women developing single versus multiple fibroids and small versus large fibroids. Additionally, these findings suggest that physical and clinical features of a woman may help to determine if and when she is at risk of developing fibroids with specific characteristics. These results can potentially guide a physician towards the best route of fibroid treatment.

Table 2.1. Demographic characteristics of study population.

Characteristics	N	Synthetic Derivative Cohort
Age (mean±SD)	2,302	45.5±12
18 to 36 (%)	510	22
37 to 42 (%)	438	19
43 to 47 (%)	442	19
48 to 54 (%)	493	21
55 to 87 (%)	419	18
BMI (kg/m^2) (mean±SD)	2,302	30.4±8
Underweight (<18.5) (%)	27	1
Normal Weight (18.5-24.9) (%)	606	26
Overweight (25-29.9) (%)	663	29
Obese (≥30) (%)	1,006	44
Race (%)	2,302	
White	1,616	70
Black	686	30
Fibroid Volume (cm^3) median (IQR)	1,307	12.6 (2.7, 47.6)
Largest Fibroid Dimension (cm) median (IQR)	1,777	2.9 (1.7, 4.8)
Fibroid Number (%)	2,149	
1	1,042	48
>1	1,107	52
Type 2 Diabetes Status (%)	2,302	
No	1,985	86
Yes	317	14
Number of Children (%)	692	
0	160	23
1	107	15
2	242	35
3	113	16
4	39	6
5	20	3
>6	11	2

BMI-body mass index; cm^3-cubic centimeters; SD-standard deviation; IQR-interquartile range.

Table 2.2. Association between fibroid number and exposures.

Exposures	N	Crude OR [95% CI]	Crude *P-value*	Adjusted [a] OR [95% CI]	Adjusted [a] *P-Value*
Age	2,149				
18 to 36	474	1.00 (referent)	-	1.00 (referent)	-
37 to 42	405	1.79 [1.36, 2.34]	<0.001	1.85 [1.41, 2.44]	<0.001
43 to 47	417	3.07 [2.34, 4.04]	<0.001	3.37 [2.55, 4.46]	<0.001
48 to 54	465	2.09 [1.61, 2.72]	<0.001	2.48 [1.89, 3.25]	<0.001
55 to 87	388	1.84 [1.40, 2.41]	<0.001	2.17 [1.64, 2.89]	<0.001
BMI (kg/m^2)	2,149				
Normal Weight*	595	1.00 (referent)	-	1.00 (referent)	-
Overweight	623	1.21 [0.96, 1.51]	0.102	1.08 [0.85, 1.36]	0.527
Obese	931	1.07 [0.87, 1.32]	0.513	0.93 [0.75, 1.16]	0.530
Race	2,149				
White	1,503	1.00 (referent)	-	1.00 (referent)	-
Black	646	1.52 [1.27, 1.84]	<0.001	1.83 [1.49, 2.24]	<0.001
Type 2 Diabetes Status	2,149				
No	1,858	1.00 (referent)	-	1.00 (referent)	-
Yes	291	0.94 [0.73, 1.20]	0.623	0.77 [0.59, 1.01]	0.055
Number of Living Children	631	0.93 [0.83, 1.04]	0.213	0.88 [0.78, 0.99]	0.036

CI-confidence interval; BMI-body mass index; kg/m^2-kilogram per meter squared.

[a] Adjusted for age quintiles, WHO BMI categories, and race, respectively.

*Due to the limited number of underweight subjects (N=27) normal weight includes underweight individuals.

Table 2.3. Association between fibroid volume and exposures.

Exposures	N	Crude Beta** [95% CI]	Crude *P-value*	Adjusted [a] Beta** [95% CI]	Adjusted [a] *P-Value*
Age	1,307				
18 to 36	313	1.00 (referent)	-	1.00 (referent)	-
37 to 42	258	1.19 [0.85, 1.68]	0.311	1.22 [0.86, 1.71]	0.262
43 to 47	261	1.40 [0.99, 1.97]	0.054	1.47 [1.05, 2.07]	0.027
48 to 54	276	1.33 [0.95, 1.86]	0.102	1.57 [1.12, 2.22]	0.010
55 to 87	199	0.70 [0.48, 1.01]	0.055	0.81 [0.56, 1.18]	0.281
BMI (kg/m^2)	1,307				
Normal Weight*	344	1.00 (referent)	-	1.00 (referent)	-
Overweight	377	1.42 [1.05, 1.93]	0.025	1.31 [0.97, 1.78]	0.083
Obese	586	1.32 [1.00, 1.75]	0.049	1.13 [0.85, 1.50]	0.408
Race	1,307				
White	844	1.00 (referent)	-	1.00 (referent)	-
Black	463	1.75 [1.38, 2.21]	<0.001	1.77 [1.38, 2.27]	<0.001
Type 2 Diabetes Status	1,307				
No	1,141	1.00 (referent)	-	1.00 (referent)	-
Yes	166	0.78 [0.55, 1.09]	0.149	0.72 [0.51, 1.02]	0.062
Number of Living Children	345	0.97 [0.82, 1.14]	0.696	0.93 [0.79, 1.11]	0.422

CI-confidence interval; BMI-body mass index; kg/m^2-kilogram per meter squared.

[a] Adjusted for age quintiles, WHO BMI categories, and race, respectively.

*Due to the limited number of underweight subjects (N=27) normal weight includes underweight individuals.

**Betas are back-transformed to cm^3.

Table 2.4. Association between largest fibroid dimension and exposures.

Exposures	N	Crude Beta** [95% CI]	Crude *P-value*	Adjusted [a] Beta** [95% CI]	Adjusted [a] *P-Value*
Age	1,777				
18 to 36	408	1.00 (referent)	-	1.00 (referent)	-
37 to 42	345	1.05 [0.94, 1.18]	0.347	1.06 [0.95, 1.19]	0.264
43 to 47	345	1.13 [1.02, 1.27]	0.026	1.16 [1.04, 1.30]	0.007
48 to 54	381	1.11 [1.00, 1.24]	0.051	1.19 [1.07, 1.32]	0.002
55 to 87	298	0.88 [0.78, 0.98]	0.024	0.94 [0.84, 1.05]	0.284
BMI (kg/m^2)	1,777				
Normal Weight*	491	1.00 (referent)	-	1.00 (referent)	-
Overweight	510	1.10 [1.00, 1.21]	0.043	1.06 [0.97, 1.17]	0.200
Obese	776	1.11 [1.02, 1.22]	0.015	1.05 [0.96, 1.14]	0.319
Race	1,777				
White	1,209	1.00 (referent)	-	1.00 (referent)	-
Black	568	1.28 [1.18, 1.38]	<0.001	1.28 [1.18, 1.38]	<0.001
Type 2 Diabetes Status	1,777				
No	1,545	1.00 (referent)	-	1.00 (referent)	-
Yes	232	0.98 [0.88, 1.08]	0.644	0.94 [0.84, 1.05]	0.278
Number of Living Children	491	0.98 [0.94, 1.03]	0.517	0.96 [0.92, 1.01]	0.155

CI-confidence interval; BMI-body mass index; kg/m^2-kilogram per meter squared.

[a] Adjusted for age quintiles, WHO BMI categories, and race, respectively.

*Due to the limited number of underweight subjects (N=27) normal weight includes underweight individuals.

**Betas are back-transformed to cm.

CHAPTER 3

ESTIMATING MISSING HERITABILITY OF UTERINE FIBROIDS USING COMMON SNPS FROM A CLINICAL COHORT COMPOSED OF EUROPEAN AMERICAN WOMEN

INTRODUCTION

As described in Chapter 1, fibroids are heritable with twin studies estimating fibroid heritability to be between 26 to 69% among EA women (19, 20). One study previously estimated the heritability of fibroids using common genetic variants derived from genome-wide SNP data (43). The study by Ge et al. (2017) had a sample size of 57,151 UK women with a fibroid prevalence of 2.77% that was self-reported (43). Ge et al. estimated that fibroids are 8.71% heritable (43).

Two prior GWAS have been performed on fibroids (50, 56) in order to discover the genetic variants associated with fibroids. The first GWAS by Cha et al. (2011) found three loci, 10q24.33 (rs7913069; OR = 1.47; $p = 8.65x10^{-14}$), 22q13.1 (rs12484776; OR = 1.23; $p = 2.79x10^{-12}$), and 11p15.5 (rs2280543; OR = 1.39; $p = 3.82x10^{-12}$) associated with fibroid risk in a Japanese cohort (50). These genetic findings have been replicated (52). For example, Edwards et al. (2013a) associated rs2280543 in *BET1L* and rs12484776 in *TNRC6B* with fibroid risk using two cohorts of EA women (52). In an additional study, Edwards et al. (2013b) found that the SNP, rs12484776, in 22q13.1 within the gene *TNRC6B* associated with increasing fibroid volume from a cohort of women of European ancestry from the *Right from the Start* cohort (53). In another GWAS, Hellwege et al. (2017) performed meta-analyses from GWAS on fibroid risk

using datasets of AA women from an EHR population. The authors observed a genome-wide significant association between rs739187 in *CYTH4* in 22q13.1 and fibroid risk (51).

GWAS are a powerful tool to find common genetic variants that associate with disease. As expected based on the available sample sizes, current GWAS of fibroids have identified only a few variants that explain a small proportion of fibroid risk. Better characterization of fibroid heritability is needed to understand the genetic architecture influencing risk for fibroids. Additionally, imaging is critical in genetic studies on fibroids because as much as 51% of women whose status is not image-confirmed and who may be asymptomatic can be misclassified as controls (2). In this study, we propose estimating and characterizing the heritability of image-confirmed fibroids in EA women using GWAS data.

MATERIALS AND METHODS

Study Population

BioVU

The BioVU DNA Repository (2007-present) is a de-identified database of EHRs that is linked to DNA. BioVU consists of stored de-identified demographic and clinical information for each patient who visits the VUMC and chooses to participate (22). A detailed description of BioVU has been previously described (22, 57). The Office of Human Research Protections and the Institutional Review Boards deemed the BioVU DNA repository as non-human subjects research (57).

Fibroid cases and controls were identified using a validated phenotyping algorithm with a positive predictive value of 96% and a negative predictive value of 98% (21). The same fibroid case phenotyping algorithm that was used in Chapter 2 on the Synthetic Derivative was applied in this study on EA women. Controls were at least 18 years old, had at least two ICD-9 or CPT

codes for pelvic imaging, and had no history of myomectomy, hysterectomy, or UAE (21) and were group matched by ancestry to cases. We also conducted a case-only analysis to evaluate the heritability of fibroid size. In order to determine fibroid size, we manually abstracted fibroid measurements from imaging and surgical reports. Patient demographic information was recorded at the time of diagnosis.

Outcome measurements include fibroid presence (case vs control), largest fibroid dimension of all measurements, and largest fibroid volume. Fibroid volume was calculated using the equation of an ellipsoid: (Length x Width x Height x 0.523). In order for the outcome to have a normal distribution, fibroid largest dimension and volume were $\log_{10}$ transformed. Approximately a third of individuals with volume measurements had their third dimension measurement imputed by taking the mean of the first two measurements.

Genotyping

Genotyping of BioVU EA individuals was performed on the Affymetrix Axiom Biobank array (Affymetrix, Inc., Santa Clara, CA) using PicoGreen (Invitrogen, Inc., Grand Island, NY) to purify and to quantify the DNA.

GWAS Quality Control (QC)

Genetic data had standard QC prior to imputation using PLINK 1.07 software (58). Pre-imputed QC included: removing subjects with low genotyping efficiency (<95%), removing related individuals (both individuals from a pair with a probability of identity by descent [IBD] > 0.95 and one from a pair with a probability of IBD from 0.2-0.95), removing individuals with inconsistency between genetic and reported sex. Additionally, we removed SNPs out of HWE ($p \leq 10^{-6}$), SNPs with a low genotyping efficiency (<95%), SNPs with a minor allele frequency

(MAF) < 0.01, and SNPs without a chromosome location. Finally, SNPs were aligned to the + strand of the 1000 Genomes (build 37, 2013).

We imputed non-genotyped SNPs using IMPUTE2.3.0 software (59) and subsequently converted the imputed genetics to PLINK 1.07 format (58) after filtering SNPs with an info score ≤0.95 and removing insertion/deletion polymorphisms. We then applied GCTA-specific QC that was performed by Hong Lee et al. (2011) (60) which included: removing SNPs with a MAF < 0.05, removing related individuals (one from a pair with a probability of IBD >0.05, and removing SNPs out of HWE (≥0.05). Post-QC SNP numbers included 4,500,362 for fibroid risk, 4,522,829 for fibroid volume, and 4,518,340 for largest fibroid dimension (**Appendix A**).

Statistical Analyses

Covariate and demographic data were summarized by Stata/SE (College Station, Texas). Principal components (PCs) were created using EIGENSTRAT4.2 software (61). To reduce admixture within BioVU, we removed all samples whose first PC (PC1) and second PC (PC2) were more than four standard deviations from the mean of PC1 and PC2 of the European (EUR) individuals (**Appendix B, Appendix C**).

We estimated the genetic variance explained by all SNPs using GCTA software (23). GCTA creates a genetic relationship matrix (GRM) between pairs of individuals within a cohort using all SNPs. GCTA uses a mixed linear model (MLM) that estimates the genetic variance explained by all SNPs for a trait using the restricted maximum likelihood approach (REML) (23). The final model for fibroid risk included the GRM which was modeled as a random effects term and the fixed effects terms age, BMI, and top five PCs. We transformed the observed estimate of heritability to the liability scale using a fibroids prevalence estimate of 70%. A prevalence estimate of 70% was used in models on fibroid risk because the mean age of our

cohort was 50.8 and previous cumulative incidence estimates were ~70% for women near 51 years of age (2). Heritability estimates for a range of prevalence estimates are shown in **Appendix D**. Secondary analyses included estimating heritability of fibroid size, namely volume and largest dimension. Final models for fibroid volume and max dimension adjusted for age, BMI, and five PCs without a prevalence estimate. Additional heritability estimates were calculated adjusting for age, BMI, and varying numbers of PCs to show that potential residual population stratification did not lead to spurious heritability estimates.

To characterize the genetic architecture of fibroids, we partitioned the genetic variance into separate GRMs by chromosomes, by genic and intergenic regions. We annotated SNP location into genic/intergenic regions using ANNOVAR software (62). We used a distance of one kilobase (kb) from a transcript as the boundary threshold for genic and intergenic SNPs. For the analysis partitioned by functional category (intergenic vs. genic), we used a two-tailed binomial test to assess the difference in the proportion of heritability explained by each partition, given the number of SNPs. Additional subanalyses involved excluding genomic regions previously implicated in fibroid risk based on GWAS. Exclusion criteria included selecting SNPs from final meta-analyses of GWAS (50, 51) with a p-value $\leq 1.00 \times 10^{-6}$ and subsequently removing all loci ± 0.5 mega bases (Mb) of these selected SNPs. Lastly, there is a common inversion that spans approximately 12 centimorgan (cM) on chromosome 8 encompassing 8p23.1-8p22 with a frequency of approximately 21% in European populations (63). We determined the effect that this inversion has on fibroid risk by removing it from chromosome-specific heritability estimates and comparing the difference in point estimates.

RESULTS

Demographic Data

There were 2,109 EA women (1,067 cases and 1,042 controls) with image-confirmed fibroid status within BioVU (**Table 3.1**). The mean age was 50.8±16 years, and the majority of women were overweight or obese (mean BMI: 28.6±7 km/m^2). There were 373 cases with fibroid volume measurements (median of largest volume: 9.6 cm^3) and 551 cases with fibroid dimension measurements (median of largest dimension: 2.6 cm).

Heritability Estimates

The heritability estimate for fibroid risk, or the proportion of phenotypic variance that was explained by genetic variance, was h^2 = 0.33±0.18, p-value 0.040 (**Table 3.2**), and this estimate was consistent after decreasing our MAF threshold to 1%. The heritability estimates for fibroid largest dimension was h^2 = 0.35±0.47, p-value 0.238, and for volume was h^2 = 0.14±0.66, p-value 0.417. These heritability estimates for each outcome were similar after adjusting for a range of PCs (risk h^2: 0.29-0.38; largest dimension h^2: 0.34-0.44; volume h^2: 0.12-0.35) (**Appendix E, Appendix F, Appendix G**).

Partitioning the genome into chromosomes and into genic and intergenic regions

The correlation (r^2) between chromosome length in base pairs and heritability per chromosome was 0.89% (p = 0.669) (**Figure 3.1**). Chromosome 8 explained the greatest proportion of variance for fibroid risk (8.76%), followed by chromosome 10 (5.57%), 11 (5.11%), and 7 (2.65%). The summation of the heritability estimate of each chromosome was 34%, which is not significantly different from the univariate estimate suggesting that residual population stratification and cryptic relatedness are well controlled (64). In addition, there is an approximately 12 cM common inversion present at a frequency of 21% in EA populations on

chromosome 8 spanning 8p23.1-8p22 (63). It is possible that this inversion could affect the disproportionate contribution of heritability for chromosome 8. Because of this, we estimated fibroid heritability on chromosome 8 excluding 8p23.1-8p22, 8p, and 8q, respectively (**Figure 3.2**). We found that excluding 8p23.1-8p22 did not change the contribution of heritability from chromosome 8. We also found that most of the heritability from chromosome 8 was located in 8q. Even though fibroids are a sex-linked disease and have a hormonal component especially with estrogen (14), we only observed slight heritability from chromosome X of about 1%.

We also partitioned the genetic data into genic and intergenic regions resulting in 2,801,797 intergenic SNPs (62.3%) and 1,694,891 genic SNPs (37.7%) (**Table 3.3**). There was no enrichment for intergenic or genic SNPs for the heritability of fibroid risk (binomial test p-value = 0.491) (**Figure 3.3**).

Heritability explained by previous GWAS on fibroid risk

The genomic regions that met the exclusion criteria were 10q24.33-10q25.1, 11p15.5, and 22q13.1-22q13.2 from the Cha et al. (2011) GWAS (50) and 3p26.1, 8p23.2, 8q21.11, and 22q12.3-22q13.1 from the Hellwege et al. (2017) GWAS (51). Excluding genetic loci previously associated with fibroid risk from all previous GWAS (50, 51) did not attenuate the estimated heritability of fibroid risk (h^2 from all studies = 0.34) (**Table 3.4**).

DISCUSSION

This was the first study to estimate and characterize the heritability of fibroids using a dataset of EA women with image-confirmed fibroid status. Estimated heritability of fibroid risk was 0.33±0.18. Chromosome 8 contributed the greatest proportion of heritability and we observed little correlation between chromosome length and chromosomal heritability, suggesting

that fibroid heritability is not evenly distributed throughout the genome. There was no enrichment of fibroid heritability for intergenic or genic regions. Finally, censoring genetic loci previously implicated with fibroid risk did little to attenuate the estimated heritability.

The heritability estimate of 33% for fibroid risk from our study is within the range of estimates from these twin studies (26-69%) (19, 20). In the first twin study, Snieder et al. (1998) found the heritability of fibroids to be 69% in a population of women who had hysterectomies (19). The authors had data on 98 MZ and 125 DZ twin pairs from the United Kingdom where at least one twin from each pair underwent a hysterectomy (19). The higher heritability from this study could be because Snieder et al. examined fibroid heritability as an indication for a hysterectomy selecting for patients who may have more severe fibroids. This higher heritability estimate could be for fibroids that cause symptoms leading to a hysterectomy. In the second twin study, Luoto et al. (2000) estimated the heritability of fibroid number to be 26% (20). This study's heritability estimates while having a much smaller sample size of 17 MZ and 16 DZ twin pairs from the *Finnish Twin Cohort* (20) was closer to our fibroid risk heritability estimate.

Our heritability estimate (33%) was much higher than previous heritability estimates from genome-wide SNP data by Ge et al. (2017) (43). While Ge et al. had a much larger sample size of 57,151 women from the UK Biobank from ages 40 to 69 (43), Ge et al. observed that fibroid prevalence was very low at 2.77% (43). For example, a study by Baird et al. (2001) observed that the cumulative incidence of fibroids to be approximately 70% for white women aged 49 years who were systematically screened by ultrasounds. The large discrepancy between prevalence estimates between the two studies could signal potential misclassification for fibroid controls resulting in a lower heritability estimate in the Ge et al. study.

This was the first estimate of fibroid heritability using a dataset of EA women with image-confirmed fibroid status. We observed that fibroid heritability in our study was higher than a previous genetic estimate using GWAS data. This result could in part be due to limiting our samples to image-confirmed cases and controls. In addition, we observed that our fibroid heritability estimate of 33% was not attenuated when censoring genetic loci previously associated with fibroid risk. This suggests that many of the genetic loci that are associated with fibroid risk in EA women are yet to be discovered. Future studies on fibroid risk are needed to understand the genetic underpinnings of fibroid heritability in EA women.

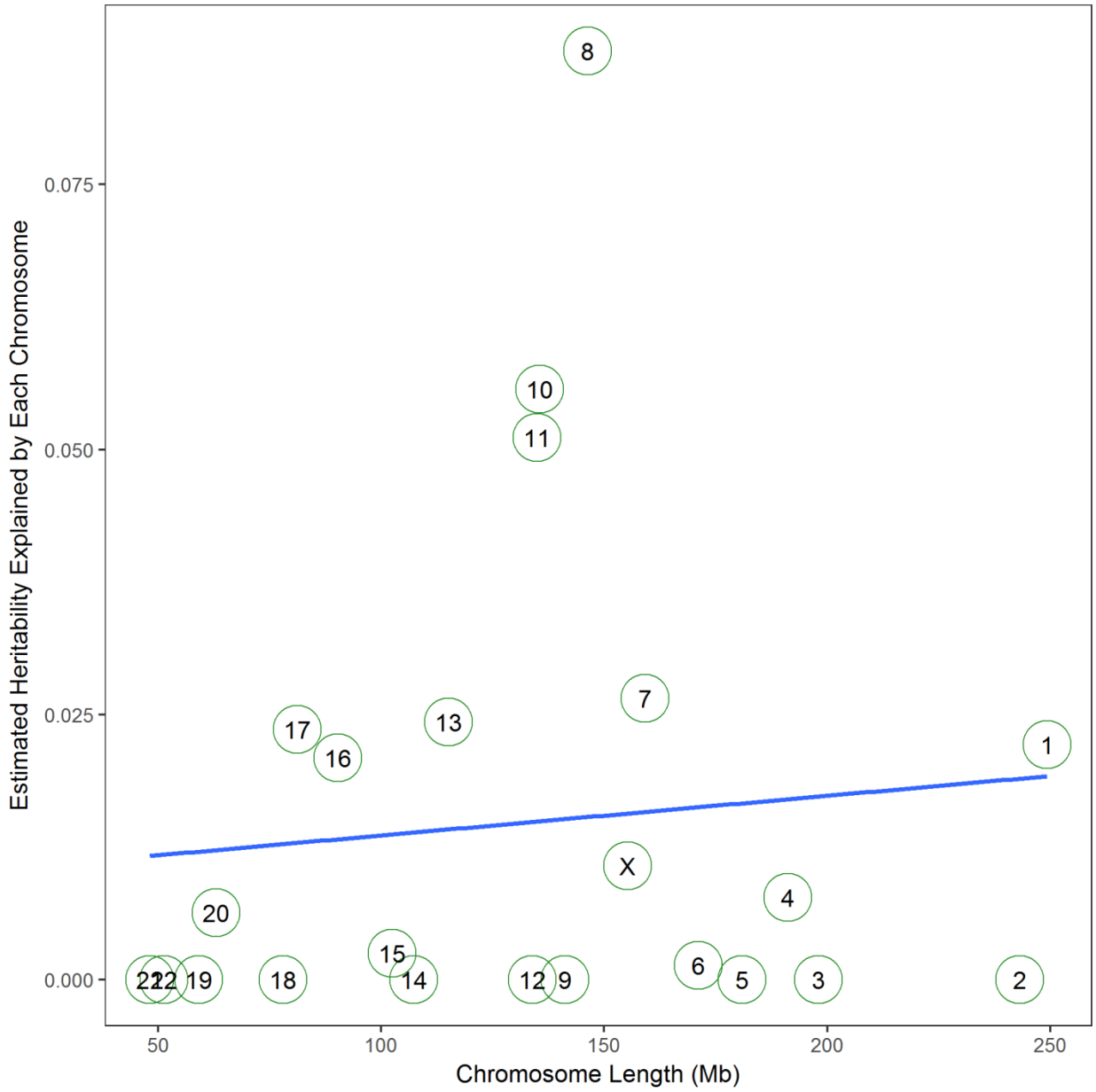


Figure 3.1. Estimated heritability of each chromosome by chromosome length in Mb. R^2 is 0.89%, and the p-value of the model is 0.669.

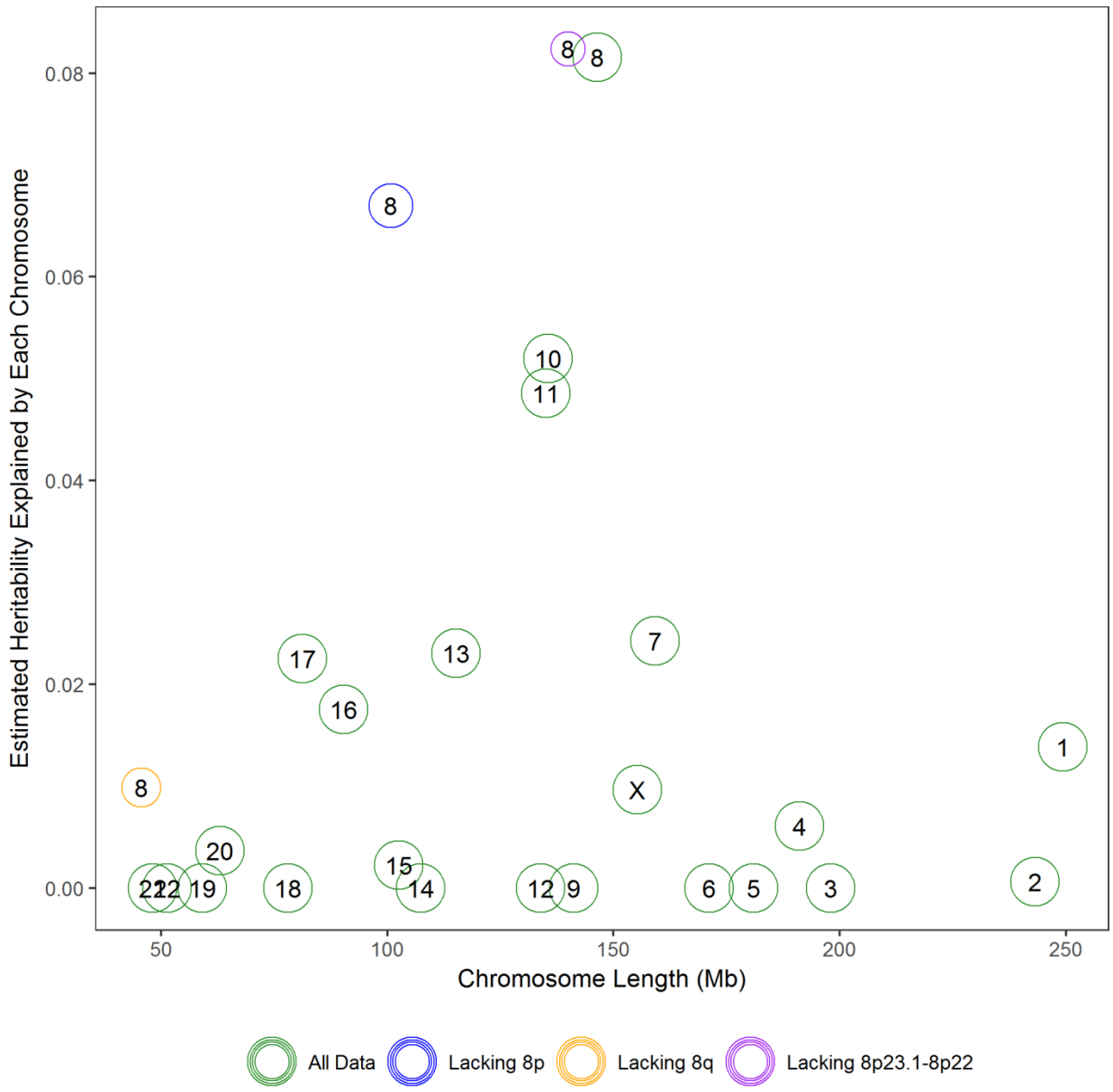


Figure 3.2. Estimated heritability of each chromosome by chromosome length in Mb with sections of chromosome 8 removed.
To help all models converge, the model used to produce this figure adjusted for age, BMI, and 7 PCs with a prevalence estimate at 70%.

