EVALUATION OF INSULIN-LIKE GROWTH FACTOR POLYMORPHISMS WITH PREVALENCE AND SIZE OF UTERINE LEIOMYOMATA

Aimee Angela D'Aloisio

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Epidemiology.

Chapel Hill
2007

Approved by:

Advisor: Jane C. Schroeder, DVM, PhD

Reader: Donna D. Baird, PhD

Reader: Kari E. North, PhD

Reader: Charles Poole, ScD

Reader: Suzanne L. West, PhD

ABSTRACT

Aimee Angela D'Aloisio: Evaluation of Insulin-like Growth Factor Polymorphisms with Prevalence and Size of Uterine Leiomyomata
(Under the direction of Jane C. Schroeder, DVM, PhD)

Genetic factors influence circulating insulin-like growth factor-one (IGF-I) and IGF binding protein-3 (IGFBP-3) levels. Prior studies with multiple *IGF-I* and *IGFBP-3* polymorphisms have been limited, especially among African Americans. We evaluated 30 *IGF-I* and 15 *IGFBP-3* single nucleotide polymorphisms (SNPs) and estimated diplotypes in relation to plasma levels of both proteins among 984 premenopausal African Americans and Caucasians from the National Institute of Environmental Health Sciences Uterine Fibroid Study. In both racial groups, *IGFBP-3* rs2854746 (Ala32Gly) was associated with plasma IGFBP-3, (CC versus GG: Caucasians: 631 ng/ml, 95% confidence interval (CI): 398, 864; African Americans: 897 ng/ml, 95% CI: 656, 1138). Relative to diplotypes with the rs2854746 CG genotype, *IGFBP-3* diplotypes with the GG genotype had lower mean plasma IGFBP-3 while *IGFBP-3* diplotypes with the CC genotype had higher mean plasma IGFBP-3. The *IGFBP-3* promoter SNP, rs2854744, which was in strong linkage disequilibrium with rs2854746 in Caucasians only, was associated with plasma IGFBP-3 in both races. Eight additional *IGFBP-3* SNPs were associated with plasma IGFBP-3, with generally consistent associations between races. Twelve *IGF-I* SNPs were associated with plasma IGF-I; however, associations were discordant between races, and were not consistent with diplotype findings.

Uterine leiomyomata (fibroids) are responsible for substantial morbidity, especially among African Americans. Gene expression studies suggest IGF-I involvement in fibroid pathogenesis; IGFBP-3 may be important based on biological interrelations with IGF-I. *IGF-I* and *IGFBP-3* polymorphisms have not been previously studied with fibroids. We evaluated the *IGF-I* and *IGFBP-3* SNPs and estimated diplotypes from our first study in association with fibroid prevalence in our African American and Caucasian study population. Relatively precise prevalence differences (PD) with *IGF-I* and *IGFBP-3* SNPs were predominantly estimated among African Americans, including *IGFBP-3* SNPs of rs9282734 (His158Pro) (PD = -0.130, 95% CI: -0.294, 0.034) and rs2475551 (splice site) (PD = 0.208, 95% CI: 0.095, 0.320) and *IGF-I* SNP rs35767 (promoter) (PD = 0.208, 95% CI: 0.095, 0.320). Associations with larger fibroids (2+ cm) were consistent or slightly weaker than with any fibroids. Diplotype associations were not consistent with SNP findings. Future research should validate our findings and examine additional genes within the IGF-I pathway.

ACKNOWLEDGEMENTS

I would like to acknowledge and express my gratitude to the following people for without their help completing my dissertation would not have been possible. My committee members were very generous with their time, experience, and insight, and I will always be grateful. I would like to thank Dr. Jane Schroeder for her support and friendship as committee chair and advisor. A special thanks to Dr. Donna Baird, my mentor at the National Institute of Environmental Health Sciences (NIEHS), for the use of the NIEHS Uterine Fibroid Study data on which my dissertation is based. Dr. Baird was always available with her unfailing support and insight through this whole process. I wish to thank Dr. Kari North for her help and the time she took to guide me through the genetics of my dissertation. I would also like to thank Dr. Charles Poole for his support and help with statistical methods, and his unfailing patience and guidance. I also wish to thank Dr. Suzanne West for not only the support and help she gave as a member of my committee, but also for providing the financial support for the genotyping through the UNC-GSK Center of Excellence in Pharmacoepidemiology & Public Health. In addition, I would like to acknowledge that this research was supported by an NIEHS Intramural Research Training Award.

Other people I would like to acknowledge are: Dr. Jason Luo from the Mammalian Genotyping Core at the Lineberger Comprehensive Cancer Center for his direction of the genotyping; the staff at Social & Scientific Systems including Dr. Mary Watson for

preparation of DNA samples for genotyping; Dr. Greg Travlos from NIEHS for conducting the immunoassays; Dr. Sue Edelstein from NIEHS Graphics for her invaluable assistance in preparing dissertation figures; the entire Epidemiology Branch at NIEHS with whom I look forward to continue working with as a postdoctoral fellow.

Finally, I would like to thank my mother for her constant support throughout this long process to my PhD.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF ABBREVIATIONS

A	Adenine
ABI	Applied Biosystems
Ala	Alanine
BMI	Body Mass Index
bp	Base Pairs
C	Cytosine
CI	Confidence Interval
DNA	Deoxyribonucleic Acid
G	Guanine
GVS	Genome Variation Server
Gly	Glycine