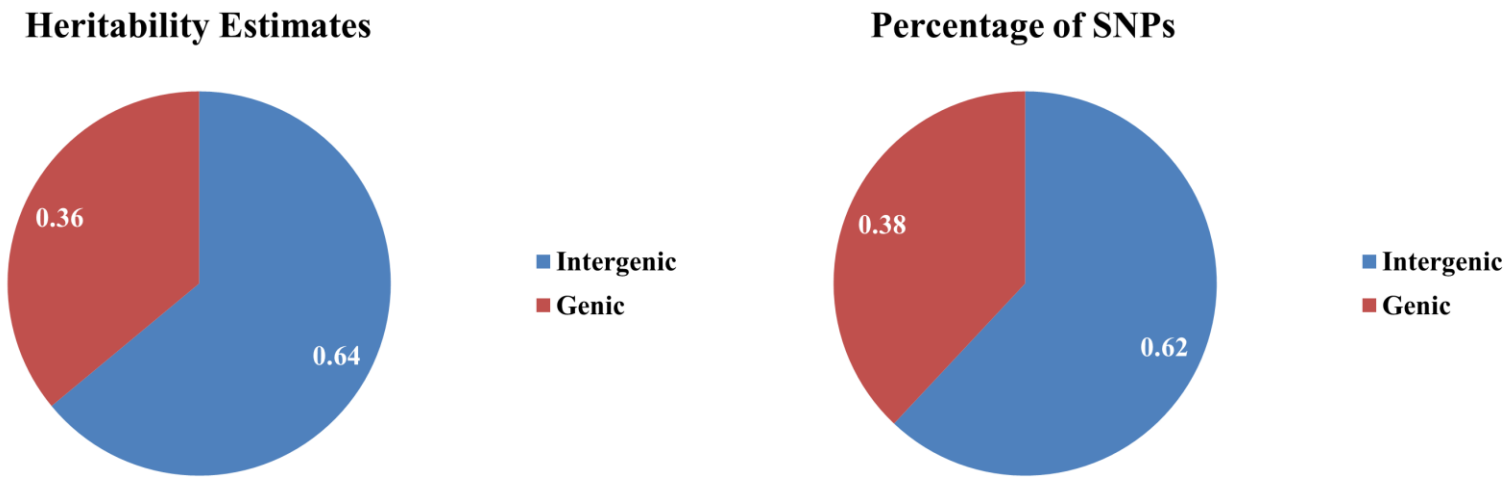


Figure 3.3. Proportion of heritability estimates and percentage of SNPs pie charts for SNPs that were partitioned into intergenic and genic SNPs.

Table 3.1. Demographics table of BioVU European American women.

Demographic Characteristics	N	BioVU
Age (mean±SD)	2,109	50.8±16
BMI (kg/m^2) (mean±SD)	2,109	28.6±7
Fibroid Risk	2,109	
Control (%)	1,042	49
Case (%)	1,067	51
Fibroid Volume (cm^3) median (IQR)	373	9.6 (2.3-39.0)
Largest Fibroid Dimension (cm) median (IQR)	551	2.6 (1.6-4.4)

BMI-body mass index; kg/m^2-kilograms per meters squared; cm^3-cubic centimeters; cm- centimeters; SD-standard deviation; IQR-interquartile range.

Table 3.2. Heritability estimates for fibroid risk using European Americans within BioVU.

Outcome	Heritability Estimate ± (Std. Err.)	P-value
Fibroid Risk [a]	0.33 ± (0.18)	0.040*
Largest Fibroid Dimension	0.35 ± (0.47)	0.238
Largest Fibroid Volume	0.14 ± (0.66)	0.417

Model: Adjusted for age, BMI, and 5 PCs.
SNPs are limited to ≥5% MAF.
a: Prevalence estimates of fibroid risk is set at 70%.
*Statistically significant

Table 3.3. Heritability estimates from partitioned SNPs into intergenic and genic regions.

Location	Heritability Estimate ± (Std. Err.)	Number of SNPs[a]
Intergenic	0.21 ± 0.15	280
Genic	0.12 ± 0.14	169

Model: Adjusted for age, BMI, and 5 PCs.
SNPs are limited to ≥5% MAF.
Prevalence estimates of fibroids are set at 70%.
a: In 10,000 SNPs

Table 3.4. Heritability estimates of fibroid risk with suggestive to significant regions of the genome removed from prior studies.

Studies	Heritability Estimate ± (Std. Err.)
No Studies	0.33 ± (0.18)
Cha et al. (2011) (50)	0.33 ± (0.18)
Hellwege et al. (2017) (51)	0.33 ± (0.18)
All Studies (50, 51)	0.34 ± (0.18)

Model: Adjusted for age, BMI, and 5 PCs.
SNPs are limited to ≥5% MAF.
Prevalence estimates of fibroids are set at 70%.

CHAPTER 4

ADMIXTURE MAPPING OF UTERINE FIBROID SIZE AND NUMBER IN AFRICAN AMERICAN WOMEN

INTRODUCTION

As discussed in Chapter 1, racial disparities in fibroid risk and characteristics exist. For example, in addition to AA women having an approximately two- to three-fold higher risk of fibroids when compared to EA women (7), AA women also have more numerous and larger fibroids (2). Additionally, AAs are two times more likely than EAs to receive surgical treatments for fibroids such as a hysterectomies (8).

While the heritability of specific fibroid characteristics, such as fibroid size and number, is unknown, heritability estimates of fibroid risk from twin studies have ranged between 26% and 69% (19, 20). Additional support for genetic etiology for fibroids comes from racial differences in fibroid risk (2, 7, 14), as well as well as the racial differences in fibroid size and number between AA and EA women. A few studies have shown a direct relationship between increasing fibroid size and gene variants (53, 65). Edwards et al. (2013) observed associations between increasing fibroid size in EAs with gene variants in *TNRC6B* and Bet1 golgi vesicular membrane trafficking protein like (*BET1L*) (53) that were originally found in a genome-wide association study (GWAS) of fibroid risk (50). Aissani et al. (2015) showed associations between fibroid risk and largest fibroid dimension when evaluating a set of candidate gene variants (65).

Admixture mapping is an analytic approach in genetics to evaluate the relationship

between genetic ancestry and disease risk. Admixture mapping analyses are performed using admixed populations such as AAs where there are known prevalence differences in disease risk across racial groups. AA women have on average approximately 80% African ancestry and 20% European ancestry (66). Admixture mapping has been successfully applied in studies examining multiple sclerosis (67), keloids (68), and prostate cancer (69) in AA populations. A few previous studies have performed admixture mapping analyses on fibroid risk using AA individuals (70, 71). In the first study by Wise et al. (2012), the authors performed an admixture mapping study using ultrasound- or image-confirmed 2,453 cases and 2,102 controls with no fibroid diagnosis from the prospective cohort, the Black Women's Health Study, with women throughout the US. Using ANCESTRYMAP (72-74) and ADMIXMAP (75), the authors found that the mean percentage of African ancestry was significantly higher in fibroid cases when compared to controls but did not find a region in the genome that was significantly associated with fibroid risk (70). The authors did, however, find suggestive associations in chromosomal regions 2q33, 4p16, and, 10q26 (70). In the second admixture mapping study on fibroid risk by Zhang et al. (2015), the authors performed a cross-sectional study using 393 ultrasound-confirmed cases and 132 ultrasound-confirmed controls from the National Institute of Environmental and Health Sciences-Uterine Fibroid Study. Using ADMIXMAP (75), the authors did not find a significant association between global ancestry and fibroid risk. Zhang et al. (2015) did find a region within chromosome 1q42.2 with suggestive to significant associations where each African allele increased risk after stratifying by BMI (71). In the most recent admixture mapping study by Giri et al. (2017), the authors performed a cross-sectional study using AA women from BioVU and the CARDIA cohorts. The authors found that BMI interacts with local European ancestry and fibroid risk in AA women in two genomic regions, 6p24 and 2q31-31 (76).

Fibroids are a heterogeneous disease. Each fibroid characteristic difference, such as single versus multiple fibroids or a small versus large fibroid, could be affected by a set of genetic loci. A study examining fibroid characteristics might have better power to detect genetic determinants of fibroid subphenotypes that might be more closely related to a potential targeted treatment than a genetic study on fibroid risk. To the best of our knowledge, no study has performed an admixture mapping analysis on fibroid characteristics in AA individuals. The objective of this study is to examine the relationship between African ancestry and fibroid characteristics, namely size and number.

MATERIALS AND METHODS

Study Population

CARDIA Cohort

The CARDIA cohort was initiated in between 1985 and 1986 with the goal of measuring risk factors for coronary heart disease in a cohort of AA and EA individuals (77). The cohort consists of 5,115 AA and EA participants between 18 and 30 years of age who were selected based on approximately equal proportions of 18 to 24 and 25 to 30 year olds, sex, race (black and white), and education status with respect to high school graduation. Cohort recruitment took place at four places in the US: Birmingham, AL, Chicago IL, Minneapolis, MN, and Oakland, CA (77).

CARDIA Women's Study (CWS) is an ancillary study of CARDIA that conducted pelvic ultrasounds among women in the CARDIA cohort at 16 years following enrollment. The goal of CWS was to evaluate the association between risk factors of polycystic ovary syndrome and cardiovascular disease. Largest fibroid dimensions, fibroid number, and other relevant data to

our project was collected and recorded by trained CWS research staff (78). A transvaginal ultrasound was performed by sonographers who were certified by the American Registry of Diagnostic Medical Sonographers (ARMDS) and who had performed at least 50 prior transvaginal ultrasound examinations. The sonographers used a 5-7.5 MHz transvaginal probe. The dimensions of the largest fibroid were measured and number of fibroids was noted (78).

Our analyses used lifestyle and sociodemographic information that was collected via self and interviewer administered questionnaires (78). Measurements for height and weight were collected using a standardized protocol described previously (79). This study was limited to AA women with fibroids only.

BioVU

The BioVU database was described in Chapter 3. We applied the same phenotyping algorithm (21) as was described in Chapter 2, and limited the algorithm to AA individuals only.

Outcome measurements for analyses include: largest dimension of all fibroid measurements, volume of largest fibroid, and number of fibroids (single vs. multiple). To obtain an accurate estimate on fibroid volume, we used the following equation to calculate the volume of an ellipsoid for both CARDIA and BioVU cohorts: (Length x Width x Height x 0.523). The product of the three dimensions was multiplied by 0.523 to estimate volume assuming an ellipsoid shape. The total volume measurement and largest dimension were $\log_{10}$ transformed to create a normally distributed outcome for regression analysis. Some individuals with volume measurements (BioVU - 33.6%) originally had only two measurements for their largest fibroid, but we imputed the third measurement by taking the average of the first two measurements.

The hypothesis of these analyses was that there are different genetic risk factors for specific fibroid characteristics, for example you may have a different genetic risk factor for

having a larger versus multiple fibroids. Since there are known racial differences in fibroid number and size and since race is genetically determined, we evaluated the role of local genetic ancestry in risk for specific fibroid characteristics. These analyses are intended to better understand racial differences in fibroid phenotypic heterogeneity. Comparing subclasses of cases to controls would not necessarily tell us if there are within case subphenotype (fibroid characteristics) differences, as evidence of an association from a case-control analysis may still be a result of fibroid risk and not risk specific to a fibroid subphenotype. Because of this, the individuals in this study are limited to fibroid cases only. This study has been evaluated and approved by the VUMC IRB.

Genotyping

CARDIA AA participants were genotyped as part of the Candidate Gene Association Resource (CARe) study using the Affymetrix 6.0 array (Affymetrix, Santa Clara, CA). BioVU AA participants were genotyped using the Affymetrix Axiom Biobank array (Affymetrix, Inc., Santa Clara, CA) and on the Axiom World Array 3 platform (Affymetrix, Inc., Santa Clara, CA). DNA was purified and quantitated by PicoGreen (Invitrogen, Inc., Grand Island, NY).

GWAS Quality Control (QC)

The same QC protocol was performed on both CARDIA and BioVU populations separately using PLINK1.7 software (58) and using the reference genome build GRCh37.p13 (**Appendix H, Appendix I**). The following steps were taken in our quality control analysis: (1) dropped subjects with inconsistent genetic versus reported sex, (2) dropped related subjects, meaning all individuals with greater than a 0.95 probability of IBD and only one individual from a pair with a probability (IBD) between 0.2 and 0.95, (3) dropped SNPs without chromosomal locations, (4) dropped SNPs with a MAF of equal to or less than 1%, (5) dropped SNPs and

subjects with low genotyping efficiency (≤95%), and (6) dropped SNPs with a Hardy-Weinberg Equilibrium p-value ≤ $1x10^{-6}$. Alleles were aligned to the genomic + strand using the 1000 Genomes (build 37, 2013).

Statistical Analyses

Ancestry Estimation

Demographic and covariate information were summarized using Stata/SE (College Station, Texas). PCs of ancestry were estimated using EIGENSTRAT4.2 (**Appendix J, Appendix K**) (61). Assigning ancestry to SNPs using allele frequencies from the 1,000 Genomes Project as proxies to population allele frequencies (Phase 3 1,000 Genomes reference panels) (80) was accomplished using LAMP-ANC (Local Ancestry in adMixed Populations - ANCestral) (81-83). We used SNPs whose allele frequency difference between African ancestry and European ancestry was ≥20% using PLINK1.7 software (58). Local ancestry was estimated using the following criteria as input: seven generations since admixture event, recombination rate at $1x10^{-8}$, average ancestry composition per individual at 0.8 for African and 0.2 for European, respectively, proportion of overlap between windows of ancestry inference at 0.2, and r-squared threshold for LD-pruning at 0.1. Finally, local ancestry was coded as the number of European ancestry calls per each locus (0, 1, or 2). Global ancestry (percentage of European ancestry) was calculated for each individual by taking the number of European ancestry calls for all markers and dividing that number by the total number of ancestry calls.

Global and Local Ancestry Analyses

Linear and logistic regression was performed for each outcome for BioVU and CARDIA, respectively, using PLINK 1.7 software (58). Outcomes included $\log_{10}$ transformed volume of largest fibroid, $\log_{10}$ transformed largest dimension of all fibroids, and number of fibroids (single

vs multiple). The exposure included global or local ancestry. Fixed effects inverse-variance weighted meta-analysis comparing global or local ancestry across CARDIA and BioVU AAs was then performed for each outcome, respectively, using METAL software (84).

Quantile-quantile (QQ) plots of local ancestry analyses were created for each outcome (**Appendix L, Appendix M, Appendix N, Appendix O, Appendix P, Appendix Q, Appendix R, Appendix S, Appendix T**). A p-value of 0.05 or less denoted significance for all analyses involving global ancestry. The significance threshold for analyses of local ancestry was determined using 10,000 permutation tests for each outcome independently using PLINK1.7 software (58). The significance threshold for largest fibroid dimension, volume of largest fibroid, and fibroid number were determined to be 2.05×10^{-5}, 1.98×10^{-5}, and 4.80×10^{-5}, respectively. The suggestive threshold was found by taking two $\log_{10}$ down from the significance threshold. We adjusted for age and BMI for all analyses involving global genetic ancestry and adjusted for age, BMI, and five PCs for all analyses involving local genetic ancestry. We adjusted for BMI because BMI was found to be a confounder in a previous admixture mapping study on fibroid risk (71). We also performed admixture mapping analyses between local ancestry and fibroid number and volume for the reported admixture mapping peaks without adjustment for BMI and note that this did not significantly alter the association signals at these chromosomes (**Appendix U, Appendix V, Appendix W, Appendix X, Appendix Y, Appendix Z, Appendix AA**). We also performed local ancestry analysis adjusting for the most significant single SNP association within the admixture mapping peak regions in order to identify any genetic/imputed SNPs that explain the ancestry peaks.

Single SNP Association Analyses

Non-genotyped SNPs in the suggestive mapping peak regions were imputed, and single SNP associations on BioVU and CARDIA were performed for each outcome, respectively. Suggestive mapping peak regions were defined as one $\log_{10}$ down from the most significant marker in a mapping peak above the suggestive threshold.

Suggestive peaks were further evaluated for single SNP associations in separate analyses by characteristics. A fixed effects inverse-variance weighted meta-analysis was performed comparing the single SNP association results across CARDIA and BioVU AAs using METAL software (84). The significance threshold for the single SNP association analyses was determined by calculating the effective number of independent SNPs among all genotyped SNPs within suggestive mapping peak regions of each outcomes, respectively, using simpleM software (85-87). The significance threshold for fibroid number and volume were determined to be 8.78×10^{-6} and 5.42×10^{-5}, respectively. We adjusted for age, BMI, and five PCs for all analyses involving single locus test of association.

RESULTS

Demographic Data

There were 171 AAs in CARDIA and 438 in BioVU with information on fibroid characteristics (**Table 4.1**). The mean age and BMI were similar among AAs in CARDIA (age: 41.3±4 years; BMI: 33.4±8) and BioVU (age: 41.5±11 years; BMI: 32.9±8), and the majority of subjects were obese (BioVU: 60%; CARDIA: 64%). The median volume of largest fibroid and largest fibroid dimension was smaller for CARDIA (volume: 5.6 cm^3; largest dimension: 2.5 cm) than for BioVU AAs (volume: vs 19.7 cm^3; largest dimension: 3.6 cm). Additionally, CARDIA AAs were more likely to have multiple fibroids than BioVU AAs (71% vs 58%). Finally, the

mean percentage of European ancestry was similar for CARDIA and BioVU AAs (14.9% vs 15.5%).

Global Ancestry Analyses

We observed a significant association between percentage of European ancestry and number of fibroids. A 10% decrease in global European ancestry was significantly associated with multiple fibroids (OR: 0.78; 95% CI: 0.66, 0.93; $p = 6.05 \times 10^{-3}$) (**Table 4.2**). There were no significant associations in the meta-analyses between percentage of global European ancestry and volume of largest fibroid or largest dimension of all fibroids for BioVU and CARDIA AAs, although a 10% decrease in global European ancestry was near significantly associated with largest dimension (Beta: -0.029; 95% CI: -0.059, 0.001) (**Table 4.2**).

Local Ancestry Analyses

We did not observe associations between local ancestry and number of fibroids, volume of largest fibroid, or largest dimension of all fibroids that were statistically significant after multiple comparisons. We did, however, observe five suggestive associations for number of fibroids ($p<4.80 \times 10^{-3}$) and one suggestive association for volume of largest fibroid ($p<1.97 \times 10^{-3}$) (**Table 4.3**). Additionally, there were several statistically significant single SNP associations within the admixture mapping regions for number of fibroids and volume of largest fibroid.

The most significant admixture mapping signal was seen in number of fibroids within 10q21.1 near inositol polyphosphate multikinase (*IPMK*), where each European chromosome decreased the odds of multiple fibroids (OR: 0.51; 95% CI: 0.36, 0.74; $p = 3.23 \times 10^{-4}$) (**Table 4.3, Figure 4.1**). The most significant single SNP in this region was rs12219990, where each effect allele decreased the odds of multiple fibroids (OR: 0.41; 95% CI: 0.28, 0.60; $p = 3.82 \times 10^{-6}$) (**Table 4.4**). After adjusting for the most significant SNP from the single SNP association

analyses, rs12219990, the original admixture mapping signal was reduced (OR: 0.64; 95% CI: 0.43, 0.94; p = 0.0214) (**Table 4.3, Figure 4.1**) suggesting the admixture mapping peak was due to this SNP. The effect allele for rs12219990 was more common among Europeans (35%) than among Africans (7%) (data not shown).

There was one suggestive admixture mapping peak in volume of largest fibroid in region 10q24.1-10q24.32 near leucine zipper tumor suppressor 2 (*LZTS2*), where each European allele decreased the probability of larger of fibroids (beta: -0.23; 95% CI: -0.27, -0.09; $p = 1.48x10^{-3}$) (**Table 4.3, Appendix BB**). The most significant single SNP in this region was rs4919512, where each effect allele decreased the probability of larger fibroids (Beta: -0.25; 95% CI: -0.37, -0.13; $p = 2.82x10^{-5}$) (**Table 4.4**). After adjusting for the most significant SNP from the single SNP association analyses, rs4919512, the original admixture mapping signal was slightly reduced (beta: -0.12; 95% CI: -0.27, 0.08; p = 0.145) (**Table 4.3, Appendix BB**).

DISCUSSION

This is the first admixture mapping study to examine the effects that African ancestry has on uterine fibroid characteristics. In this study we observed a strong inverse association between mean European ancestry and fibroid number, where increasing global African ancestry increases the odds of having multiple fibroids. While there were no statistically significant local ancestry analyses of any fibroid characteristic, there were multiple suggestive regions of fibroid number (**Figure 4.1, Appendix CC, Appendix DD, Appendix EE, Appendix FF**) and volume (**Appendix BB**). Additionally, there were two statistically significant single SNP associations: one on 10q21.1 for fibroid number (rs12219990) and another on 10q24.31 for fibroid volume

(rs4919512). Furthermore, the local ancestry analyses highlighted genes with potential biological significance in etiology of fibroid characteristics.

There were five genes (sin3A association protein 130 [*SAP130*], proteasome 26S subunit, non-ATPase 6 [*PSMD6*], plexin A4 [*PLXNA4*], *IPMK*, and SNW domain containing 1 [*SNW1*]) identified by local ancestry analyses of fibroid number and one gene (*LZTS2*) identified from our analyses of fibroid volume (**Table 4.3**). Four out of the five genes identified in the fibroid number analyses (*SAP130*, *PLXNA4*, *IPMK*, and *SNW1*) have been previously associated with cancer susceptibility or tumor growth (88-91). While the fifth gene (*PSMD6*) was previously associated with delay in DNA repair when depleted (92). The only gene, *LZTS2*, that was identified by our analyses of fibroid volume has also been attributed in many cancers (93) where depletion of *LZTS2* increases probability of tumorigenesis (94). Further analyses evaluating the expression level of these genes in the Genotype-Tissue Expression (GTEx) project database, which aims to characterize the relationship between tissue-specific gene expression and genotype (95), demonstrated that *LZTS2*, *PSMD6*, *SAP130*, and *SNW1* were expressed (reads per kilobase of transcript per million mapped reads [RPKM] > 5) in uterine tissue (GTEx Analysis Release V6p (dbGaP Accession phs000424.v6.p1)) (95). Out of 53 tissues in total, *LZTS2* was expressed more in the uterus than any other tissue. This supports our observed data that genetic variation around *LZTS2* may decrease gene expression leading to a susceptibility of fibroid tumor growth.

We also evaluated potential candidate regions that have been implicated in prior studies of fibroids and observed that the region 12q14.1-q14.3 contained a small admixture mapping peak associated with fibroid number ($p = 7.21 \times 10^{-3}$) and included *HMGA2* (**Appendix GG, Appendix HH, Appendix II**), a gene that was previously implicated in fibroid risk in studies of the eker rat. The eker rat represents an animal model that spontaneously forms fibroid tumors

similar to that of humans (47). Both the human *HMGA2* and eker rat homolog are atypically expressed in fibroid tumors (47). The HMGA proteins are part a family of transcription factors (47, 96). Progesterone and estrogen receptor activity has been shown to be regulated by the HGMA subfamily (47, 97-100).

The prior admixture mapping studies by Wise et al. (2012), Zhang et al (2015), and Giri et al. (2017) identified several suggestive associations (70, 71, 76); none overlapped with our study findings. However, the Wise et al. (2012) global ancestry analyses showed a significant inverse association between global European ancestry and fibroid risk (70), consistent with our findings suggesting that African ancestry is not only associated with the development of a fibroid (or fibroid risk) but with the development of multiple fibroids. The second admixture study by Zhang et al. (2015) used 393 AA cases finding no association between global European ancestry and fibroid risk (71).

A difference between the two prior admixture mapping studies on uterine fibroids (70, 71) and ours included that our study used GWAS data, allowing us to conduct single SNP association analyses within identified regions. Wise et al. (2012) and Zhang et al. (2015) used panels of ancestry informative markers.

Previous estimations of ancestry place AA individuals having about 20% European ancestry and 80% African ancestry (66). The mean percentage of European ancestry for our study, however, was 15.5% for BioVU AAs and 14.9% for CARDIA AAs. The lower average percentage of European ancestry in our study could be an artifact of enriching for a trait (i.e. fibroids) that is more common in AAs (2).

There were limitations to our study. Our sample size was modest for each cohort (BioVU N=438; CARDIA N=171) matching the cohort size for admixture mapping study by Zhang et al.

(2015) (N=525). We, however, performed meta-analyses between BioVU and CARDIA individuals which served as a form of replication and validation for our findings. Additionally, all fibroid characteristic information was assessed either via ultrasound or surgery leading to a decrease in outcome misclassification. In addition, we did not perform subanalyses limiting to either image- or surgery-confirmed fibroid cases since all CARDIA individuals in this study had only ultrasounds to assess fibroid characteristics and since only some BioVU individuals had this information recorded during the abstraction process. It could be possible that BioVU individuals whose fibroids were discovered during surgery had a different tumor characteristic profile than women whose fibroids were confirmed via ultrasounds which could introduce bias in the analyses. Including women who had surgery could possibly inflate the effect sizes since these women may be more likely to have severe symptoms that may be due to larger size and/or number of fibroids. Finally, there is no overlap between the previous two previous admixture mapping studies on fibroid risk using AAs (70, 71) and our study. This could be because our study examined genetic risk factors via ancestry for differences in fibroid size and number and that these genetic factors via ancestry differ from the genetic factors for fibroid incidence.

Our study was the first admixture mapping study to access the association between ancestry and fibroid characteristics using AA populations. We found that global ancestry influences the number of fibroids and found many suggestive mapping peaks influencing fibroid number and one mapping peak influencing fibroid volume. Further studies needs to be performed to understand the biological mechanisms underlying the observed genetic associations.

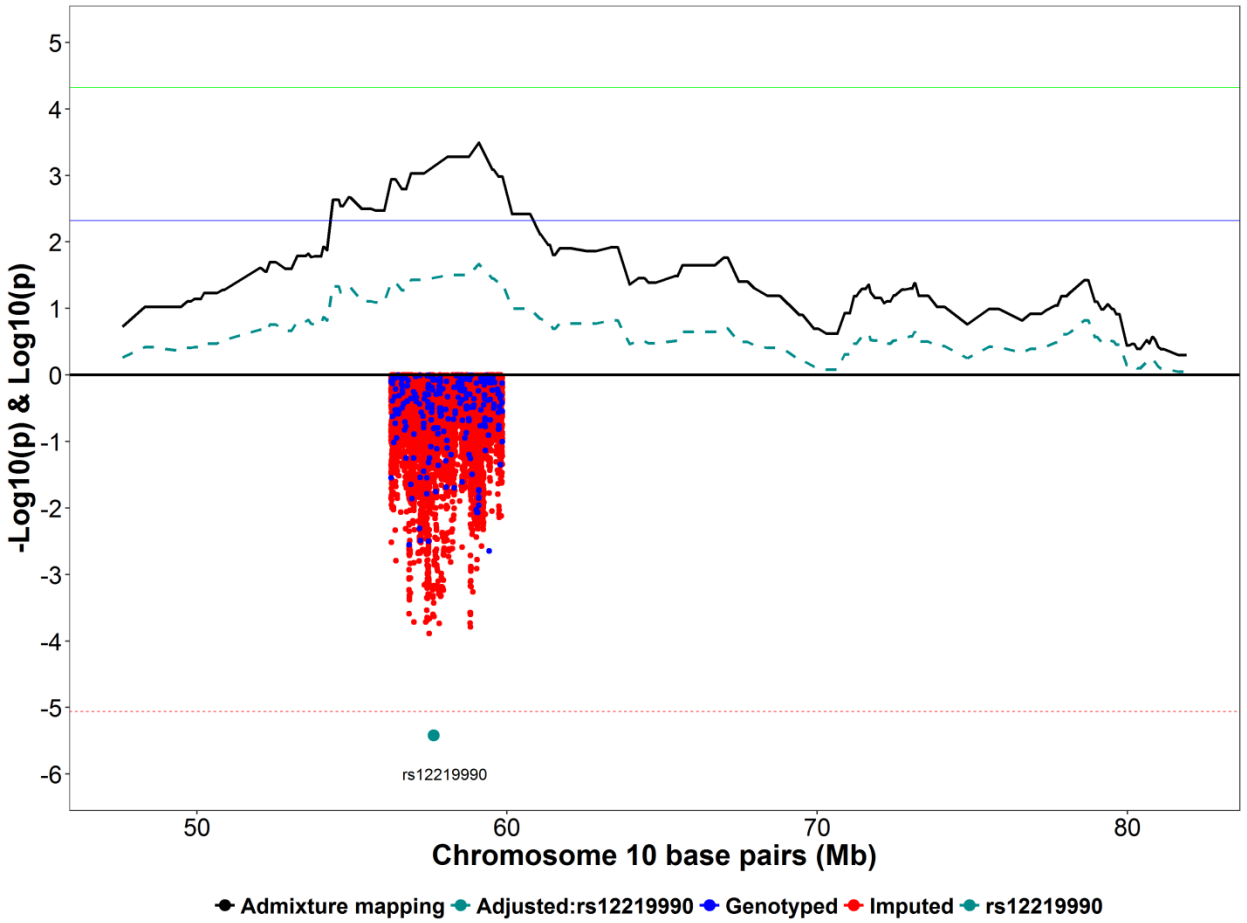


Figure 4.1. Admixture mapping analysis of chromosome 10 with overlapping single SNP association analysis results.

The X-axis indicates the genomic position along chromosome 10 in Mb. The top of the Y-axis indicates -$\log_{10}$(*p* value) from the meta-analysis between BioVU and CARDIA AA from logistic regression of admixture mapping values that were generated from LAMP-ANC (solid black line – not conditioned for rs12219990; dashed green line – conditioned on rs12219990) with fibroid number (single vs multiple) as the outcome. The admixture mapping analysis was adjusted for age, BMI, 5 PCs. The solid green line represents the significance threshold. The solid blue line represents the suggestive threshold. The bottom portion of the Y-axis indicates the $\log_{10}$(*p* value) for the single SNP association analyses (blue circles indicates genotyped SNPs; red circles indicates imputed SNPs) with fibroid number (single vs multiple) as the outcome. The imputed region was found by taking 1 –log down from the most significant mapping peak above the suggest threshold. The imputed region encompassed 56,258,202 to 59,854,651 bp. The single SNP association analysis was adjusted for age, BMI, and 5 PCs. The SNP, rs12219990, was imputed (BioVU info score – 0.921; CARDIA info score – 0.922). The dotted red line represents the significance threshold.

Table 4.1. Demographics of CARDIA and BioVU African Americans.

Demographic Characteristics	N	All (N=609)	BioVU (N=438)	CARDIA (N=171)
Age (mean±SD)	609	41.4±9	41.5±11	41.3±4
BMI (kg/m2) (mean±SD)	606	33.1±8	32.9±8	33.4±8
Underweight (<18.5) (%)	6	1	1	1
Normal weight (18.5-24.9) (%)	87	14	16	11
Overweight (25-29.9) (%)	142	23	23	24
Obese (≥30) (%)	371	61	60	64
Fibroid Volume (cm^3) median (IQR)	492	14.1 (3.6-47.2)	19.6 (5.3-76.0)	5.6 (2.1-19.3)
Largest Fibroid Dimension (cm) median (IQR)	562	3.1 (2.0-5.0)	3.5 (2.2-5.7)	2.5 (1.8-3.7)
Fibroid Number	588			
1 (%)	226	38	42	29
>1 (%)	362	62	58	71
Percentage of European Ancestry (mean±SD)	609	15.3±10	15.5±10	14.9±10

BMI-body mass index; kg/m^2-kilograms per meters squared; cm^3-cubic centimeters; cm- centimeters; SD-standard deviation; IQR-interquartile range.

Table 4.2. Associations between exposure of mean percentage of European ancestry and outcome.

Number of Fibroids [b]	
Cohort	**Adjusted OR [95% CI] [a]**
BioVU	0.77 [0.62, 0.94]*
CARDIA	0.83 [0.59, 1.16]
Meta-analysis	0.78 [0.66, 0.93]**
Volume of Largest Fibroid [c]	
Cohort	**Adjusted Beta [95% CI] [a]**
BioVU	-0.043 [-0.137, 0.052]
CARDIA	0.022 [-0.081, 0.126]
Meta-analysis	-0.013 [-0.082, 0.056]
Largest Dimension of All Fibroids [d]	
Cohort	**Adjusted Beta [95% CI] [a]**
BioVU	-0.029 [-0.059, 0.001]
CARDIA	0.009 [-0.024, 0.043]
Meta-analysis	-0.012 [-0.034, 0.011]

a Adjusted for age and BMI;
b Number is coded as single vs multiple fibroids;
c Volume is coded as $\log_{10}$ transformed largest fibroid volume in cm^3;
d Largest Dimension is coded as $\log_{10}$ transformed largest fibroid dimension of all fibroid measurements in cm;
* $p<0.05$, ** $p<0.007$;
Exposure is in 10% increments of mean percentage of European ancestry; OR-odds ratio; CI-confidence interval; p-p-value, cm^3-cubic centimeters; cm- centimeters.

Table 4.3. Admixture mapping for number of fibroids and volume of largest fibroid in African American women.

Number [c]					
		10q21.1			
Nearby Gene	**Cohort**	**OR [95% CI] [a]**	**p [a]**	**OR [95% CI] [b]**	**p [b]**
	BioVU	0.54 [0.62, 1.44]	$3.36x10^{-3}$	0.68 [0.44, 1.04]	0.0764
IPMK	CARDIA	0.41 [0.18, 0.92]	0.030	0.50 [0.21, 1.17]	0.110
	Meta-analysis	0.51 [0.36, 0.74]	$3.23x10^{-4}$	0.64 [0.43, 0.94]	0.0214
		14q24.2-14q24.3			
Nearby Gene	**Cohort**	**OR [95% CI] [a]**	**p [a]**	**OR [95% CI] [b]**	**p [b]**
	BioVU	2.07 [1.34, 3.18]	$9.89x10^{-4}$	1.81 [1.17, 2.82]	$8.12x10^{-3}$
SNW1	CARDIA	1.62 [0.74, 3.55]	0.226	1.54 [0.69, 3.44]	0.289
	Meta-analysis	1.95 [1.34, 2.85]	$5.25x10^{-4}$	1.75 [1.19, 2.57]	$4.66x10^{-3}$
		2q14.3-2q21.1			
Nearby Gene	**Cohort**	**OR [95% CI] [a]**	**p [a]**	**OR [95% CI] [b]**	**p [b]**
	BioVU	0.64 [0.43, 0.95]	0.026	0.74 [0.48, 1.13]	0.163
SAP130	CARDIA	0.47 [0.24, 0.93]	0.030	0.85 [0.37, 1.96]	0.701
	Meta-analysis	0.59 [0.42, 0.83]	$2.60x10^{-3}$	0.76 [0.52, 1.11]	0.157
		7q32.2-7q33			
Nearby Gene	**Cohort**	**OR [95% CI] [a]**	**p [a]**	**OR [95% CI] [b]**	**p [b]**
	BioVU	2.02 [1.33, 3.08]	$1.05x10^{-3}$	1.85 [1.20, 2.86]	$5.70x10^{-3}$
PLXNA4	CARDIA	1.02 [0.46, 2.25]	0.964	0.51 [0.20, 1.27]	0.149
	Meta-analysis	1.74 [1.20, 2.52]	$3.54x10^{-3}$	1.46 [0.98, 2.16]	0.0604
		3p14.2-3p14.1			
Nearby Gene	**Cohort**	**OR [95% CI] [a]**	**p [a]**	**OR [95% CI] [b]**	**p [b]**
	BioVU	1.56 [1.01, 2.40]	0.045	1.45 [0.94, 2.25]	0.095
PSMD6	CARDIA	2.96 [1.25, 7.05]	0.014	2.23 [0.92, 5.37]	0.074
	Meta-analysis	1.77 [1.20, 2.60]	$3.86x10^{-3}$	1.58 [1.07, 2.34]	0.022
Volume [d]					
		10q24.1-10q24.32			
Nearby Gene	**Cohort**	**Beta [95% CI] [a]**	**p [a]**	**Beta [95% CI] [b]**	**p [b]**
	BioVU	-0.26 [-0.43, -0.08]	$4.62x10^{-3}$	-0.16 [-0.35, 0.04]	0.116
LZTS2	CARDIA	-0.18 [-0.42, 0.06]	0.138	-0.04 [-0.30, 0.22]	0.754
	Meta-analysis	-0.23 [-0.27, -0.09]	$1.48x10^{-3}$	-0.12 [-0.27, 0.08]	0.145

a Model 1: Adjusted for age, BMI, and 5 PCs;
b Model 2: Adjusted for age, BMI, and 5 PCs + adjustment for most significant single SNP;
c Number is coded as single vs multiple fibroids;
d Volume is coded as $\log_{10}$ transformed largest fibroid volume in cm^3;
OR-odds ratio; CI-confidence interval; p-p-value; cm^3-cubic centimeters.

Table 4.4. Single SNP signals in admixture mapping regions for number of fibroids and volume of largest fibroid in African American women.

Number [b]				
rs12219990 in 10q21.1*				
Cohort	**OR [95% CI]** [a]	**p**	**EA/RA**	**BioVU/CARDIA EAF**
BioVU	0.41 [0.27, 0.63]	$3.29x10^{-5}$	C/A	0.14/0.11
CARDIA	0.39 [0.15, 0.97]	0.042		
Meta-analysis	0.41 [0.28, 0.60] [a]	$3.82x10^{-6}$		
rs11394508 in 14q24.2-14q24.3				
Cohort	**OR [95% CI]** [a]	**p**	**EA/RA**	**BioVU/CARDIA EAF**
BioVU	0.50 [0.37, 0.68]	$6.02x10^{-6}$	CA/C	0.37/0.37
CARDIA	0.80 [0.46, 1.36]	0.402		
Meta-analysis	0.56 [0.43, 0.73]	$1.29x10^{-5}$		
rs6723563 in 2q14.3-2q21.1				
Cohort	**OR [95% CI]** [a]	**p**	**EA/RA**	**BioVU/CARDIA EAF**
BioVU	0.52 [0.31, 0.88]	0.015	A/G	0.09/0.07
CARDIA	0.15 [0.06, 0.40]	$1.41x10^{-4}$		
Meta-analysis	0.40 [0.25, 0.63]	$7.68x10^{-5}$		
rs782525 in 7q32.2-7q33				
Cohort	**OR [95% CI]** [a]	**p**	**EA/RA**	**BioVU/CARDIA EAF**
BioVU	1.47 [1.07, 2.01]	0.017	G/A	0.27/0.27
CARDIA	2.50 [1.48, 4.24]	$6.39x10^{-4}$		
Meta-analysis	1.69 [1.29, 2.21]	$1.43x10^{-4}$		
rs112428319 in 3p14.2-3p14.1				
Cohort	**OR [95% CI]** [a]	**p**	**EA/RA**	**BioVU/CARDIA EAF**
BioVU	0.40 [0.21, 0.05]	$8.01x10^{-3}$	G/A	0.05/0.06
CARDIA	0.14 [0.79, 0.41]	$3.59x10^{-4}$		
Meta-analysis	0.30 [0.17, 0.53]	$3.74x10^{-5}$		
Volume [c]				
rs4919512 in 10q24.1-10q24.32*				
Cohort	**Beta [95% CI]** [a]	**p**	**EA/RA**	**BioVU/CARDIA EAF**
BioVU	-0.25 [-0.40, -0.10]	$8.16x10^{-4}$	C/T	0.29/0.22
CARDIA	-0.25 [-0.45, -0.06]	0.012		
Meta-analysis	-0.25 [-0.37, -0.13]	$2.82x10^{-5}$		

a Model 1: Adjusted for age, BMI, and 5 PCs;
b Number is coded as single vs multiple fibroids;
c Volume is coded as $\log_{10}$ transformed largest fibroid volume in cm^3;
*significant meta-analysis after correction for multiple testing
OR-odds ratio; CI-confidence interval; p-p-value; EA-effect allele; RA-reference allele; EAF-effect allele frequency; cm^3-cubic centimeters.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

The experiments described in this dissertation aimed to highlight epidemiologic and genetic risk factors for uterine fibroid characteristics, namely number and size, as well as to estimate fibroid heritability of uterine fibroids. Most studies on fibroids focus on fibroid presence. Part of the reason for the lack of research on fibroid characteristics is the requirement of each patient to be systematically screened via ultrasounds or surgeries to measure and count fibroids. By using datasets within BioVU and CARDIA, we were able to perform a large scale epidemiologic analysis on fibroid characteristics, to perform the first GWAS and admixture mapping analysis on fibroid characteristics to scan the genome for associations with fibroid number and size, and to estimate and characterize fibroid heritability.

The aim of Chapter 2 was to evaluate physical and clinical features of women with fibroids in relation to selected fibroid characteristics. Using a dataset of patient EHRs from the Synthetic Derivative consisting of 2,302 women with image- or surgery-confirmed fibroids, we performed the largest association analyses between fibroid risk factors and fibroid characteristics. The fibroid risk factors included age, BMI, race (black or white), type 2 diabetes status, and number of living children (proxy to parity) and were abstracted from patient EHRs. The outcomes included fibroid number (single vs. multiple), volume, and largest dimension, which were manually abstracted of patient image- or surgery-reports. After observing no effect

modification by race in analyses of age, we observed that increasing age was associated with multiple and larger fibroids and that this effect changes nonlinearly with age. In addition, we observed that black race was associated with both multiple and larger fibroids. We observed no associations between fibroid characteristics and BMI or type 2 diabetes. These findings suggest that physical and clinical features of a woman may help to determine if and when she is at risk of developing fibroids with specific characteristics. These results could help pave the way towards the best route of fibroid treatment.

In Chapter 3, we were the first to estimate and characterize the heritability of fibroids using a dataset of 2,109 EA women from BioVU with image-confirmed fibroid status. We estimated heritability of fibroid risk was 0.33, being much higher than previous genetic estimates on fibroid heritability of approximately 9%. In addition, we observed that our fibroid heritability estimate of 33% was not attenuated when censoring genetic loci previously associated with fibroid risk. This suggests that many of the genetic loci that are associated with fibroid risk in EA women have yet to be discovered. We observed that fibroid heritability was not evenly distributed throughout the genome. Chromosome 8 was the most heritable chromosome, with most of the heritability of chromosome 8 being in 8q. Lastly, there was no enrichment of fibroid heritability for intergenic or genic regions.

The aim of Chapter 4 was to evaluate the relationship between genetic ancestry and uterine fibroid characteristics. We performed the first admixture mapping study to examine the effects that African ancestry has on uterine fibroid characteristics using 609 AA women from BioVU and CARDIA. In this study we observed a strong significant inverse association between

mean European ancestry and fibroid number, where increasing global African ancestry increases the odds of having multiple fibroids. While there were no statistically significant local ancestry analyses of any fibroid characteristic, there were six suggestive regions of fibroid number and volume. Additionally, there were two statistically significant single SNP associations within the admixture mapping peaks: one on 10q21.1 for fibroid number (rs12219990) and another on 10q24.31 for fibroid volume (rs4919512).

Future Directions

These Chapters explored the epidemiologic and genetic risk factors of uterine fibroid characteristics using a mixture of prospective cohort data (CARDIA) and de-identified EHRs (Synthetic Derivative and BioVU). Epidemiologic and genetic data from the Synthetic Derivative and BioVU is an invaluable resource. In particular, by using this resource we successfully performed the largest epidemiologic analyses of uterine fibroid characteristics using data from 2,302 individuals. There are some difficulties with handling this rich data resource. This includes the variable abstraction process. All the EHRs include digitalized hospital notes which can vary greatly depending on the nurse or physician recording the information. In addition, there was human error where information within a single patient's record was sometimes inconsistent. In studies that use this new resource consisting of patient EHRs, there needs to be credence that large sample sizes will be needed to offset of loss of power due to minor misclassification of data.

This study also performed the first admixture mapping analysis on fibroid characteristics. We observed many genetic association with fibroid size and number that were not associated previously associated with fibroid risk. This suggests that different genetic etiologies might exist between fibroid size and number and fibroid risk. Further genetic studies are needed to explore the underlying mechanisms observed in these genetic associations.

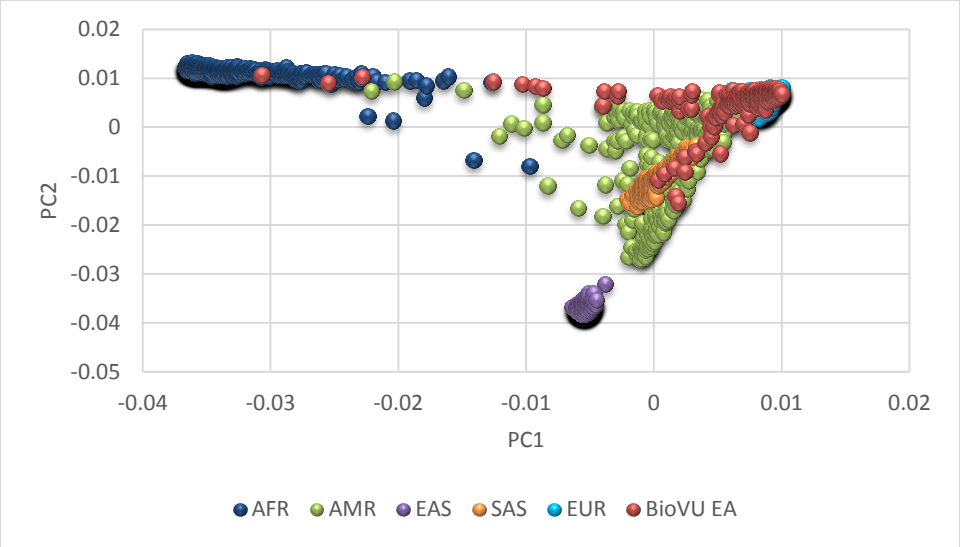
Lastly, we estimated the heritability of fibroids in a population of EA women with image-confirmed fibroid status. We observed that chromosome 8q explained the greatest proportion of variance for fibroid risk along with chromosomes 11 and 10. In addition, after censoring loci previously implicated with fibroid risk in populations of Japanese and AA women, there was no attenuation fibroid heritability. This entails that there are many genetic loci that increase fibroid risk in EA women that have yet to be discovered. Future genetic studies on fibroid risk are needed to understand fibroid heritability in EA women.

APPENDICES

Appendix A. Final SNP count for each outcome.

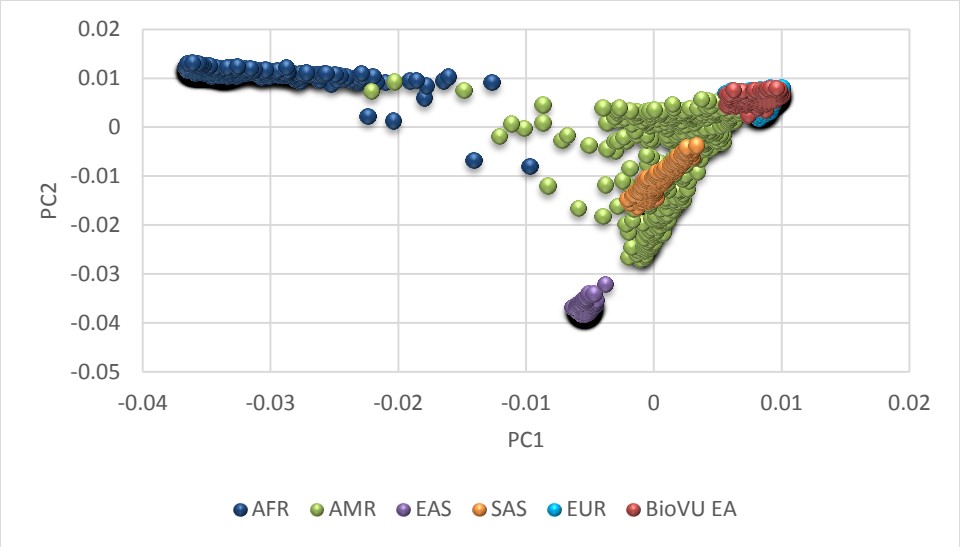
Outcome	Sample Size	Post QC SNP Number - MAF 5%
Fibroid Risk		4,500,362*
Case	1,075	
Control	1,059	
Volume	375	4,522,829*
Max Dimension	554	4,518,340*

*Post QC SNP Numbers - These are slightly different between studies because we removed SNPs out of HWE at the end for each outcome independently.



Appendix B. Principal Components Plot of BioVU EAs with Reference Population before Pruning Outliers.

AFR-African; AMR-Ad Mixed American; EAS-East Asian; SAS-South Asian; EUR-European.



Appendix C. Principal Components Plot of BioVU EAs with Reference Population after Pruning Outliers.

AFR-African; AMR-Ad Mixed American; EAS-East Asian; SAS-South Asian; EUR-European.

Appendix D. Heritability estimates for fibroid risk using European Americans within BioVU.

Heritability Estimate ± (Std. Err.)	P-value	Prevalence
0.24 ± (0.13)	0.040*	0.10
0.29 ± (0.16)	0.040*	0.20
0.33 ± (0.18)	0.040*	0.30
0.35 ± (0.19)	0.040*	0.40
0.35 ± (0.20)	0.040*	0.50
0.35 ± (0.19)	0.040*	0.60
0.33 ± (0.18)	0.040*	0.70
0.29 ± (0.16)	0.040*	0.80
0.24 ± (0.13)	0.040*	0.90

Model: Adjusted for age, BMI, and 5 PCs.
SNPs are limited to ≥5% MAF.
*Statistically significant

Appendix E. Heritability estimates for fibroid risk using European Americans within BioVU.

Heritability Estimate ± (Std. Err.)	P-value	PC Adjustments
0.38 ± (0.18)	0.013*	0
0.34 ± (0.18)	0.031*	1
0.35 ± (0.18)	0.027*	2
0.33 ± (0.18)	0.037*	3
0.33 ± (0.18)	0.038*	4
0.33 ± (0.18)	0.040*	5
0.33 ± (0.18)	0.040*	6
0.29 ± (0.18)	0.064	7
0.29 ± (0.18)	0.058	8
0.30 ± (0.18)	0.053	9
0.30 ± (0.18)	0.056	10

Model: Adjusted for age and BMI and noted PC number in table.
SNPs are limited to ≥5% MAF.
Prevalence estimates of fibroids are set at 70%.
*Statistically significant

Appendix F. Heritability estimates for largest fibroid dimension using European Americans within BioVU.

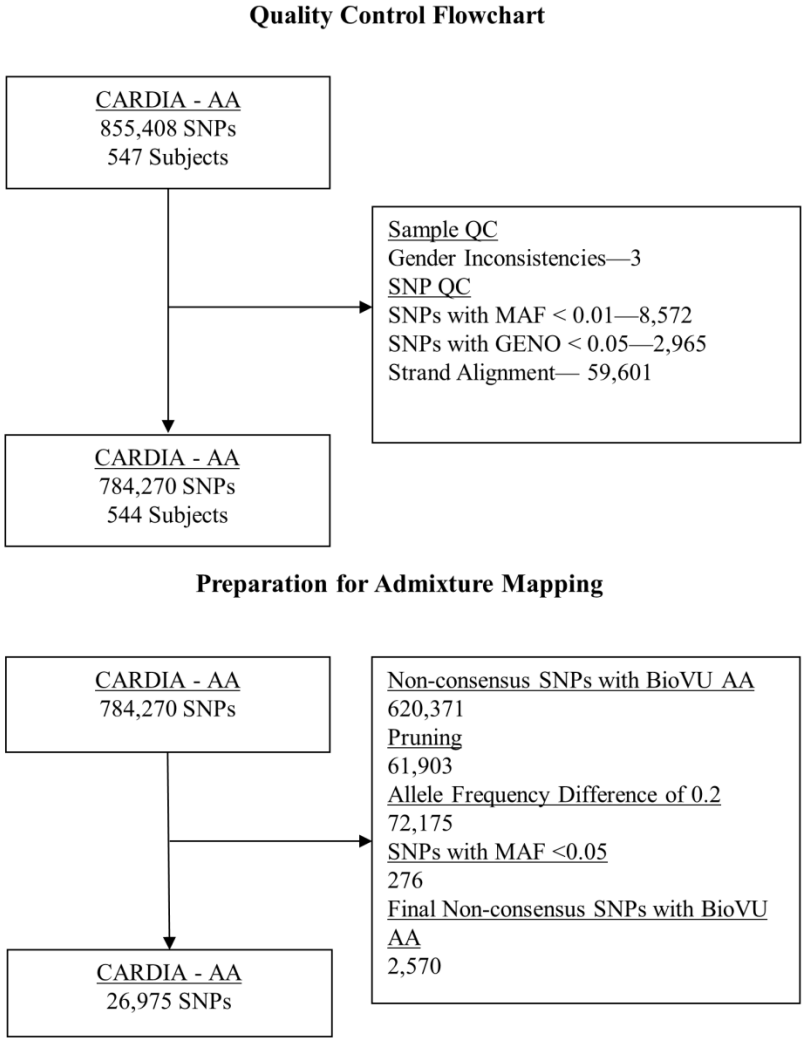
Heritability Estimate ± (Std. Err.)	P-value	PC Adjustments
0.34 ± (0.47)	0.246	0
0.40 ± (0.47)	0.209	1
0.40 ± (0.47)	0.209	2
0.40 ± (0.47)	0.209	3
0.39 ± (0.47)	0.212	4
0.35 ± (0.47)	0.238	5
0.36 ± (0.48)	0.233	6
0.38 ± (0.48)	0.224	7
0.39 ± (0.48)	0.214	8
0.43 ± (0.48)	0.195	9
0.44 ± (0.48)	0.190	10

Model: Adjusted for age and BMI and noted PC number in table.
SNPs are limited to ≥5% MAF.

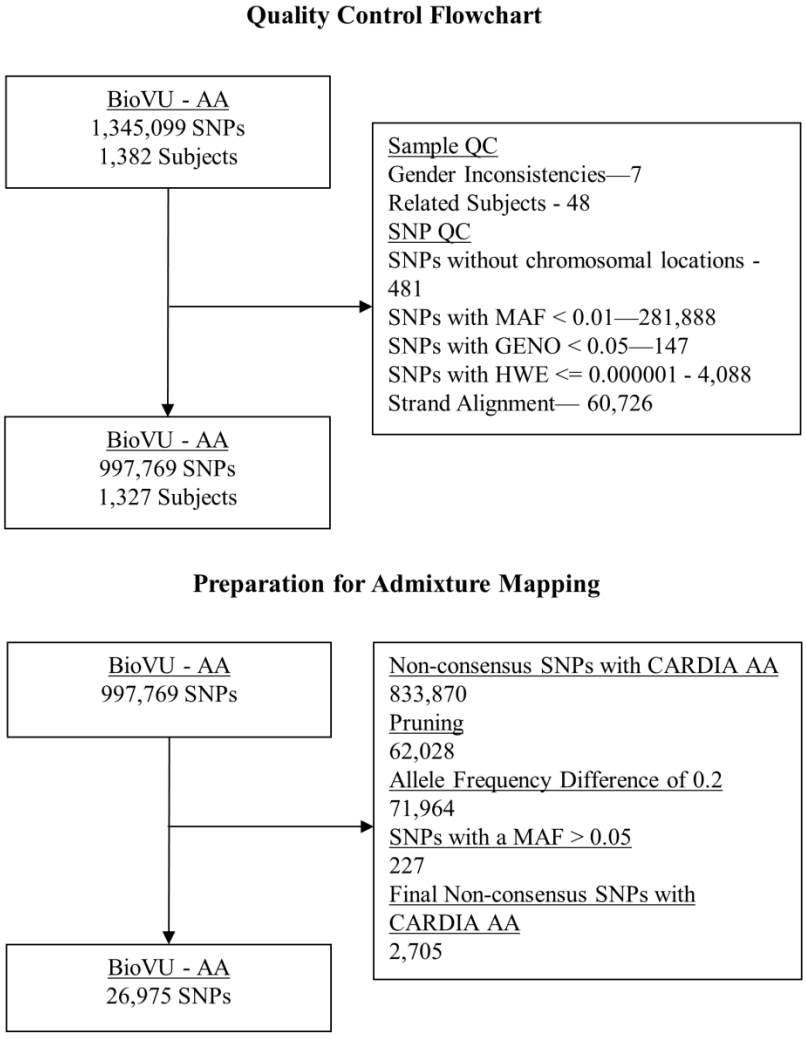
Appendix G. Heritability estimates for largest fibroid volume using European Americans within BioVU.

Heritability Estimate ± (Std. Err.)	P-value	PC Adjustments
0.12 ± (0.65)	0.426	0
0.18 ± (0.65)	0.391	1
0.18 ± (0.65)	0.389	2
0.12 ± (0.66)	0.427	3
0.13 ± (0.66)	0.423	4
0.14 ± (0.66)	0.417	5
0.16 ± (0.65)	0.404	6
0.17 ± (0.67)	0.399	7
0.24 ± (0.67)	0.355	8
0.28 ± (0.67)	0.334	9
0.35 ± (0.67)	0.300	10

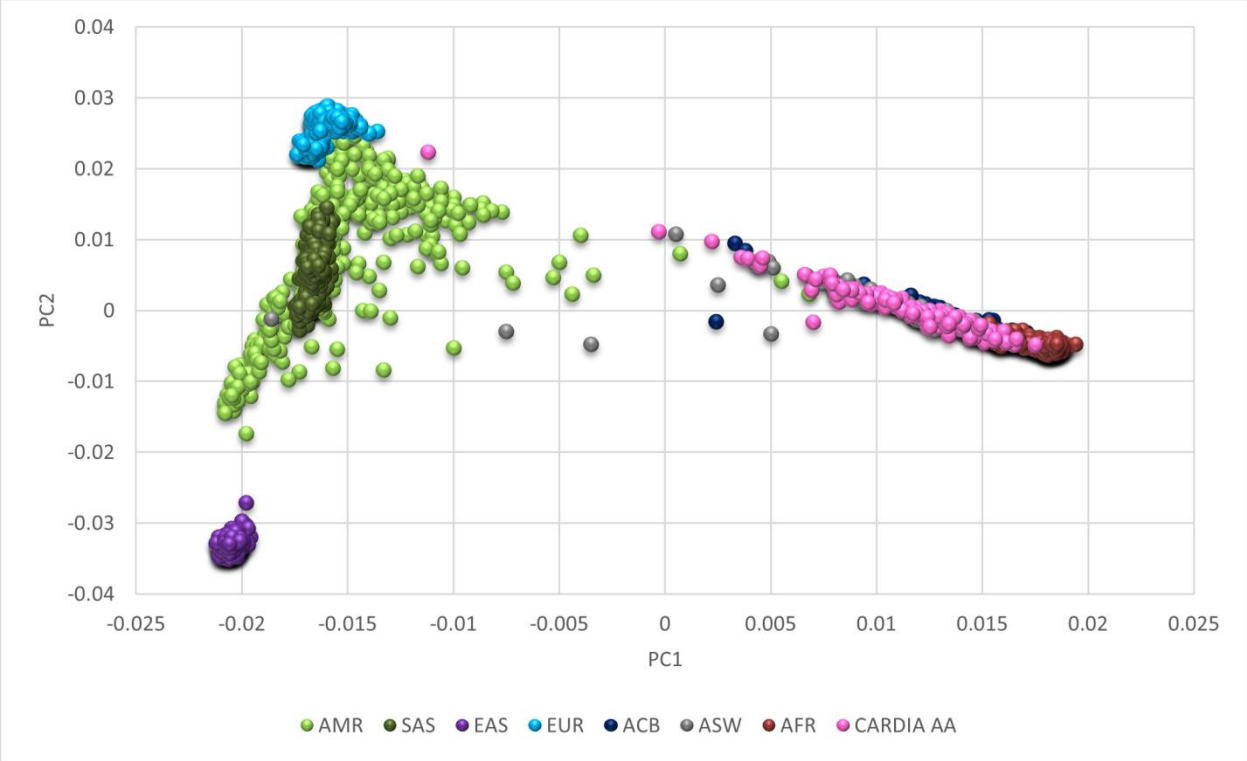
Model: Adjusted for age and BMI and noted PC number in table.
SNPs are limited to ≥5% MAF.



Appendix H. Quality control flowchart of CARDIA AA.

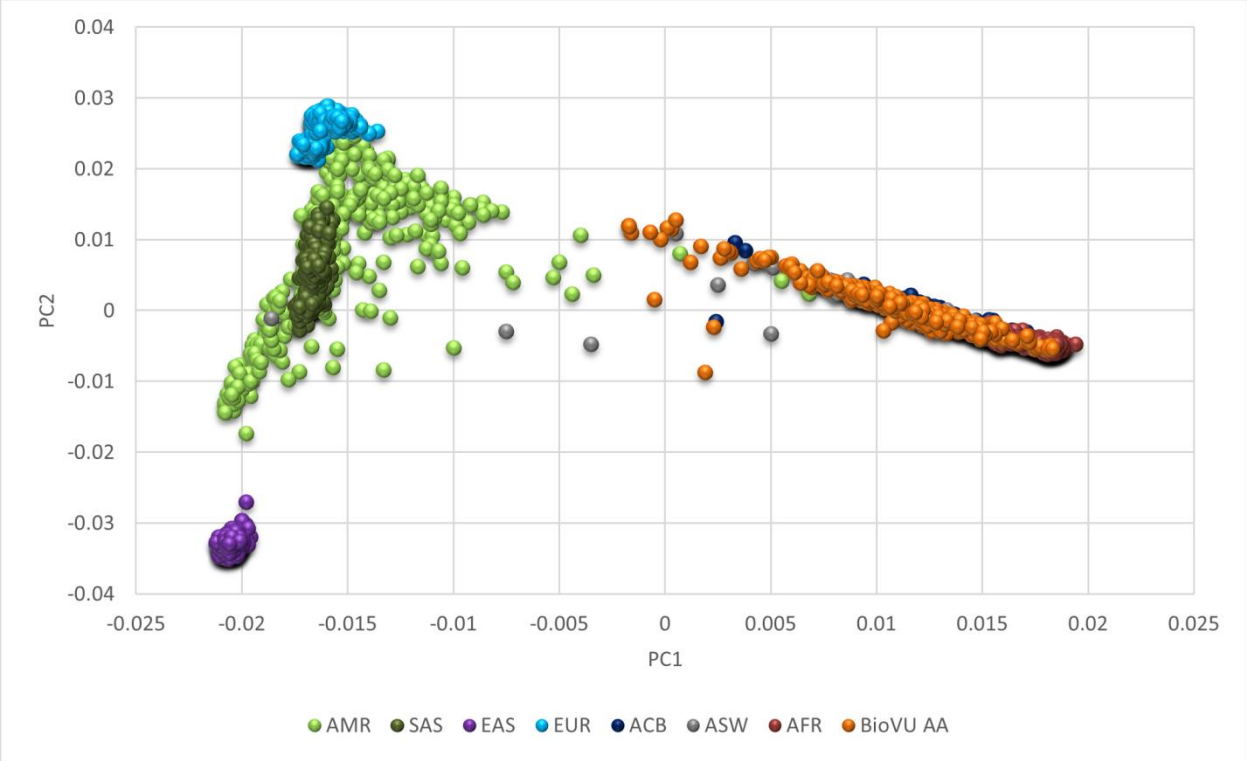


Appendix I. Quality control flowchart of BioVU AA.



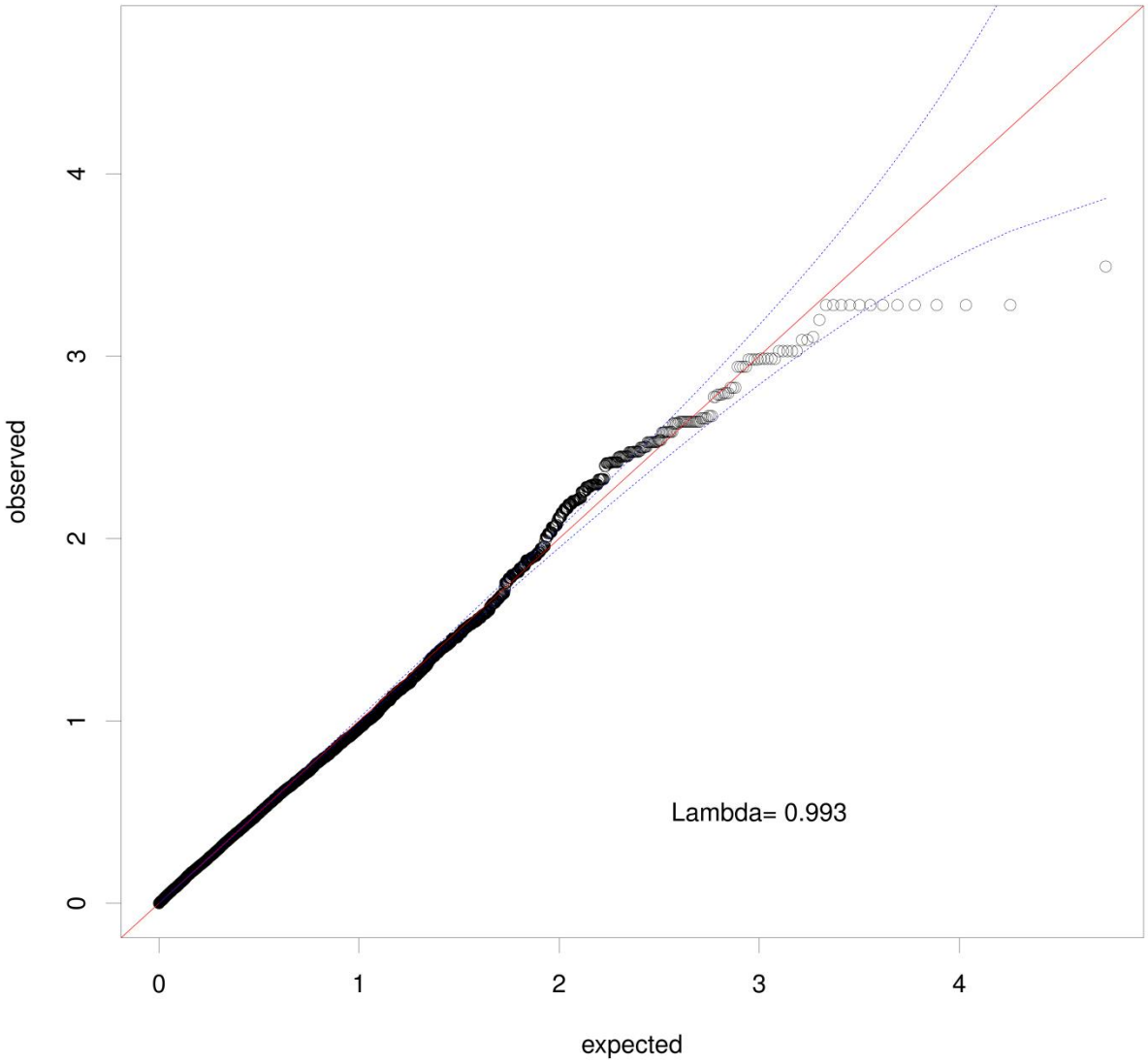
Appendix J. Graph of the principal component analysis of CARDIA AA cases plotted with individuals from the 1,000 Genomes Project.

AMR-Ad Mixed American; EAS-East Asian; EUR-European; SAS-South Asian. ACB-African Caribbeans in Barbados; ASW-Americans of African Ancestry in SW USA; AFR-African.

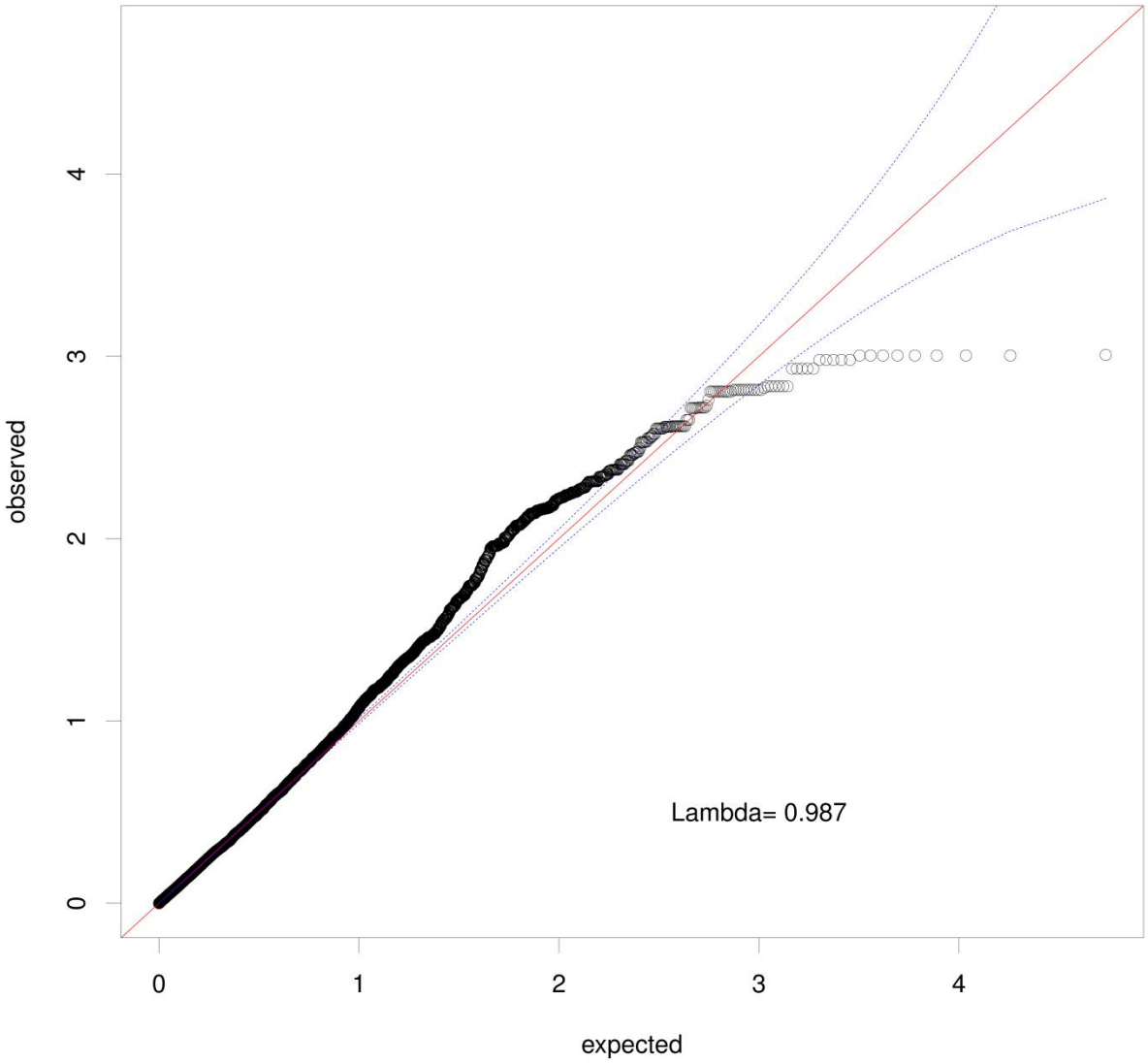


Appendix K. Graph of the principal component analysis of BioVU AA cases plotted with individuals from the 1,000 Genomes Project.

AMR-Ad Mixed American; EAS-East Asian; EUR-European; SAS-South Asian. ACB-African Caribbeans in Barbados; ASW-Americans of African Ancestry in SW USA; AFR-African.

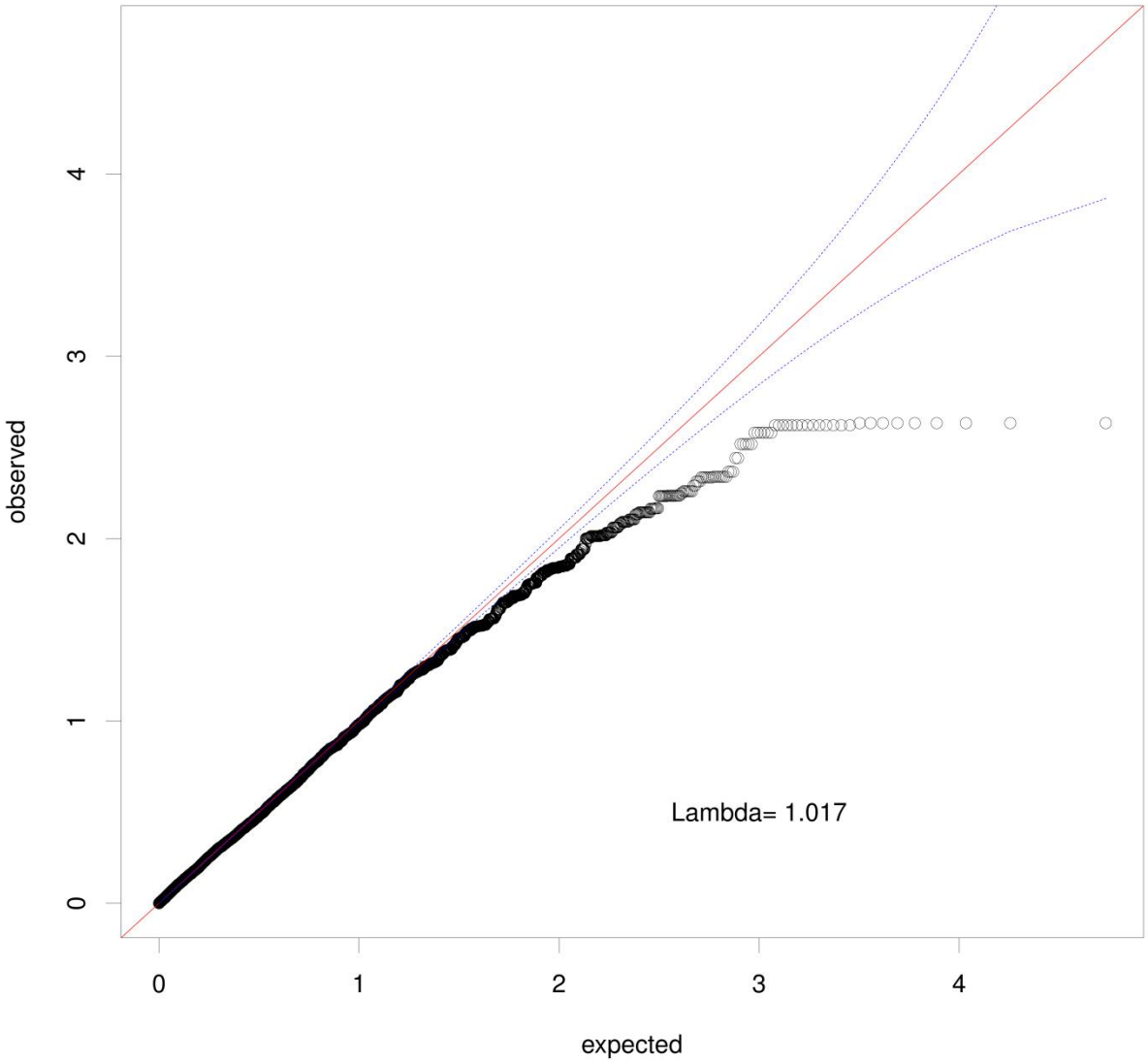


Appendix L. The QQ plot for the meta-analysis of the admixture mapping results for fibroid number (single vs multiple) between BioVU and CARDIA AAs.
The regression analyses were adjusted for age, BMI, five PCs. Markers with a MAF of at least 5% were used in the admixture mapping analysis.



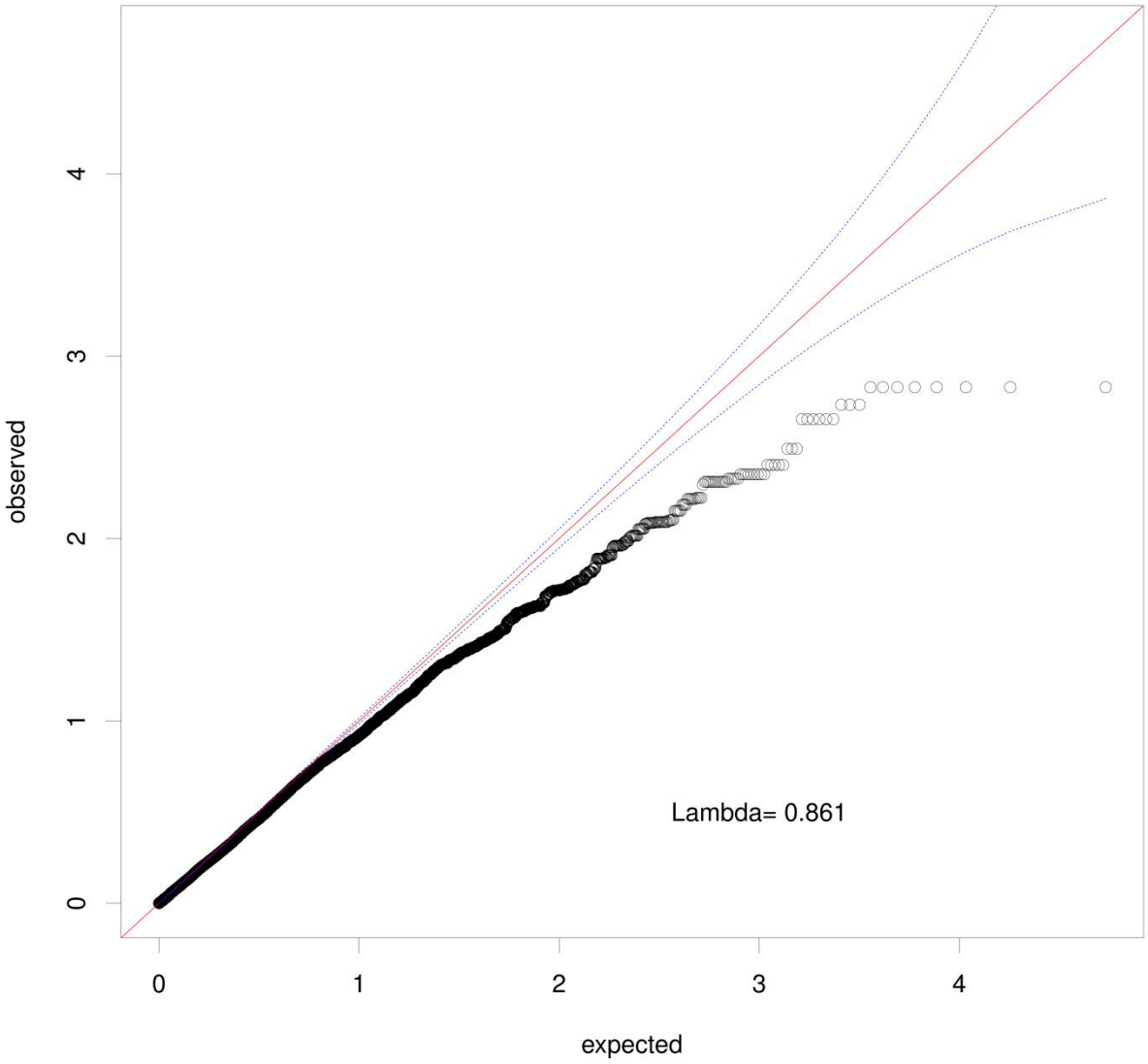
Appendix M. The QQ plot for the admixture mapping results for fibroid number (single vs multiple) for BioVU AAs.

The regression analyses were adjusted for age, BMI, five PCs. Markers with a MAF of at least 5% were used in the admixture mapping analysis.

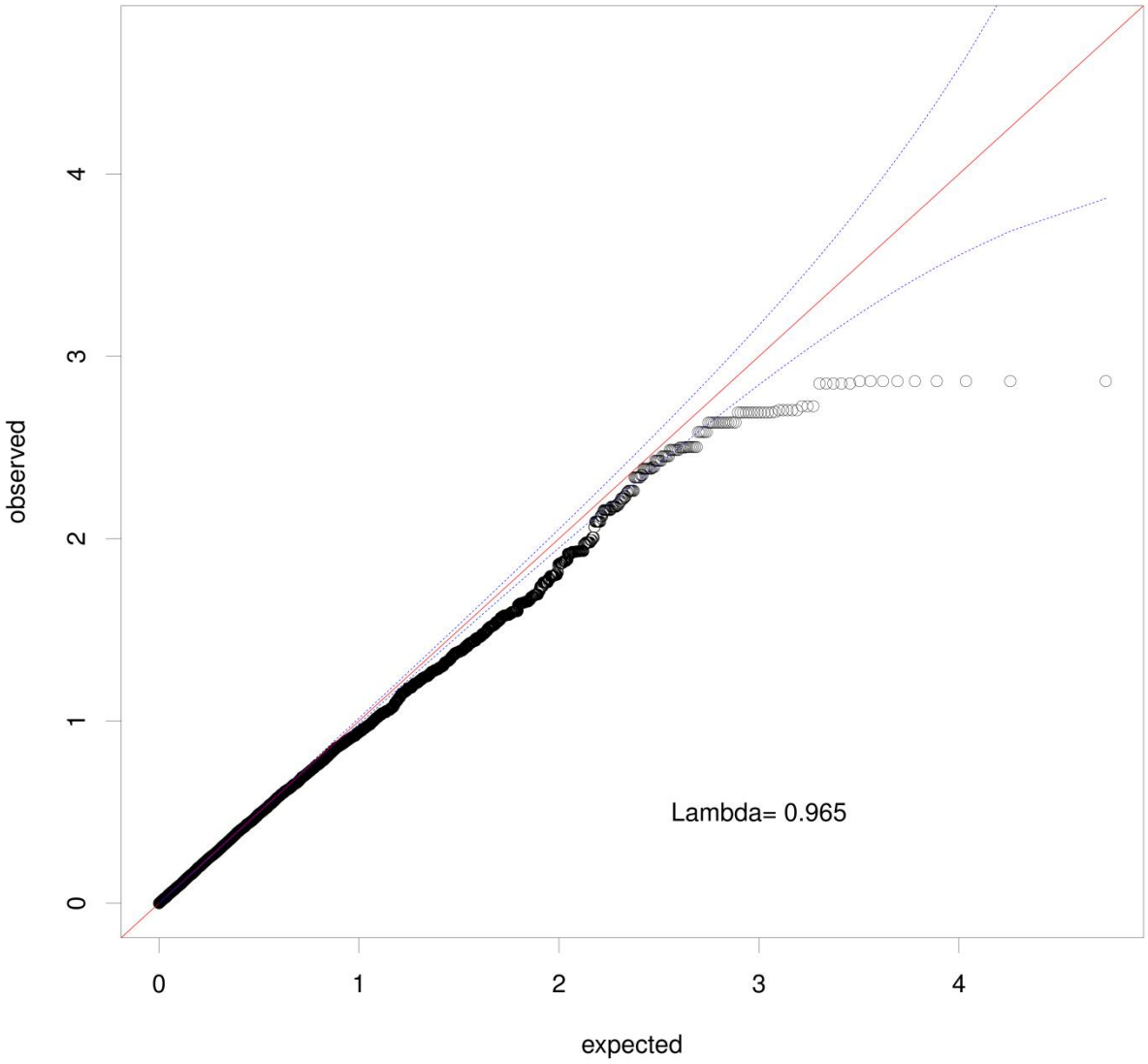


Appendix N. The QQ plot for the admixture mapping results for fibroid number (single vs multiple) for CARDIA AAs.

The regression analyses were adjusted for age, BMI, five PCs. Markers with a MAF of at least 5% were used in the admixture mapping analysis.

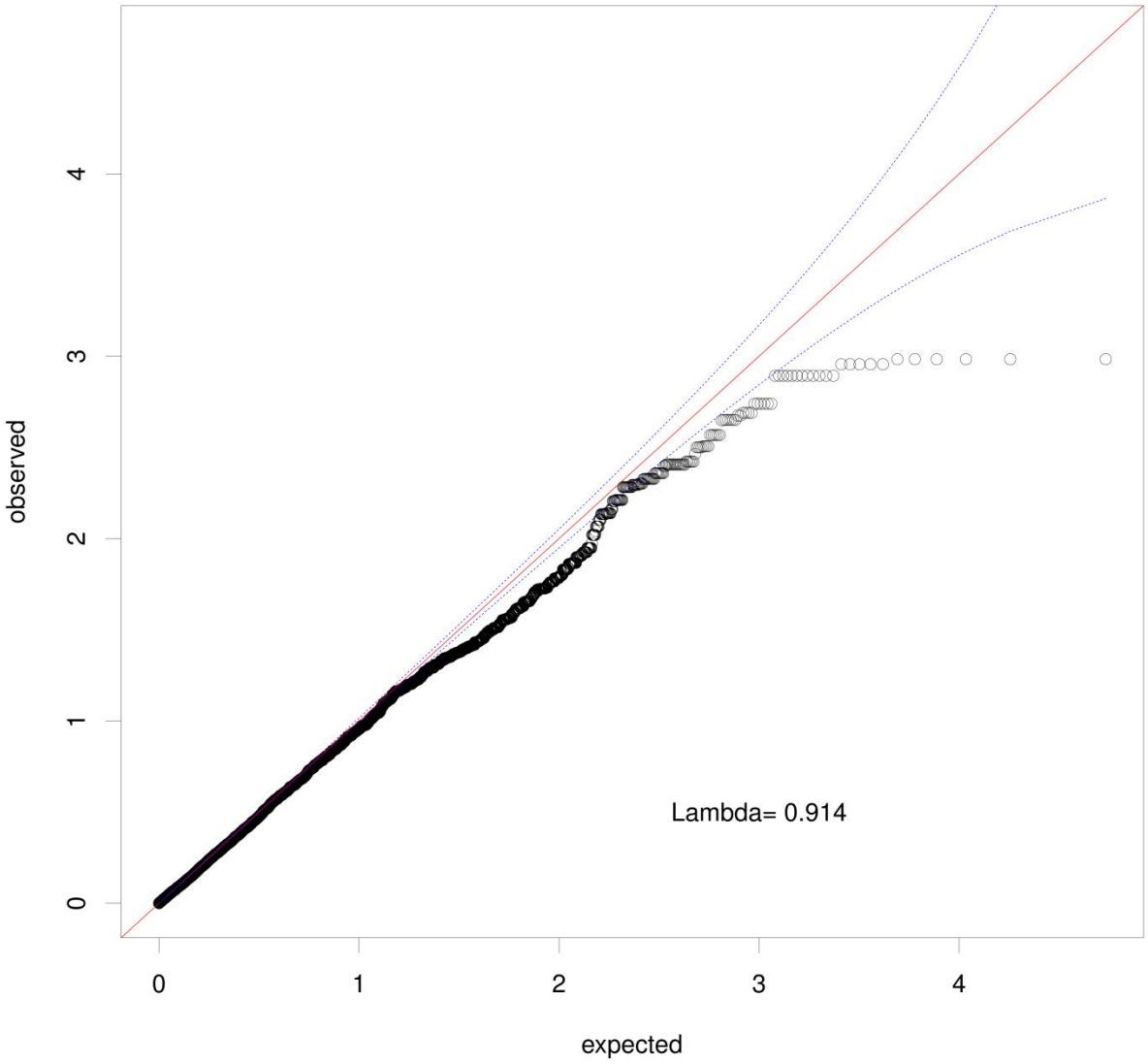


Appendix O. The QQ plot for the meta-analysis of the admixture mapping results for log_{10} transformed volume of largest fibroid between BioVU and CARDIA AAs.
The regression analyses were adjusted for age, BMI, five PCs. Markers with a MAF of at least 5% were used in the admixture mapping analysis.



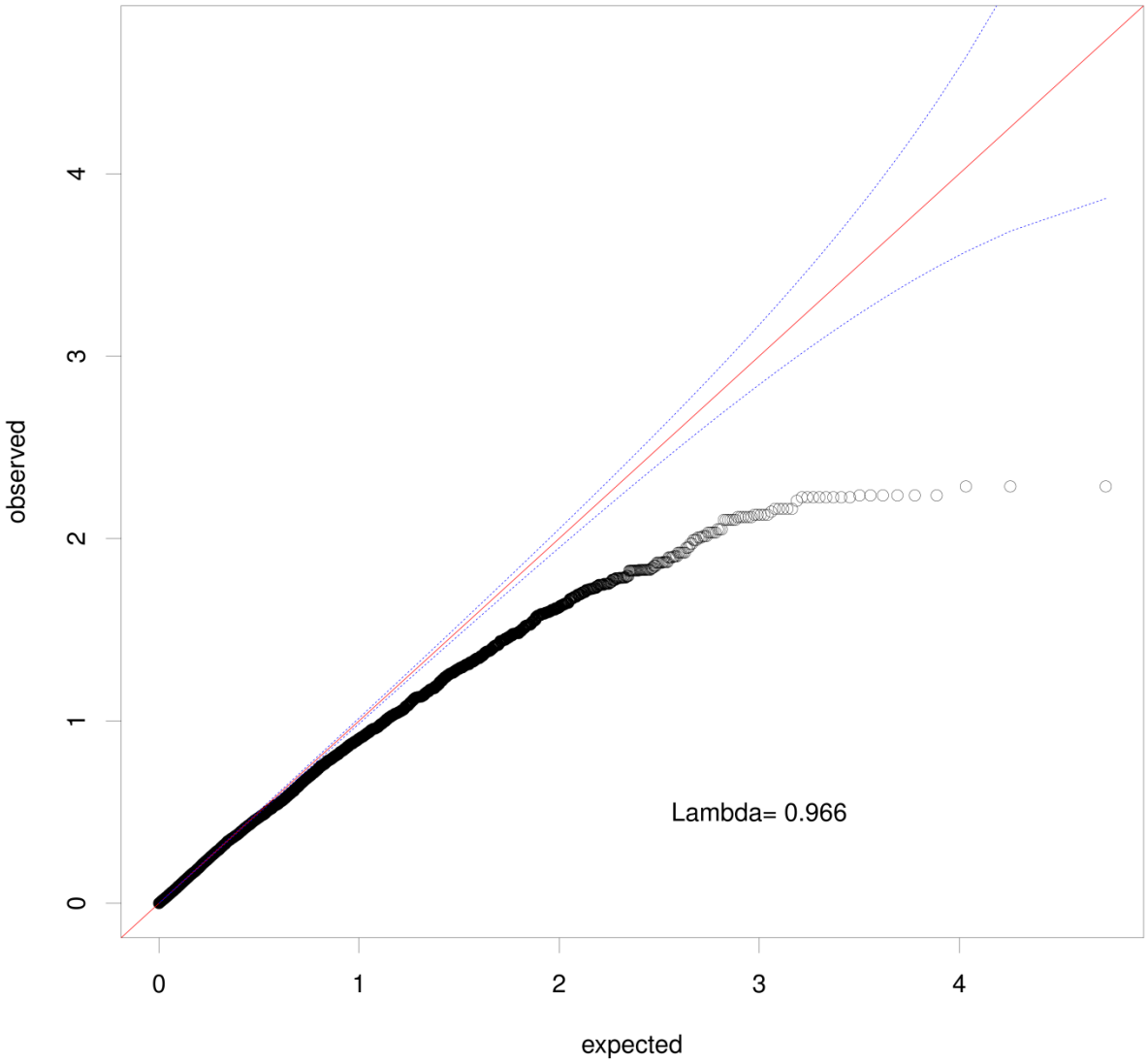
Appendix P. The QQ plot for the admixture mapping results for log_{10} transformed volume of largest fibroid for BioVU AAs.

The regression analyses were adjusted for age, BMI, five PCs. Markers with a MAF of at least 5% were used in the admixture mapping analysis.

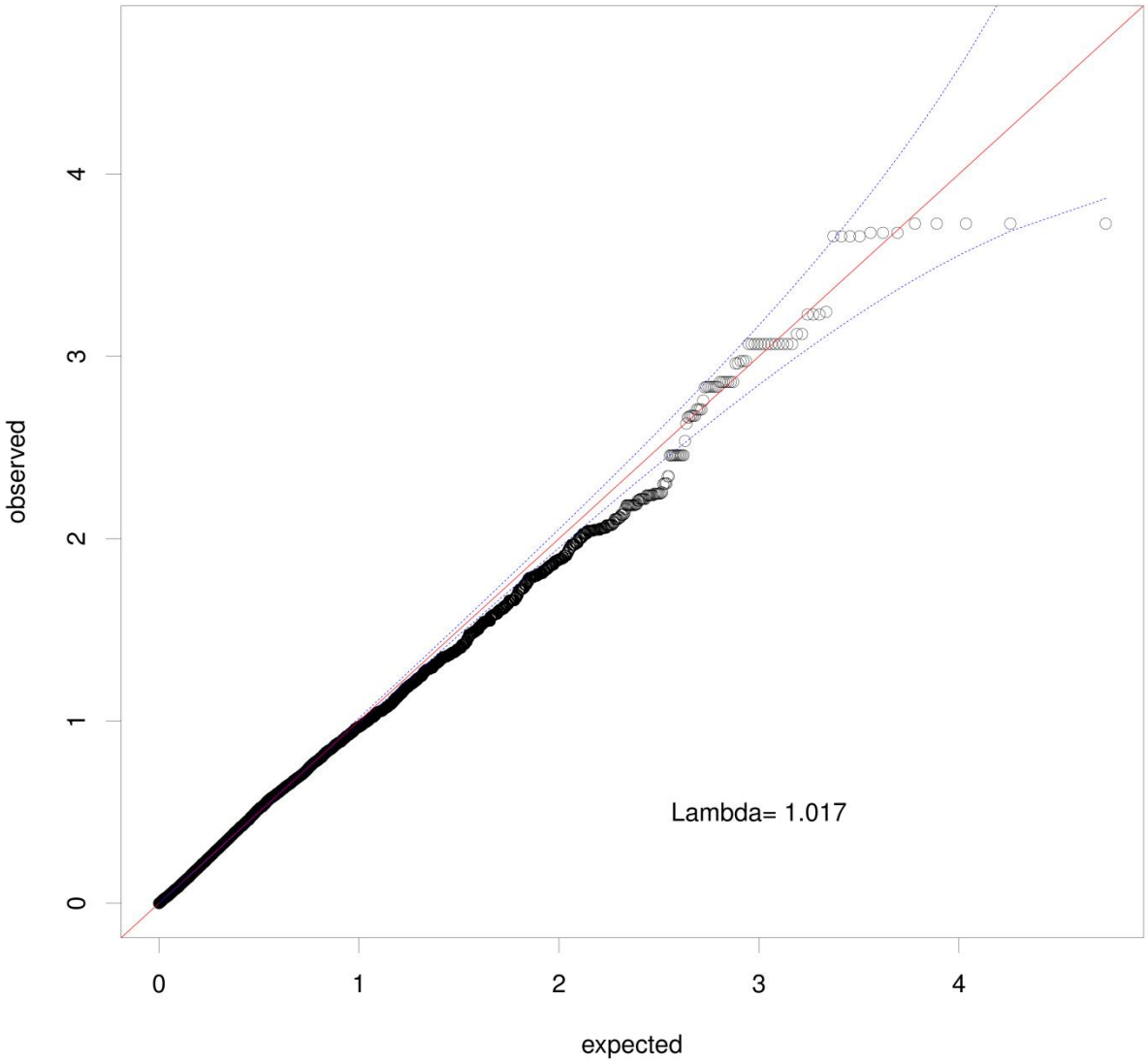


Appendix Q. The QQ plot for the admixture mapping results for log_{10} transformed volume of largest fibroid for CARDIA AAs.

The regression analyses were adjusted for age, BMI, five PCs. Markers with a MAF of at least 5% were used in the admixture mapping analysis.

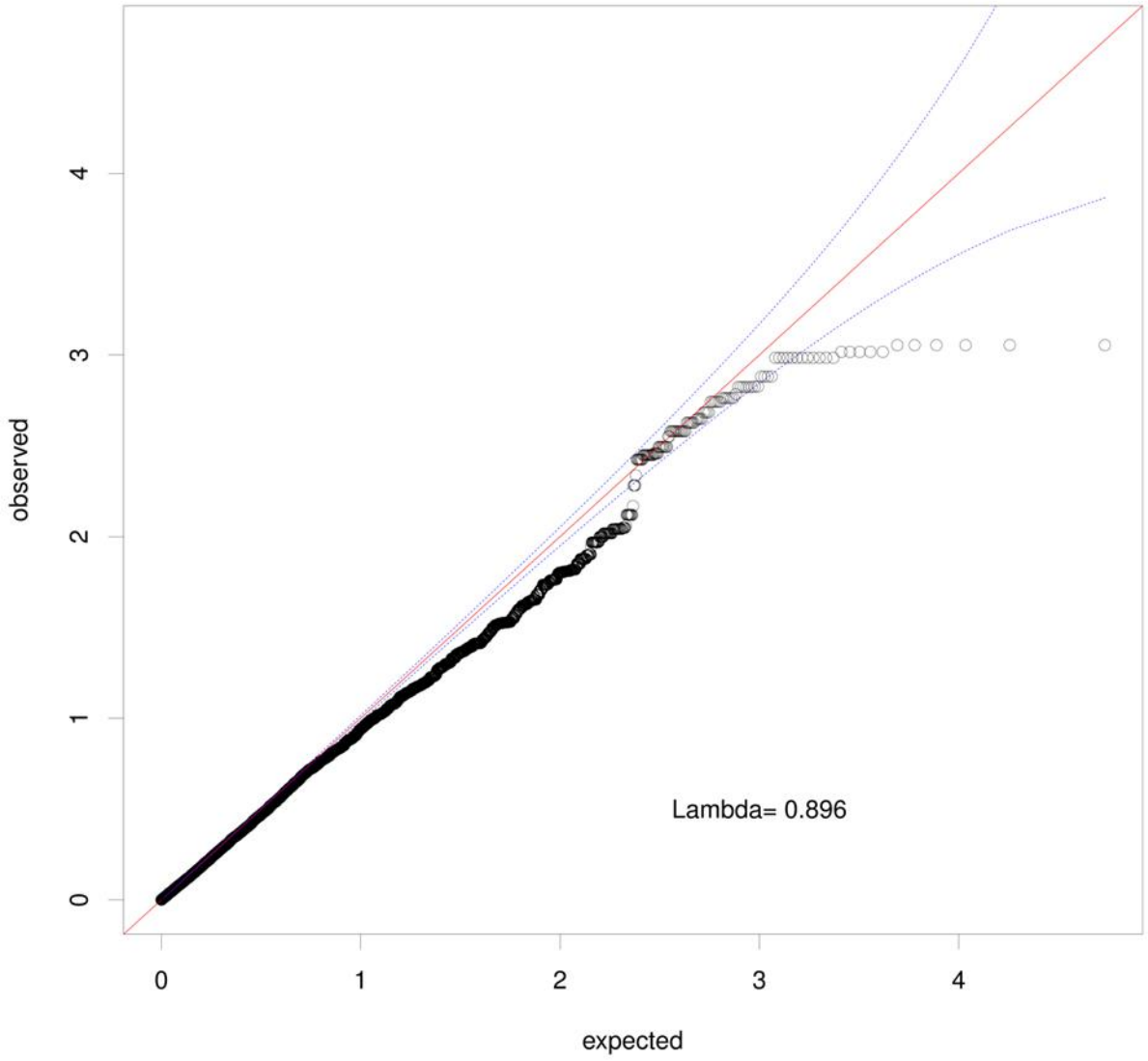


Appendix R. The QQ plot for the meta-analysis of the admixture mapping results for $\log_{10}$ transformed largest dimension of all fibroids between BioVU and CARDIA AAs.
The regression analyses were adjusted for age, BMI, five PCs. Markers with a MAF of at least 5% were used in the admixture mapping analysis.



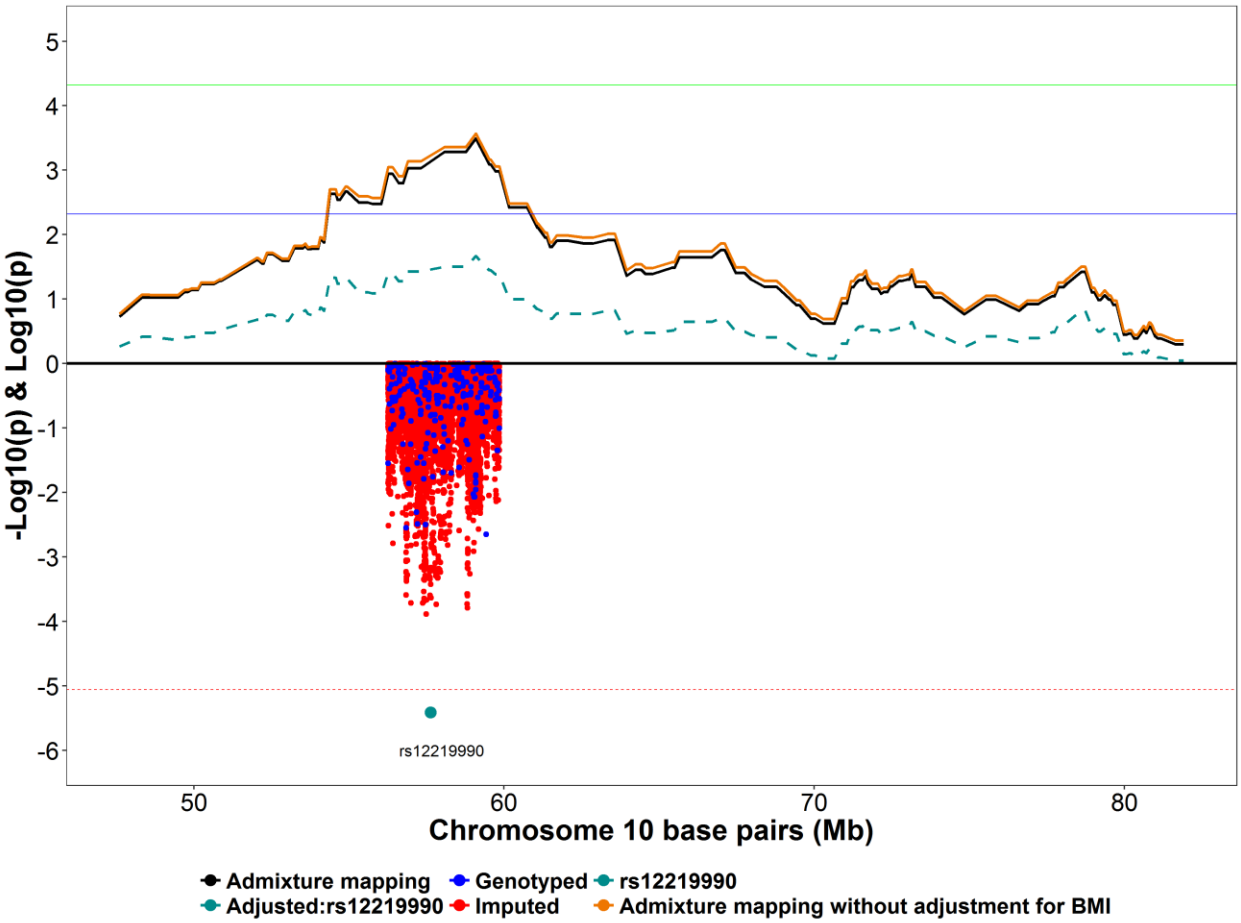
Appendix S. The QQ plot for the admixture mapping results for log_{10} transformed largest dimension of all fibroids for BioVU AAs.

The regression analyses were adjusted for age, BMI, five PCs. Markers with a MAF of at least 5% were used in the admixture mapping analysis.

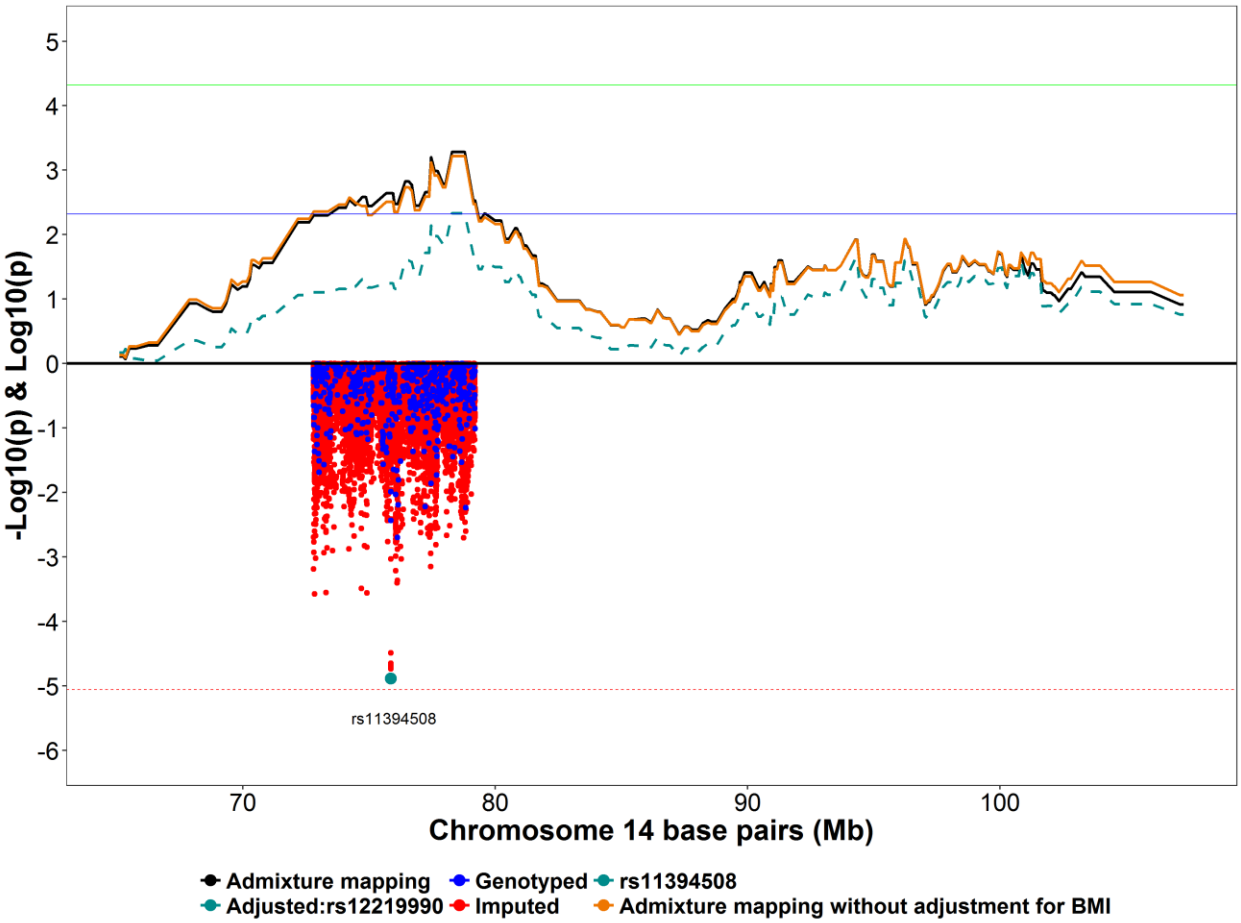


Appendix T. The QQ plot for the admixture mapping results for log_{10} transformed largest dimension of all fibroids for CARDIA AAs.

The regression analyses were adjusted for age, BMI, five PCs. Markers with a MAF of at least 5% were used in the admixture mapping analysis.

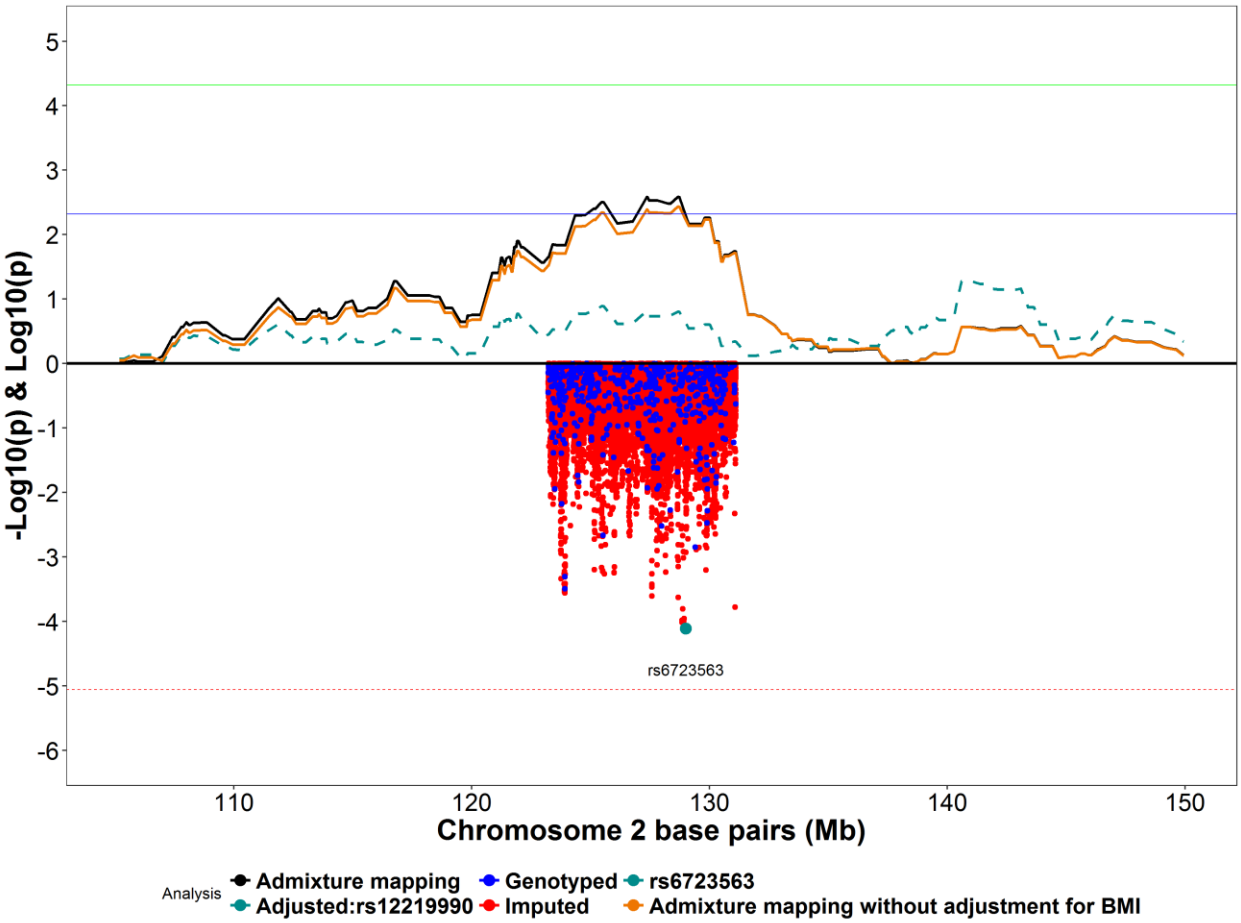


Appendix U. Admixture mapping analysis of chromosome 10 with overlapping single SNP association analysis results with and without adjustment for BMI.
The X-axis indicates the genomic position along chromosome 10 in Mb. The top of the Y-axis indicates -$\log_{10}$(*p* value) from the meta-analysis between BioVU and CARDIA AA from logistic regression of admixture mapping values that were generated from LAMP-ANC (solid black line – not conditioned for rs12219990; dashed green line – conditioned on rs12219990) with fibroid number (single vs multiple) as the outcome. The admixture mapping analysis was adjusted for age, BMI, five PCs. The solid orange line represents the admixture mapping analysis that was adjusted for age and five PCs only. The solid green line represents the significance threshold. The solid blue line represents the suggestive threshold. The bottom portion of the Y-axis indicates the $\log_{10}$(*p* value) for the single SNP association analyses (blue circles indicates genotyped SNPs; red circles indicates imputed SNPs) with fibroid number (single vs multiple) as the outcome. The imputed region was found by taking 1 –log down from the most significant mapping peak above the suggest threshold. The imputed region encompassed 56,258,202 to 59,854,651 bp. The single SNP association analysis was adjusted for age, BMI, and five PCs. The SNP, rs12219990, was imputed (BioVU info score – 0.921; CARDIA info score – 0.922). The dotted red line represents the significance threshold.

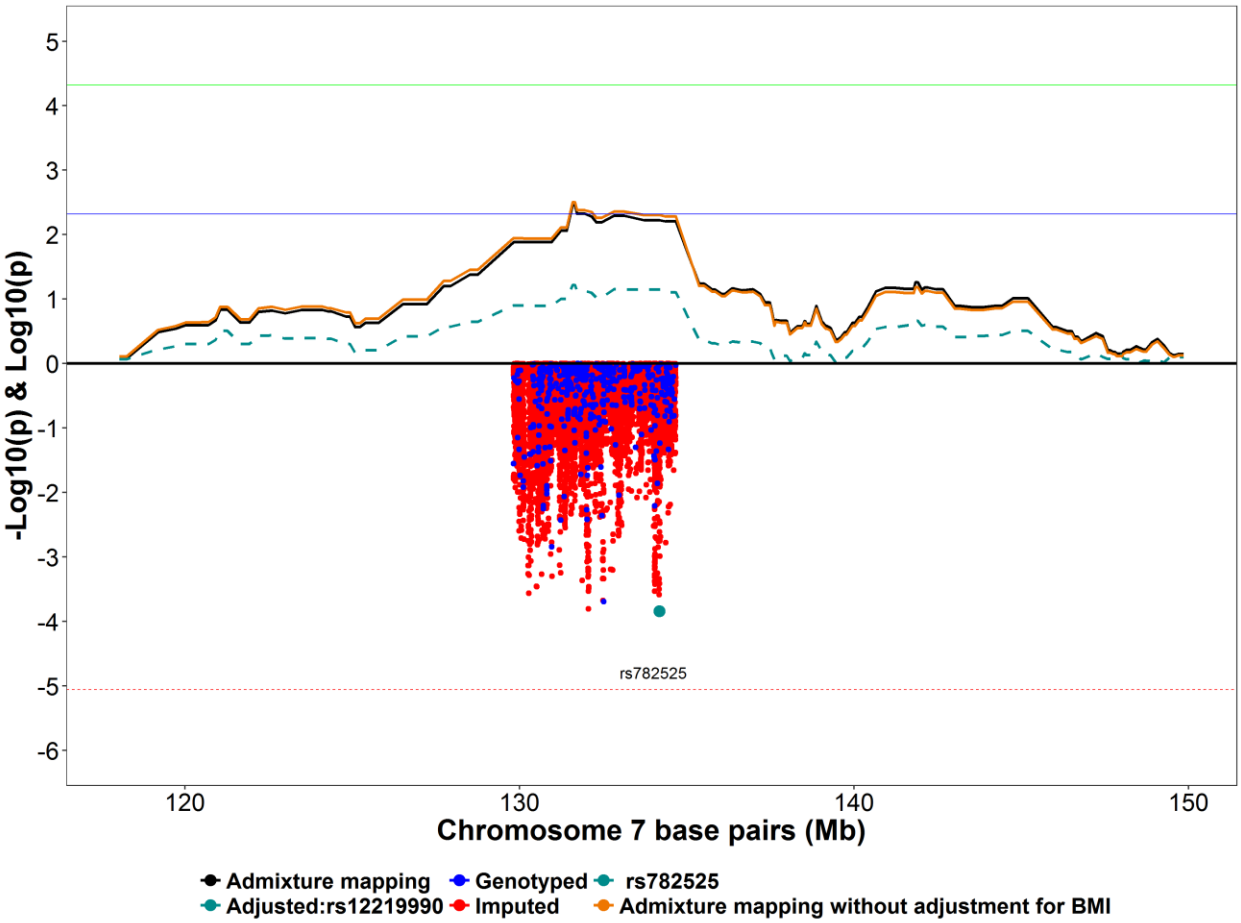


Appendix V. Admixture mapping analysis of chromosome 14 with overlapping single SNP association analysis results with and without adjustment for BMI.

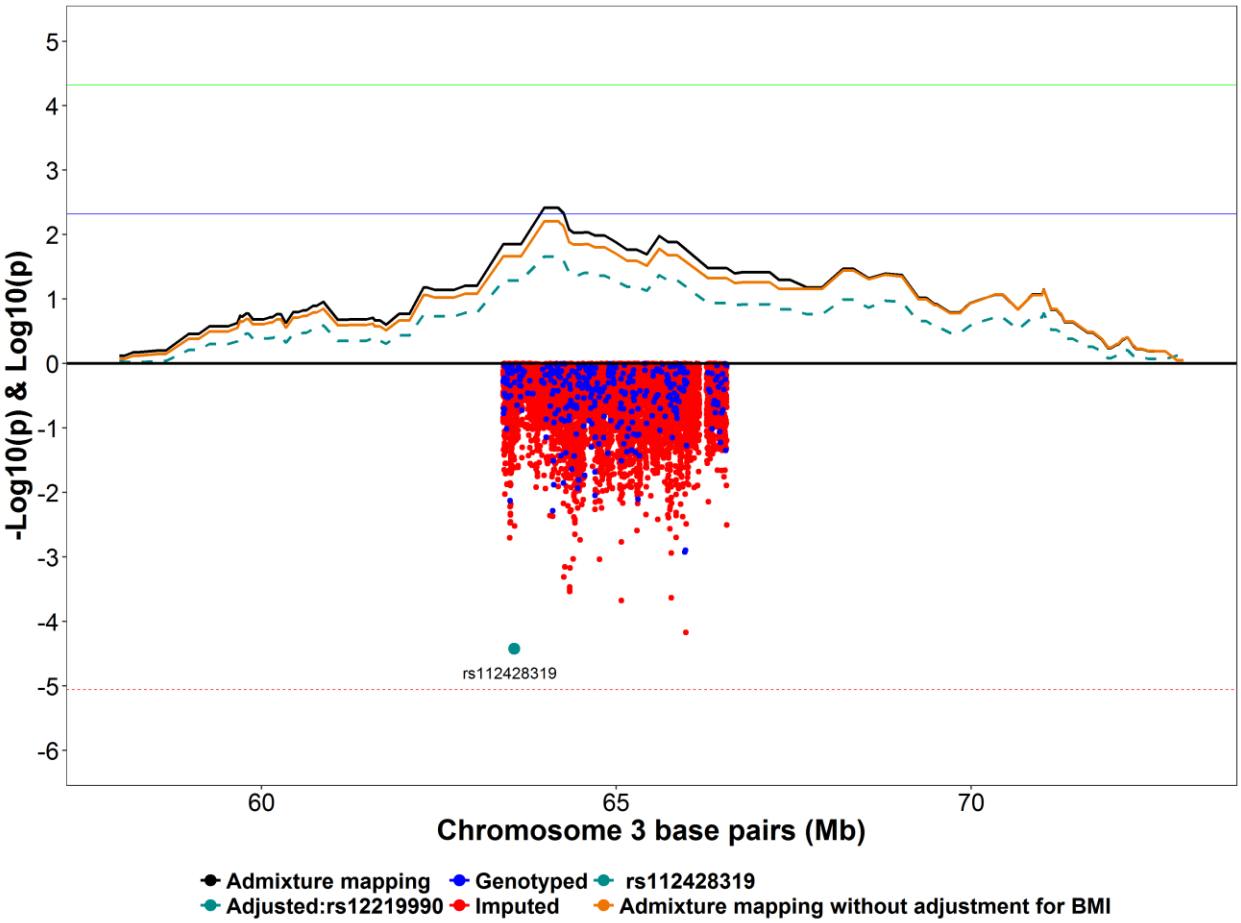
The X-axis indicates the genomic position along chromosome 14 in Mb. The top of the Y-axis indicates -$\log_{10}$(*p* value) from the meta-analysis between BioVU and CARDIA AA from logistic regression of admixture mapping values that were generated from LAMP-ANC (solid black line – not conditioned for rs11394508; dashed green line – conditioned on rs11394508) with fibroid number (single vs multiple) as the outcome. The admixture mapping analysis was adjusted for age, BMI, five PCs. The solid orange line represents the admixture mapping analysis that was adjusted for age and five PCs only. The solid green line represents the significance threshold. The solid blue line represents the suggestive threshold. The bottom portion of the Y-axis indicates the $\log_{10}$(*p* value) for the single SNP association analyses (blue circles indicates genotyped SNPs; red circles indicates imputed SNPs) with fibroid number (single vs multiple) as the outcome. The imputed region was found by taking 1 –log down from the most significant mapping peak above the suggest threshold. The imputed region encompassed 72,802,666 to 79,205,619 bp. The single SNP association analysis was adjusted for age, BMI, and five PCs. The SNP, rs11394508, was imputed (BioVU info score – 0.995; CARDIA info score – 0.956). The dotted red line represents the significance threshold.



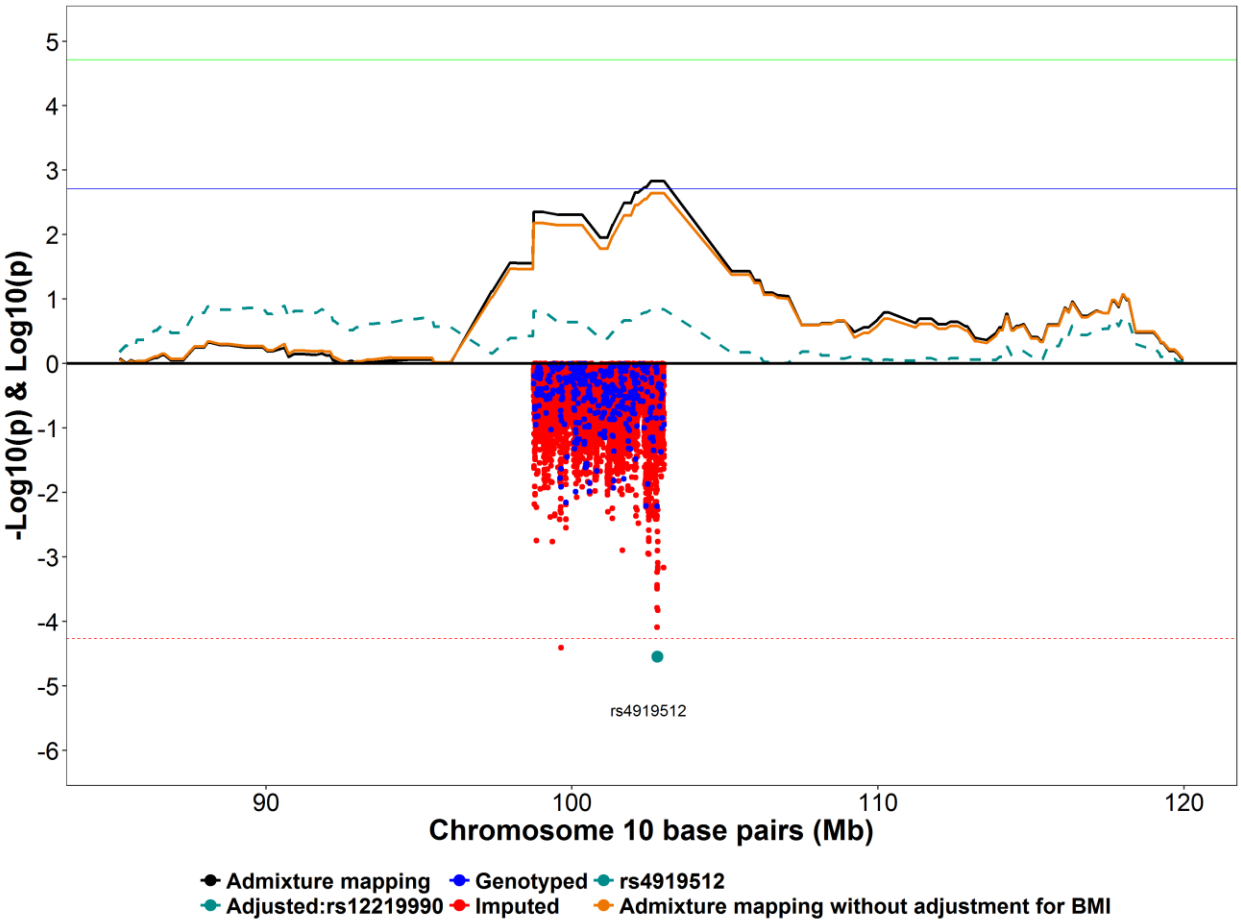
Appendix W. Admixture mapping analysis of chromosome 2 with overlapping single SNP association analysis results with and without adjustment for BMI.
The X-axis indicates the genomic position along chromosome 2 in Mb. The top of the Y-axis indicates -$\log_{10}$(*p* value) from the meta-analysis between BioVU and CARDIA AA from logistic regression of admixture mapping values that were generated from LAMP-ANC (solid black line – not conditioned for rs6723563; dashed green line – conditioned on rs6723563) with fibroid number (single vs multiple) as the outcome. The admixture mapping analysis was adjusted for age, BMI, five PCs. The solid orange line represents the admixture mapping analysis that was adjusted for age and five PCs only. The solid green line represents the significance threshold. The solid blue line represents the suggestive threshold. The bottom portion of the Y-axis indicates the $\log_{10}$(*p* value) for the single SNP association analyses (blue circles indicates genotyped SNPs; red circles indicates imputed SNPs) with fibroid number (single vs multiple) as the outcome. The imputed region was found by taking 1 –log down from the most significant mapping peak above the suggest threshold. The imputed region encompassed 123,221,985 to 131,110,097 bp. The single SNP association analysis was adjusted for age, BMI, and five PCs. The SNP, rs6723563, was imputed (BioVU info score – 0.932; CARDIA info score – 0.885). The dotted red line represents the significance threshold.



Appendix X. Admixture mapping analysis of chromosome 7 with overlapping single SNP association analysis results with and without adjustment for BMI.
The X-axis indicates the genomic position along chromosome 7 in Mb. The top of the Y-axis indicates -$\log_{10}$(*p* value) from the meta-analysis between BioVU and CARDIA AA from logistic regression of admixture mapping values that were generated from LAMP-ANC (solid black line – not conditioned for rs782525; dashed green line – conditioned on rs7872525) with fibroid number (single vs multiple) as the outcome. The admixture mapping analysis was adjusted for age, BMI, five PCs. The solid orange line represents the admixture mapping analysis that was adjusted for age and five PCs only. The solid green line represents the significance threshold. The solid blue line represents the suggestive threshold. The bottom portion of the Y-axis indicates the $\log_{10}$(*p* value) for the single SNP association analyses (blue circles indicates genotyped SNPs; red circles indicates imputed SNPs) with fibroid number (single vs multiple) as the outcome. The imputed region was found by taking 1 –log down from the most significant mapping peak above the suggest threshold. The imputed region encompassed 129,822,797 to 134,670,462 bp. The single SNP association analysis was adjusted for age, BMI, and five PCs. The SNP, rs782525, was imputed (BioVU info score – 0.999; CARDIA info score – 0.985). The dotted red line represents the significance threshold.

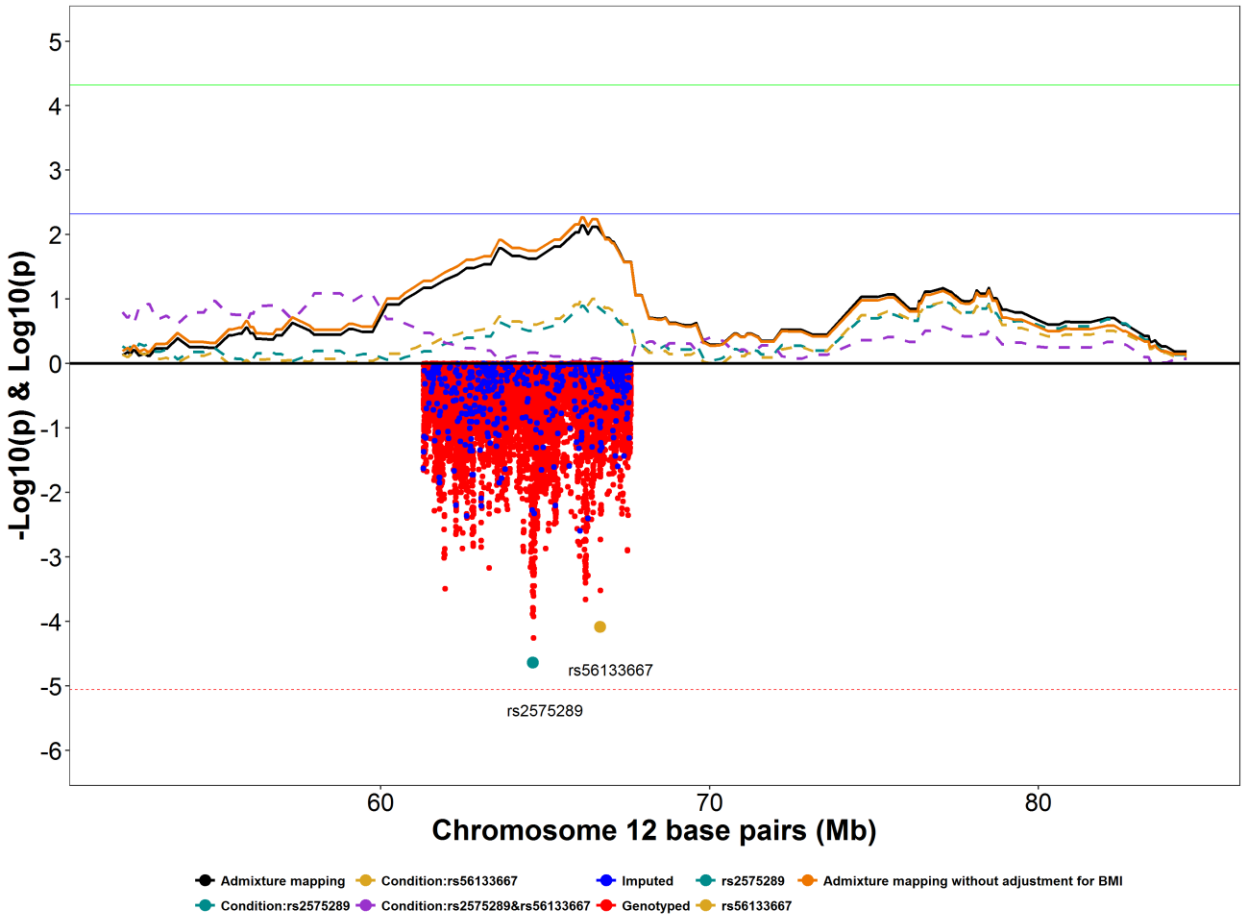


Appendix Y. Admixture mapping analysis of chromosome 3 with overlapping single SNP association analysis results with and without adjustment for BMI.
The X-axis indicates the genomic position along chromosome 3 in Mb. The top of the Y-axis indicates -$\log_{10}$(*p* value) from the meta-analysis between BioVU and CARDIA AA from logistic regression of admixture mapping values that were generated from LAMP-ANC (solid black line – not conditioned for rs112428319; dashed green line – conditioned on rs112428319) with fibroid number (single vs multiple) as the outcome. The admixture mapping analysis was adjusted for age, BMI, five PCs. The solid orange line represents the admixture mapping analysis that was adjusted for age and five PCs only. The solid green line represents the significance threshold. The solid blue line represents the suggestive threshold. The bottom portion of the Y-axis indicates the $\log_{10}$(*p* value) for the single SNP association analyses (blue circles indicates genotyped SNPs; red circles indicates imputed SNPs) with fibroid number (single vs multiple) as the outcome. The imputed region was found by taking 1 –log down from the most significant mapping peak above the suggest threshold. The imputed region encompassed 63,405,151 to 66,558,036 bp. The single SNP association analysis was adjusted for age, BMI, and five PCs. The SNP, rs112428319, was imputed (BioVU info score – 0.9998; CARDIA info score – 0.994). The dotted red line represents the significance threshold.



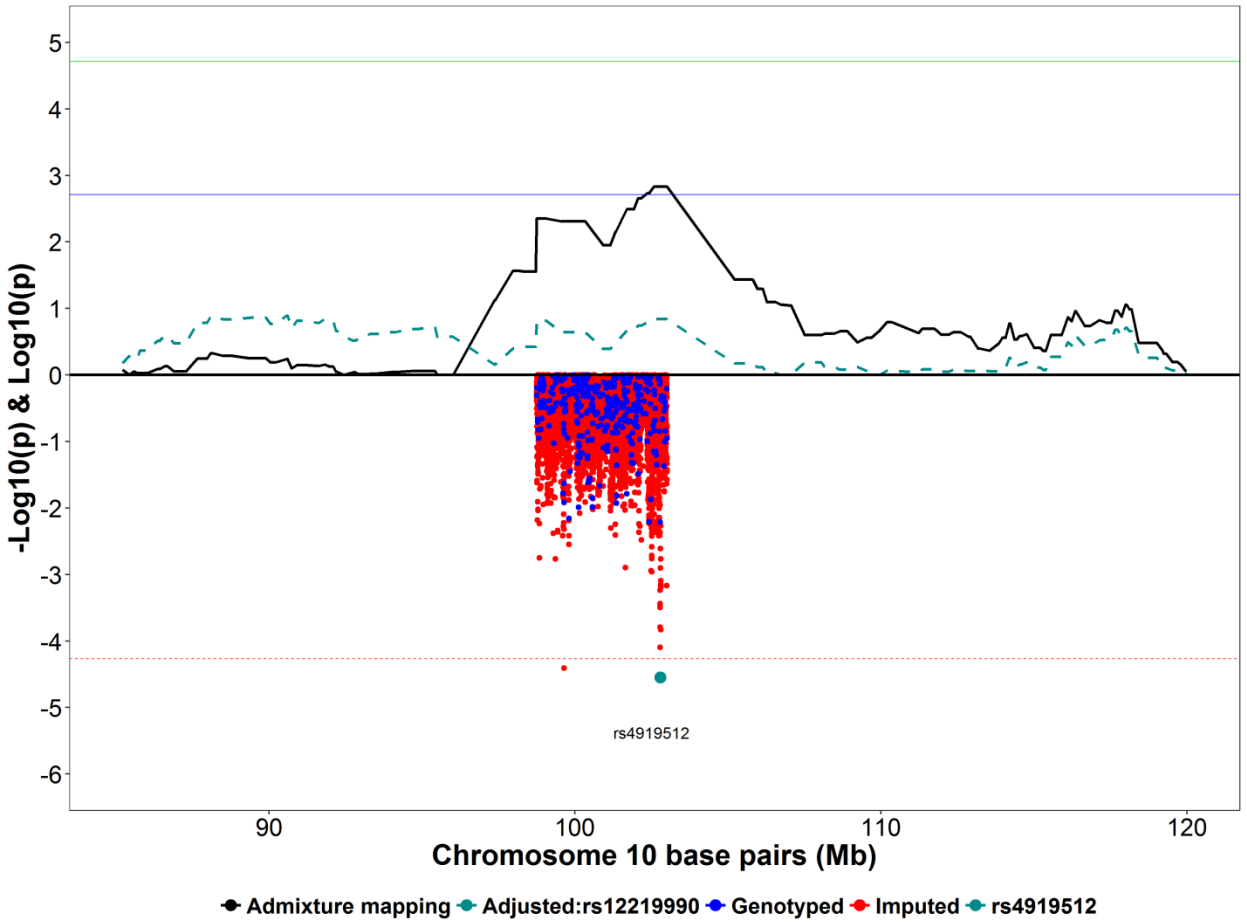
Appendix Z. Admixture mapping analysis of chromosome 10 with overlapping single SNP association analysis results with and without adjustment for BMI.

The X-axis indicates the genomic position along chromosome 10 in Mb. The top of the Y-axis indicates -$\log_{10}$(*p* value) from the meta-analysis between BioVU and CARDIA AA from logistic regression of admixture mapping values that were generated from LAMP-ANC (solid black line – not conditioned for rs4919512; dashed green line – conditioned on rs4919512) with fibroid volume ($\log_{10}$ transformed) as the outcome. The admixture mapping analysis was adjusted for age, BMI, five PCs. The solid orange line represents the admixture mapping analysis that was adjusted for age and five PCs only. The solid green line represents the significance threshold. The solid blue line represents the suggestive threshold. The bottom portion of the Y-axis indicates the $\log_{10}$(*p* value) for the single SNP association analyses (blue circles indicates genotyped SNPs; red circles indicates imputed SNPs) with fibroid volume ($\log_{10}$ transformed) as the outcome. The imputed region was found by taking 1 –log down from the most significant mapping peak above the suggest threshold. The imputed region encompassed 98,747,114 to 103,009,908 bp. The single SNP association analysis was adjusted for age, BMI, and five PCs. The SNP, rs4919512, was genotyped for BioVU but imputed for CARDIA (BioVU info score – 1; CARDIA info score – 0.801). The dotted red line represents the significance threshold.



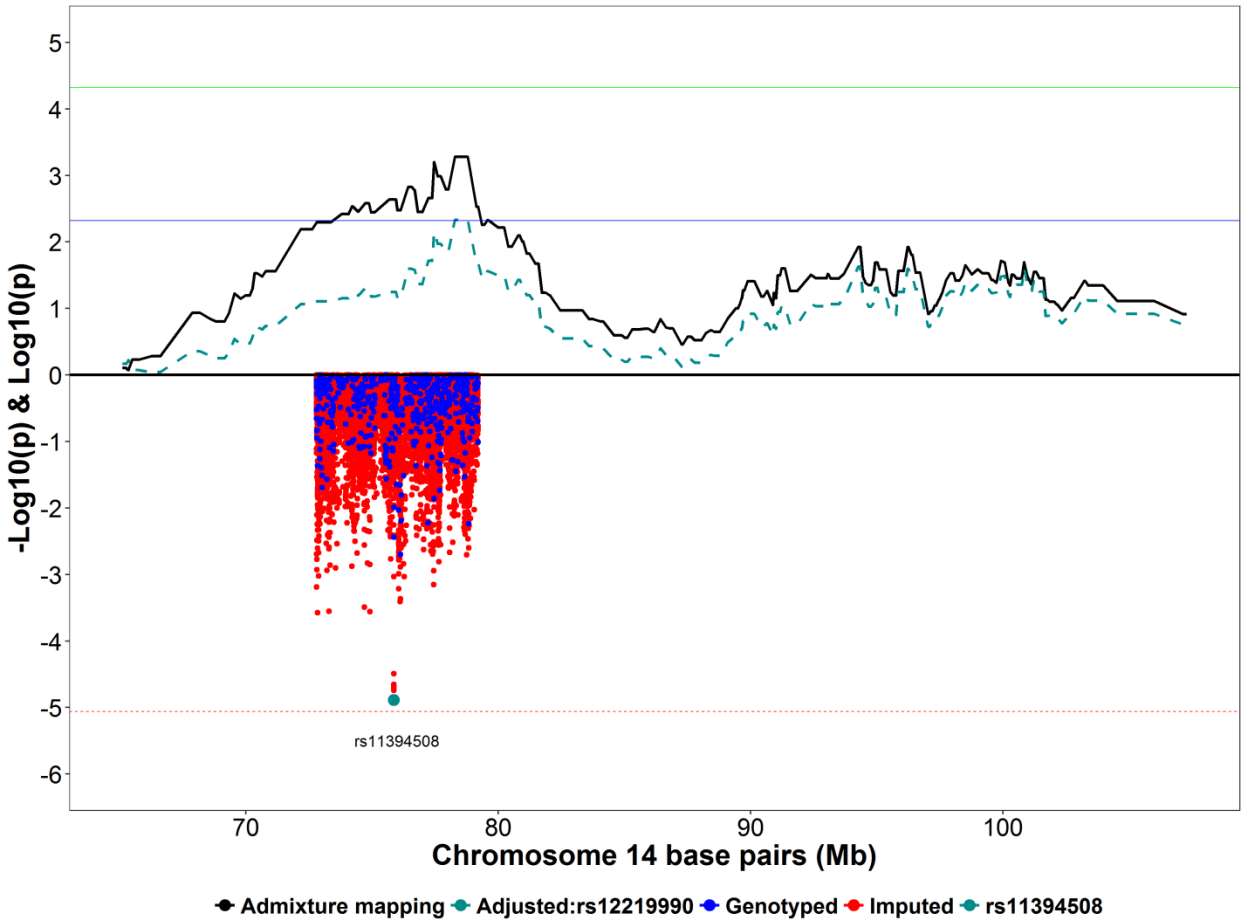
Appendix AA. Admixture mapping analysis of chromosome 12 with overlapping single SNP association analysis results with and without adjustment for BMI.

The X-axis indicates the genomic position along chromosome 10 in Mb. The top of the Y-axis indicates -$\log_{10}$(*p* value) from the meta-analysis between BioVU and CARDIA AA from logistic regression of admixture mapping values that were generated from LAMP-ANC (solid black line – not conditioned for rs2575289 or rs56133667; dashed teal line – conditioned on rs2575289; dashed yellow line - conditioned on rs56133667; dashed purple line - conditioned on rs2575289 and rs56133667) with fibroid number (single vs multiple) as the outcome. The admixture mapping analysis was adjusted for age, BMI, five PCs. The solid orange line represents the admixture mapping analysis that was adjusted for age and five PCs only. The solid green line represents the significance threshold. The solid blue line represents the suggestive threshold. The bottom portion of the Y-axis indicates the $\log_{10}$(*p* value) for the single SNP association analyses (blue circles indicates genotyped SNPs; red circles indicates imputed SNPs) with fibroid number (single vs multiple) as the outcome. The imputed region was found by taking 1 –log down from the most significant mapping peak above the suggest threshold. The imputed region encompassed 61,300,873 to 67,610,913 bp. The single SNP association analysis was adjusted for age, BMI, and five PCs. The SNP, rs2575289, was imputed (BioVU info score – 0.976; CARDIA info score – 0.991). The SNP, rs56133667, was imputed (BioVU info score > 0.999; CARDIA info score > 0.999). The dotted red line represents the significance threshold of the single SNP association analysis.



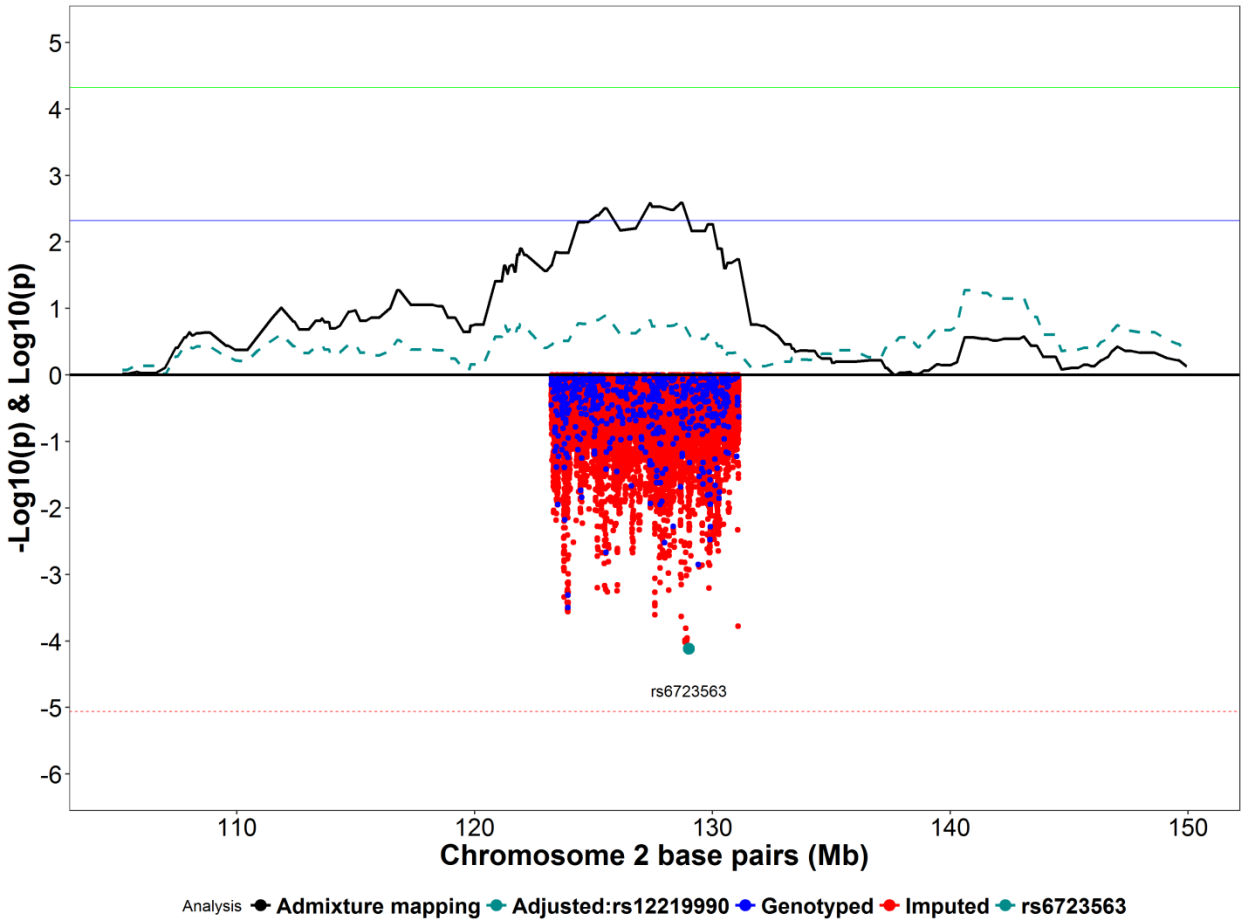
Appendix BB. Admixture mapping analysis of chromosome 10 with overlapping single SNP association analysis results.

The X-axis indicates the genomic position along chromosome 10 in Mb. The top of the Y-axis indicates -$\log_{10}$(*p* value) from the meta-analysis between BioVU and CARDIA AA from logistic regression of admixture mapping values that were generated from LAMP-ANC (solid black line – not conditioned for rs4919512; dashed green line – conditioned on rs4919512) with fibroid volume ($\log_{10}$ transformed) as the outcome. The admixture mapping analysis was adjusted for age, BMI, five PCs. The solid green line represents the significance threshold. The solid blue line represents the suggestive threshold. The bottom portion of the Y-axis indicates the $\log_{10}$(*p* value) for the single SNP association analyses (blue circles indicates genotyped SNPs; red circles indicates imputed SNPs) with fibroid volume ($\log_{10}$ transformed) as the outcome. The imputed region was found by taking 1 –log down from the most significant mapping peak above the suggest threshold. The imputed region encompassed 98,747,114 to 103,009,908 bp. The single SNP association analysis was adjusted for age, BMI, and five PCs. The SNP, rs4919512, was genotyped for BioVU but imputed for CARDIA (BioVU info score – 1; CARDIA info score – 0.801). The dotted red line represents the significance threshold.



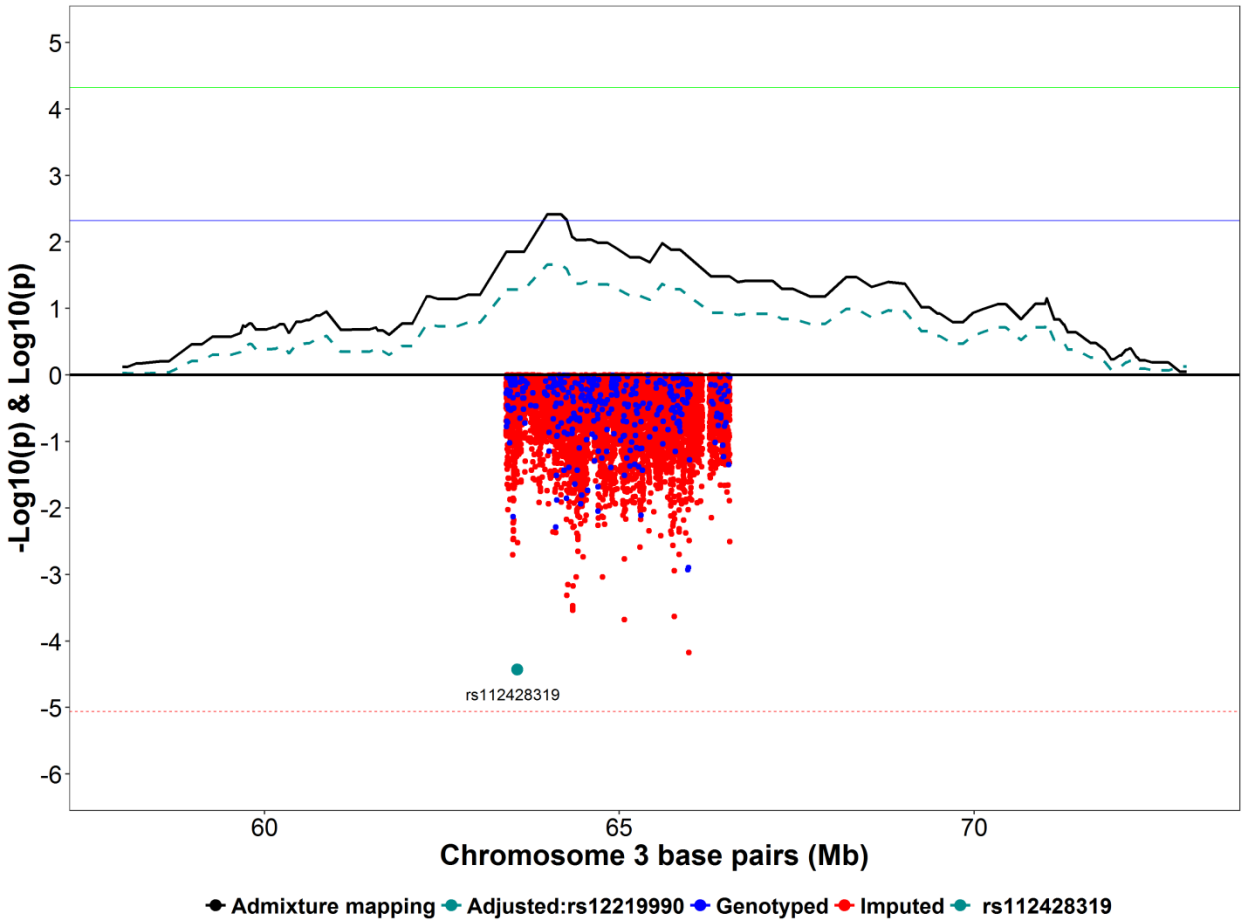
Appendix CC. Admixture mapping analysis of chromosome 14 with overlapping single SNP association analysis results.

The X-axis indicates the genomic position along chromosome 14 in Mb. The top of the Y-axis indicates -$\log_{10}$(*p* value) from the meta-analysis between BioVU and CARDIA AA from logistic regression of admixture mapping values that were generated from LAMP-ANC (solid black line – not conditioned for rs11394508; dashed green line – conditioned on rs11394508) with fibroid number (single vs multiple) as the outcome. The admixture mapping analysis was adjusted for age, BMI, five PCs. The solid green line represents the significance threshold. The solid blue line represents the suggestive threshold. The bottom portion of the Y-axis indicates the $\log_{10}$(*p* value) for the single SNP association analyses (blue circles indicates genotyped SNPs; red circles indicates imputed SNPs) with fibroid number (single vs multiple) as the outcome. The imputed region was found by taking 1 –log down from the most significant mapping peak above the suggest threshold. The imputed region encompassed 72,802,666 to 79,205,619 bp. The single SNP association analysis was adjusted for age, BMI, and five PCs. The SNP, rs11394508, was imputed (BioVU info score – 0.995; CARDIA info score – 0.956). The dotted red line represents the significance threshold.



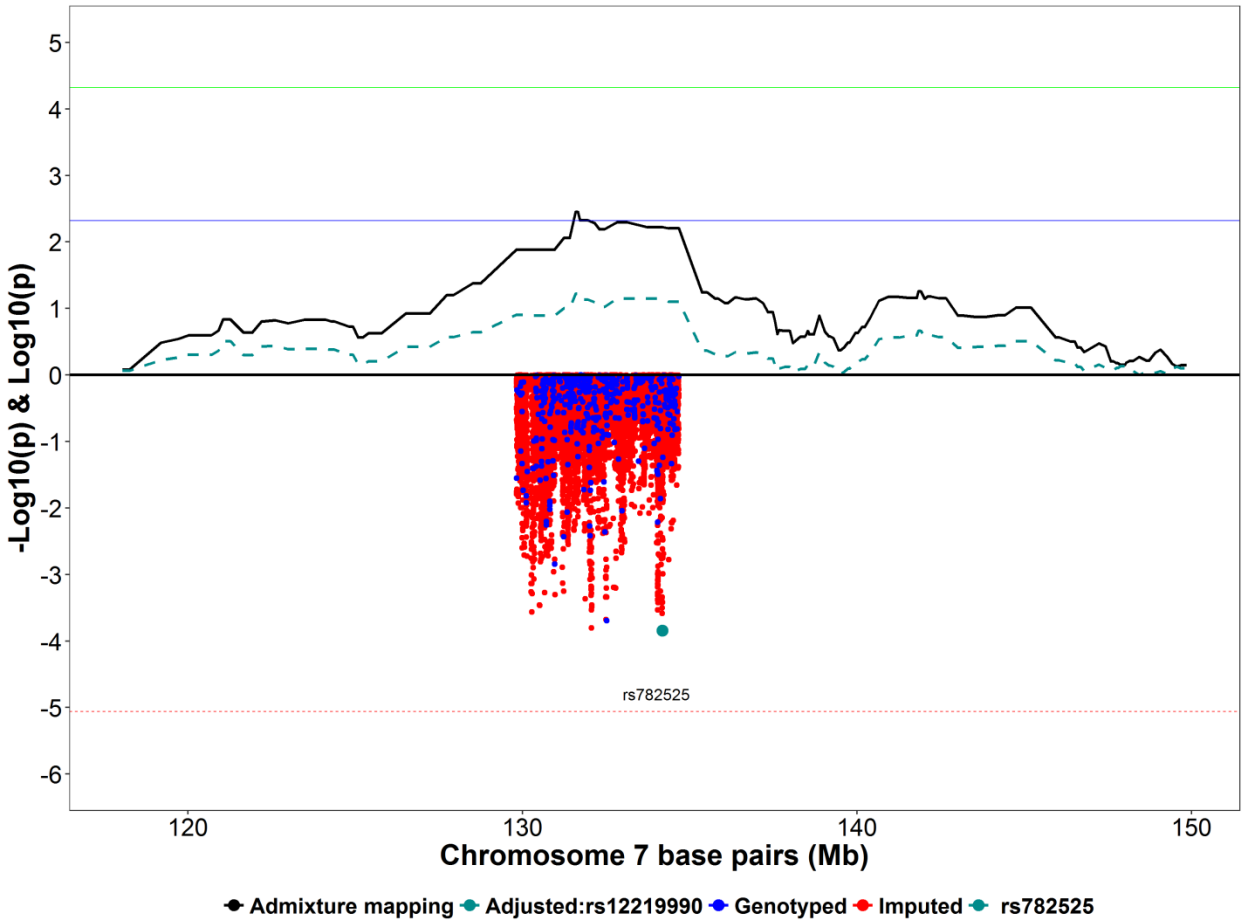
Appendix DD. Admixture mapping analysis of chromosome 2 with overlapping single SNP association analysis results.

The X-axis indicates the genomic position along chromosome 2 in Mb. The top of the Y-axis indicates -$\log_{10}$(*p* value) from the meta-analysis between BioVU and CARDIA AA from logistic regression of admixture mapping values that were generated from LAMP-ANC (solid black line – not conditioned for rs6723563; dashed green line – conditioned on rs6723563) with fibroid number (single vs multiple) as the outcome. The admixture mapping analysis was adjusted for age, BMI, five PCs. The solid green line represents the significance threshold. The solid blue line represents the suggestive threshold. The bottom portion of the Y-axis indicates the $\log_{10}$(*p* value) for the single SNP association analyses (blue circles indicates genotyped SNPs; red circles indicates imputed SNPs) with fibroid number (single vs multiple) as the outcome. The imputed region was found by taking 1 –log down from the most significant mapping peak above the suggest threshold. The imputed region encompassed 123,221,985 to 131,110,097 bp. The single SNP association analysis was adjusted for age, BMI, and five PCs. The SNP, rs6723563, was imputed (BioVU info score – 0.932; CARDIA info score – 0.885). The dotted red line represents the significance threshold.



Appendix EE. Admixture mapping analysis of chromosome 3 with overlapping single SNP association analysis results.

The X-axis indicates the genomic position along chromosome 3 in Mb. The top of the Y-axis indicates -$\log_{10}$(*p* value) from the meta-analysis between BioVU and CARDIA AA from logistic regression of admixture mapping values that were generated from LAMP-ANC (solid black line – not conditioned for rs112428319; dashed green line – conditioned on rs112428319) with fibroid number (single vs multiple) as the outcome. The admixture mapping analysis was adjusted for age, BMI, five PCs. The solid green line represents the significance threshold. The solid blue line represents the suggestive threshold. The bottom portion of the Y-axis indicates the $\log_{10}$(*p* value) for the single SNP association analyses (blue circles indicates genotyped SNPs; red circles indicates imputed SNPs) with fibroid number (single vs multiple) as the outcome. The imputed region was found by taking 1 –log down from the most significant mapping peak above the suggest threshold. The imputed region encompassed 63,405,151 to 66,558,036 bp. The single SNP association analysis was adjusted for age, BMI, and five PCs. The SNP, rs112428319, was imputed (BioVU info score – 0.9998; CARDIA info score – 0.994). The dotted red line represents the significance threshold.



Appendix FF. Admixture mapping analysis of chromosome 7 with overlapping single SNP association analysis results.

The X-axis indicates the genomic position along chromosome 7 in Mb. The top of the Y-axis indicates -$\log_{10}$(*p* value) from the meta-analysis between BioVU and CARDIA AA from logistic regression of admixture mapping values that were generated from LAMP-ANC (solid black line – not conditioned for rs782525; dashed green line – conditioned on rs7872525) with fibroid number (single vs multiple) as the outcome. The admixture mapping analysis was adjusted for age, BMI, five PCs. The solid green line represents the significance threshold. The solid blue line represents the suggestive threshold. The bottom portion of the Y-axis indicates the $\log_{10}$(*p* value) for the single SNP association analyses (blue circles indicates genotyped SNPs; red circles indicates imputed SNPs) with fibroid number (single vs multiple) as the outcome. The imputed region was found by taking 1 –log down from the most significant mapping peak above the suggest threshold. The imputed region encompassed 129,822,797 to 134,670,462 bp. The single SNP association analysis was adjusted for age, BMI, and five PCs. The SNP, rs782525, was imputed (BioVU info score – 0.999; CARDIA info score – 0.985). The dotted red line represents the significance threshold.

Appendix GG. Admixture mapping for number of fibroids in African American women in candidate regions that have been implicated in prior studies.

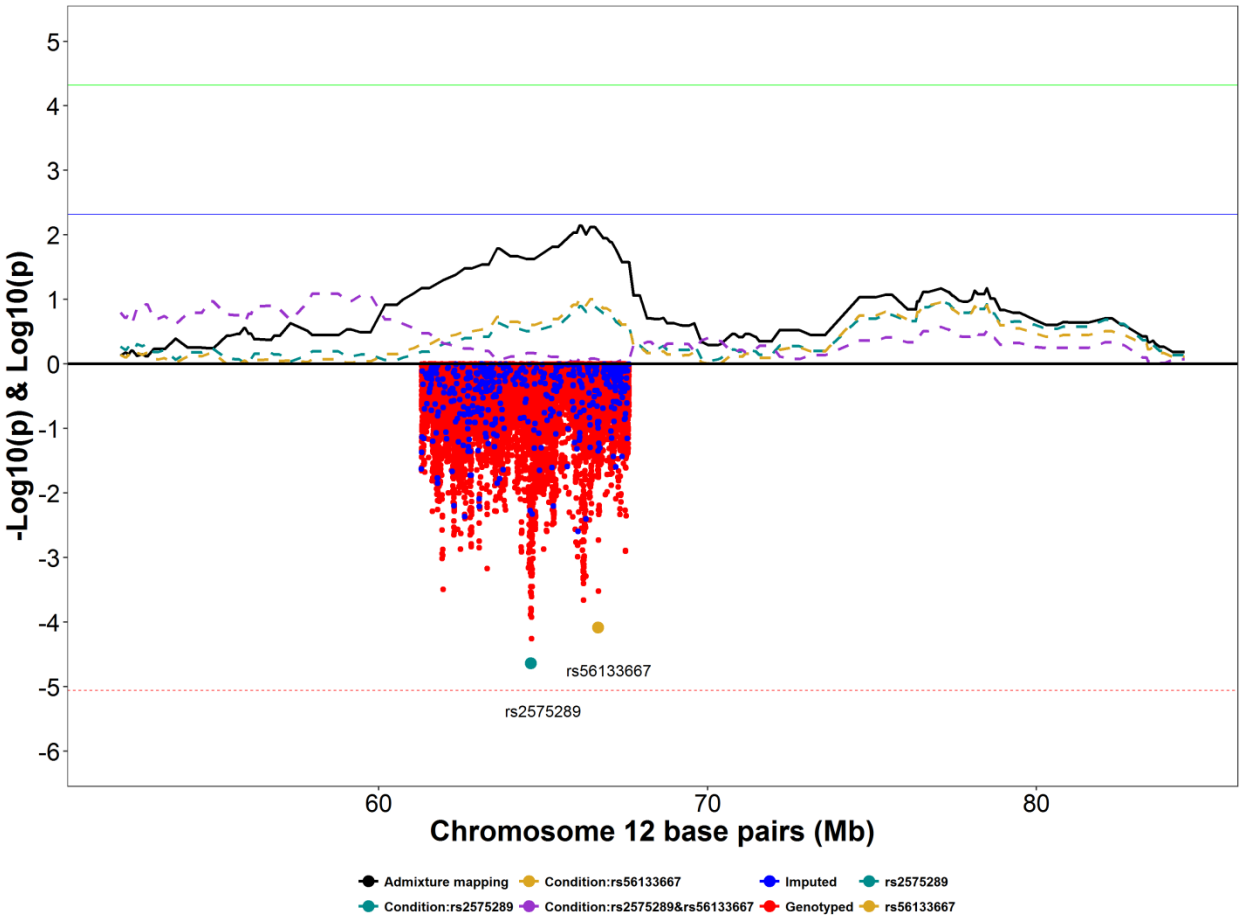
		Number [e]							
		12q14.1-q14.3							
Nearby Gene	**Cohort**	**OR [95% CI] [a]**	**p [a]**	**OR [95% CI] [b]**	**p [b]**	**OR [95% CI] [c]**	**p [c]**	**OR [95% CI] [d]**	**p [d]**
	BioVU	1.45 [0.98, 2.17]	0.066	1.25 [0.82, 1.90]	0.294	1.25 [0.82, 1.90]	0.294	1.02 [0.65, 1.60]	0.941
HMGA2	CARDIA	3.58 [1.36, 9.42]	$9.60x10^{-3}$	2.11 [0.77, 5.79]	0.149	2.62 [0.92, 7.41]	0.070	1.26 [0.39, 4.06]	0.694
	Meta-analysis	1.66 [1.15, 2.40]	$7.21x10^{-3}$	1.35 [0.92, 1.99]	0.128	1.39 [0.94, 2.05]	0.099	1.05 [0.69, 1.59]	0.834

a Model 1: Adjusted for age, BMI, and 5 PCs;
b Model 2: Adjusted for age, BMI, and 5 PCs + adjustment for most significant single SNP;
c Model 3: Adjusted for age, BMI, and 5 PCs + adjustment for second most significant single SNP;
d Model 4: Adjusted for age, BMI, and 5 PCs + adjustment for most and second most significant single SNP;
e Number is coded as single vs multiple fibroids;
CI-confidence interval; p-p-value.

Appendix HH. Single SNP signals in admixture mapping regions for number of fibroids in African American women in candidate regions that have been implicated in prior studies.

Number [b]				
rs2575289 in 12q14.1-q14.3				
Cohort	**OR [95% CI] [a]**	**p**	**EA/RA**	**BioVU/CARDIA EAF**
BioVU	0.49 [0.16, 0.82]	$3.63x10^{-3}$	A/G	0.24/0.26
CARDIA	1.13 [0.50, 1.73]	$4.05x10^{-4}$		
Meta-analysis	0.63 [0.34, 0.92]	$2.29x10^{-5}$		
rs56133667 in 12q14.1-q14.3				
Cohort	**OR [95% CI] [a]**	**p**	**EA/RA**	**BioVU/CARDIA EAF**
BioVU	1.32 [0.36, 2.28]	$7.30x10^{-3}$	T/G	0.08/0.05
CARDIA	0.98 [0.33, 1.64]	$3.29x10^{-3}$		
Meta-analysis	1.09 [0.55, 1.63]	$8.19x10^{-5}$		

a Model 1: Adjusted for age, BMI, and 5 PCs;
b Number is coded as single vs multiple fibroids;
CI-confidence interval; p-p-value; EA-effect allele; RA-reference allele; EAF-effect allele frequency.



Appendix II. Admixture mapping analysis of chromosome 12 with overlapping single SNP association analysis results.

The X-axis indicates the genomic position along chromosome 10 in Mb. The top of the Y-axis indicates -$\log_{10}$(*p* value) from the meta-analysis between BioVU and CARDIA AA from logistic regression of admixture mapping values that were generated from LAMP-ANC (solid black line – not conditioned for rs2575289 or rs56133667; dashed teal line – conditioned on rs2575289; dashed yellow line - conditioned on rs56133667; dashed purple line - conditioned on rs2575289 and rs56133667) with fibroid number (single vs multiple) as the outcome. The admixture mapping analysis was adjusted for age, BMI, five PCs. The solid green line represents the significance threshold. The solid blue line represents the suggestive threshold. The bottom portion of the Y-axis indicates the $\log_{10}$(*p* value) for the single SNP association analyses (blue circles indicates genotyped SNPs; red circles indicates imputed SNPs) with fibroid number (single vs multiple) as the outcome. The imputed region was found by taking 1 –log down from the most significant mapping peak above the suggest threshold. The imputed region encompassed 61,300,873 to 67,610,913 bp. The single SNP association analysis was adjusted for age, BMI, and five PCs. The SNP, rs2575289, was imputed (BioVU info score – 0.976; CARDIA info score – 0.991). The SNP, rs56133667, was imputed (BioVU info score > 0.999; CARDIA info score > 0.999). The dotted red line represents the significance threshold of the single SNP association analysis.

REFERENCES

1. Vollenhoven B. Introduction: the epidemiology of uterine leiomyomas. Bailliere's clinical obstetrics and gynaecology 1998;12:169-76.
2. Baird DD, Dunson DB, Hill MC, Cousins D, Schectman JM. High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence. American journal of obstetrics and gynecology 2003;188:100-7.
3. De La Cruz MS, Buchanan EM. Uterine Fibroids: Diagnosis and Treatment. American family physician 2017;95:100-7.
4. Zimmermann A, Bernuit D, Gerlinger C, Schaefers M, Geppert K. Prevalence, symptoms and management of uterine fibroids: an international internet-based survey of 21,746 women. BMC women's health 2012;12:6.
5. Farquhar CM, Steiner CA. Hysterectomy rates in the United States 1990-1997. Obstetrics and gynecology 2002;99:229-34.
6. Cardozo ER, Clark AD, Banks NK, Henne MB, Stegmann BJ, Segars JH. The estimated annual cost of uterine leiomyomata in the United States. American journal of obstetrics and gynecology 2012;206:211 e1-9.
7. Marshall LM, Spiegelman D, Barbieri RL, Goldman MB, Manson JE, Colditz GA *et al.* Variation in the incidence of uterine leiomyoma among premenopausal women by age and race. Obstetrics and gynecology 1997;90:967-73.
8. Wilcox LS, Koonin LM, Pokras R, Strauss LT, Xia Z, Peterson HB. Hysterectomy in the United States, 1988-1990. Obstetrics and gynecology 1994;83:549-55.
9. Velez Edwards DR, Baird DD, Hartmann KE. Association of age at menarche with increasing number of fibroids in a cohort of women who underwent standardized ultrasound assessment. American journal of epidemiology 2013;178:426-33.
10. Dandolu V, Singh R, Lidicker J, Harmanli O. BMI and uterine size: is there any relationship? International journal of gynecological pathology : official journal of the International Society of Gynecological Pathologists 2010;29:568-71.
11. Ligon AH, Morton CC. Genetics of uterine leiomyomata. Genes, chromosomes & cancer 2000;28:235-45.
12. Ligon AH, Morton CC. Leiomyomata: heritability and cytogenetic studies. Human reproduction update 2001;7:8-14.
13. Morton CC. Genetic approaches to the study of uterine leiomyomata. Environmental health perspectives 2000;108 Suppl 5:775-8.
14. Walker CL, Stewart EA. Uterine fibroids: the elephant in the room. Science (New York, NY) 2005;308:1589-92.
15. Parazzini F. Risk factors for clinically diagnosed uterine fibroids in women around menopause. Maturitas 2006;55:174-9.
16. Parazzini F, Negri E, La Vecchia C, Chatenoud L, Ricci E, Guarnerio P. Reproductive factors and risk of uterine fibroids. Epidemiology (Cambridge, Mass) 1996;7:440-2.
17. Wise LA, Palmer JR, Stewart EA, Rosenberg L. Polycystic ovary syndrome and risk of uterine leiomyomata. Fertility and sterility 2007;87:1108-15.
18. Velez Edwards DR, Hartmann KE, Wellons M, Shah A, Xu H, Edwards TL. Evaluating the role of race and medication in protection of uterine fibroids by type 2 diabetes exposure. BMC women's health 2017;17:28.

19. Snieder H, MacGregor AJ, Spector TD. Genes control the cessation of a woman's reproductive life: a twin study of hysterectomy and age at menopause. The Journal of clinical endocrinology and metabolism 1998;83:1875-80.
20. Luoto R, Kaprio J, Rutanen EM, Taipale P, Perola M, Koskenvuo M. Heritability and risk factors of uterine fibroids--the Finnish Twin Cohort study. Maturitas 2000;37:15-26.
21. Feingold-Link L, Edwards TL, Jones S, Hartmann KE, Velez Edwards DR. Enhancing uterine fibroid research through utilization of biorepositories linked to electronic medical record data. Journal of women's health 2014;23:1027-32.
22. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balser JR *et al.* Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clinical pharmacology and therapeutics 2008;84:362-9.
23. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. American journal of human genetics 2011;88:76-82.
24. Bray MJ, Edwards TL, Wellons MF, Jones SH, Hartmann KE, Velez Edwards DR. Admixture mapping of uterine fibroid size and number in African American women. Fertility and sterility 2017;108:1034-42 e26.
25. Stewart EA. Uterine fibroids. Lancet 2001;357:293-8.
26. Lagana AS, Vergara D, Favilli A, La Rosa VL, Tinelli A, Gerli S *et al.* Epigenetic and genetic landscape of uterine leiomyomas: a current view over a common gynecological disease. Archives of gynecology and obstetrics 2017.
27. Linder D, Gartler SM. Glucose-6-phosphate dehydrogenase mosaicism: utilization as a cell marker in the study of leiomyomas. Science (New York, NY) 1965;150:67-9.
28. Cramer SF, Patel A. The frequency of uterine leiomyomas. American journal of clinical pathology 1990;94:435-8.
29. Buttram VC, Jr., Reiter RC. Uterine leiomyomata: etiology, symptomatology, and management. Fertility and sterility 1981;36:433-45.
30. Ciavattini A, Clemente N, Delli Carpini G, Di Giuseppe J, Giannubilo SR, Tranquilli AL. Number and size of uterine fibroids and obstetric outcomes. The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstet 2015;28:484-8.
31. Parker WH. Uterine myomas: management. Fertility and sterility 2007;88:255-71.
32. Vilos GA, Allaire C, Laberge PY, Leyland N, Special C. The management of uterine leiomyomas. Journal of obstetrics and gynaecology Canada : JOGC = Journal d'obstetrique et gynecologie du Canada : JOGC 2015;37:157-78.
33. Roberts A. Magnetic resonance-guided focused ultrasound for uterine fibroids. Seminars in interventional radiology 2008;25:394-405.
34. Flake GP, Andersen J, Dixon D. Etiology and pathogenesis of uterine leiomyomas: a review. Environmental health perspectives 2003;111:1037-54.
35. Marshall LM, Spiegelman D, Goldman MB, Manson JE, Colditz GA, Barbieri RL *et al.* A prospective study of reproductive factors and oral contraceptive use in relation to the risk of uterine leiomyomata. Fertility and sterility 1998;70:432-9.
36. Schneider J, Bradlow HL, Strain G, Levin J, Anderson K, Fishman J. Effects of obesity on estradiol metabolism: decreased formation of nonuterotropic metabolites. The Journal of clinical endocrinology and metabolism 1983;56:973-8.

37. Ross RK, Pike MC, Vessey MP, Bull D, Yeates D, Casagrande JT. Risk factors for uterine fibroids: reduced risk associated with oral contraceptives. British medical journal 1986;293:359-62.
38. Marshall LM, Spiegelman D, Manson JE, Goldman MB, Barbieri RL, Stampfer MJ *et al.* Risk of uterine leiomyomata among premenopausal women in relation to body size and cigarette smoking. Epidemiology (Cambridge, Mass) 1998;9:511-7.
39. Moorman PG, Leppert P, Myers ER, Wang F. Comparison of characteristics of fibroids in African American and white women undergoing premenopausal hysterectomy. Fertility and sterility 2013;99:768-76.e1.
40. Okolo S. Incidence, aetiology and epidemiology of uterine fibroids. Best practice & research Clinical obstetrics & gynaecology 2008;22:571-88.
41. Parazzini F, La Vecchia C, Negri E, Cecchetti G, Fedele L. Epidemiologic characteristics of women with uterine fibroids: a case-control study. Obstetrics and gynecology 1988;72:853-7.
42. Baird DD, Travlos G, Wilson R, Dunson DB, Hill MC, D'Aloisio AA *et al.* Uterine leiomyomata in relation to insulin-like growth factor-I, insulin, and diabetes. Epidemiology (Cambridge, Mass) 2009;20:604-10.
43. Ge T, Chen CY, Neale BM, Sabuncu MR, Smoller JW. Phenome-wide heritability analysis of the UK Biobank. PLoS genetics 2017;13:e1006711.
44. Styer AK, Rueda BR. The Epidemiology and Genetics of Uterine Leiomyoma. Best practice & research Clinical obstetrics & gynaecology 2016;34:3-12.
45. Gross KL, Morton CC. Genetics and the development of fibroids. Clinical obstetrics and gynecology 2001;44:335-49.
46. Sandberg AA. Updates on the cytogenetics and molecular genetics of bone and soft tissue tumors: leiomyoma. Cancer genetics and cytogenetics 2005;158:1-26.
47. Walker CL, Hunter D, Everitt JI. Uterine leiomyoma in the Eker rat: a unique model for important diseases of women. Genes, chromosomes & cancer 2003;38:349-56.
48. Taatjes DJ. The human Mediator complex: a versatile, genome-wide regulator of transcription. Trends in biochemical sciences 2010;35:315-22.
49. Makinen N, Mehine M, Tolvanen J, Kaasinen E, Li Y, Lehtonen HJ *et al.* MED12, the mediator complex subunit 12 gene, is mutated at high frequency in uterine leiomyomas. Science (New York, NY) 2011;334:252-5.
50. Cha PC, Takahashi A, Hosono N, Low SK, Kamatani N, Kubo M *et al.* A genome-wide association study identifies three loci associated with susceptibility to uterine fibroids. Nature genetics 2011;43:447-50.
51. Hellwege JN, Jeff JM, Wise LA, Gallagher CS, Wellons M, Hartmann KE *et al.* A multi-stage genome-wide association study of uterine fibroids in African Americans. Human genetics 2017.
52. Edwards TL, Michels KA, Hartmann KE, Velez Edwards DR. BET1L and TNRC6B associate with uterine fibroid risk among European Americans. Human genetics 2013;132:943-53.
53. Edwards TL, Hartmann KE, Velez Edwards DR. Variants in BET1L and TNRC6B associate with increasing fibroid volume and fibroid type among European Americans. Human genetics 2013;132:1361-9.
54. Kho AN, Hayes MG, Rasmussen-Torvik L, Pacheco JA, Thompson WK, Armstrong LL *et al.* Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes

within a genome-wide association study. Journal of the American Medical Informatics Association : JAMIA 2012;19:212-8.
55. . Obesity: preventing and managing the global epidemic. Report of a WHO consultation. World Health Organization technical report series 2000;894:i-xii, 1-253.
56. Eggert SL, Huyck KL, Somasundaram P, Kavalla R, Stewart EA, Lu AT *et al.* Genome-wide linkage and association analyses implicate FASN in predisposition to Uterine Leiomyomata. American journal of human genetics 2012;91:621-8.
57. Pulley J, Clayton E, Bernard GR, Roden DM, Masys DR. Principles of human subjects protections applied in an opt-out, de-identified biobank. Clinical and translational science 2010;3:42-8.
58. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics 2007;81:559-75.
59. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS genetics 2009;5:e1000529.
60. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. American journal of human genetics 2011;88:294-305.
61. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics 2006;38:904-9.
62. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research 2010;38:e164.
63. Broman KW, Matsumoto N, Giglio S, Martin CL, Roseberry JA, Zuffardi O *et al.* Common long human inversion polymorphism on chromosome 8p. Lecture Notes-Monograph Series 2003:237-45.
64. Davis LK, Yu D, Keenan CL, Gamazon ER, Konkashbaev AI, Derks EM *et al.* Partitioning the heritability of Tourette syndrome and obsessive compulsive disorder reveals differences in genetic architecture. PLoS genetics 2013;9:e1003864.
65. Aissani B, Zhang K, Wiener H. Follow-up to genome-wide linkage and admixture mapping studies implicates components of the extracellular matrix in susceptibility to and size of uterine fibroids. Fertility and sterility 2015;103:528-34 e13.
66. Winkler CA, Nelson GW, Smith MW. Admixture mapping comes of age. Annual review of genomics and human genetics 2010;11:65-89.
67. Reich D, Patterson N, De Jager PL, McDonald GJ, Waliszewska A, Tandon A *et al.* A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. Nature genetics 2005;37:1113-8.
68. Velez Edwards DR, Tsosie KS, Williams SM, Edwards TL, Russell SB. Admixture mapping identifies a locus at 15q21.2-22.3 associated with keloid formation in African Americans. Human genetics 2014;133:1513-23.
69. Freedman ML, Haiman CA, Patterson N, McDonald GJ, Tandon A, Waliszewska A *et al.* Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. Proceedings of the National Academy of Sciences of the United States of America 2006;103:14068-73.

70. Wise LA, Ruiz-Narvaez EA, Palmer JR, Cozier YC, Tandon A, Patterson N *et al.* African ancestry and genetic risk for uterine leiomyomata. American journal of epidemiology 2012;176:1159-68.
71. Zhang K, Wiener H, Aissani B. Admixture mapping of genetic variants for uterine fibroids. Journal of human genetics 2015;60:533-8.
72. Patterson N, Hattangadi N, Lane B, Lohmueller KE, Hafler DA, Oksenberg JR *et al.* Methods for high-density admixture mapping of disease genes. American journal of human genetics 2004;74:979-1000.
73. Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG *et al.* Control of confounding of genetic associations in stratified populations. American journal of human genetics 2003;72:1492-504.
74. McKeigue PM, Carpenter JR, Parra EJ, Shriver MD. Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. Annals of human genetics 2000;64:171-86.
75. Hoggart CJ, Shriver MD, Kittles RA, Clayton DG, McKeigue PM. Design and analysis of admixture mapping studies. American journal of human genetics 2004;74:965-78.
76. Giri A, Edwards TL, Hartmann KE, Torstenson ES, Wellons M, Schreiner PJ *et al.* African genetic ancestry interacts with body mass index to modify risk for uterine fibroids. PLoS genetics 2017;13:e1006871.
77. Friedman GD, Cutter GR, Donahue RP, Hughes GH, Hulley SB, Jacobs DR, Jr. *et al.* CARDIA: study design, recruitment, and some characteristics of the examined subjects. Journal of clinical epidemiology 1988;41:1105-16.
78. Wellons MF, Lewis CE, Schwartz SM, Gunderson EP, Schreiner PJ, Sternfeld B *et al.* Racial differences in self-reported infertility and risk factors for infertility in a cohort of black and white women: the CARDIA Women's Study. Fertility and sterility 2008;90:1640-8.
79. Lewis CE, Smith DE, Wallace DD, Williams OD, Bild DE, Jacobs DR, Jr. Seven-year trends in body weight and associations with lifestyle and behavioral characteristics in black and white young adults: the CARDIA study. American journal of public health 1997;87:635-42.
80. Genomes Project C, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM *et al.* A map of human genome variation from population-scale sequencing. Nature 2010;467:1061-73.
81. Baran Y, Pasaniuc B, Sankararaman S, Torgerson DG, Gignoux C, Eng C *et al.* Fast and accurate inference of local ancestry in Latino populations. Bioinformatics 2012;28:1359-67.
82. Sankararaman S, Sridhar S, Kimmel G, Halperin E. Estimating local ancestry in admixed populations. American journal of human genetics 2008;82:290-303.
83. Pasaniuc B, Sankararaman S, Kimmel G, Halperin E. Inference of locus-specific ancestry in closely related populations. Bioinformatics 2009;25:i213-21.
84. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190-1.
85. Gao X. Multiple testing corrections for imputed SNPs. Genetic epidemiology 2011;35:154-8.
86. Gao X, Becker LC, Becker DM, Starmer JD, Province MA. Avoiding the high Bonferroni penalty in genome-wide association studies. Genetic epidemiology 2010;34:100-5.
87. Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genetic epidemiology 2008;32:361-9.

88. Seifert L, Werba G, Tiwari S, Giao Ly NN, Alothman S, Alqunaibit D *et al.* The necrosome promotes pancreatic oncogenesis via CXCL1 and Mincle-induced immune suppression. Nature 2016;532:245-9.
89. Yamada D, Watanabe S, Kawahara K, Maeda T. Plexin A1 signaling confers malignant phenotypes in lung cancer cells. Biochemical and biophysical research communications 2016;480:75-80.
90. Kim E, Beon J, Lee S, Park J, Kim S. IPMK: A versatile regulator of nuclear signaling events. Advances in biological regulation 2016;61:25-32.
91. Sato N, Maeda M, Sugiyama M, Ito S, Hyodo T, Masuda A *et al.* Inhibition of SNW1 association with spliceosomal proteins promotes apoptosis in breast cancer cells. Cancer medicine 2015;4:268-77.
92. Narayanaswamy PB, Hodjat M, Haller H, Dumler I, Kiyan Y. Loss of urokinase receptor sensitizes cells to DNA damage and delays DNA repair. PloS one 2014;9:e101529.
93. Cui QZ, Tang ZP, Zhang XP, Zhao HY, Dong QZ, Xu K *et al.* Leucine zipper tumor suppressor 2 inhibits cell proliferation and regulates Lef/Tcf-dependent transcription through Akt/GSK3beta signaling pathway in lung cancer. The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society 2013;61:659-70.
94. Johnson DT, Luong R, Lee SH, Peng Y, Shaltouki A, Lee JT *et al.* Deletion of leucine zipper tumor suppressor 2 (Lzts2) increases susceptibility to tumor development. The Journal of biological chemistry 2013;288:3727-38.
95. . Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science (New York, NY) 2015;348:648-60.
96. Reeves R. Molecular biology of HMGA proteins: hubs of nuclear function. Gene 2001;277:63-81.
97. Boonyaratanakornkit V, Melvin V, Prendergast P, Altmann M, Ronfani L, Bianchi ME *et al.* High-mobility group chromatin proteins 1 and 2 functionally interact with steroid hormone receptors to enhance their DNA binding in vitro and transcriptional activity in mammalian cells. Molecular and cellular biology 1998;18:4471-87.
98. Verrier CS, Roodi N, Yee CJ, Bailey LR, Jensen RA, Bustin M *et al.* High-mobility group (HMG) protein HMG-1 and TATA-binding protein-associated factor TAF(II)30 affect estrogen receptor-mediated transcriptional activation. Molecular endocrinology (Baltimore, Md) 1997;11:1009-19.
99. Romine LE, Wood JR, Lamia LA, Prendergast P, Edwards DP, Nardulli AM. The high mobility group protein 1 enhances binding of the estrogen receptor DNA binding domain to the estrogen response element. Molecular endocrinology (Baltimore, Md) 1998;12:664-74.
100. Zhang CC, Krieg S, Shapiro DJ. HMG-1 stimulates estrogen response element binding by estrogen receptor from stably transfected HeLa cells. Molecular endocrinology (Baltimore, Md) 1999;13:632-43.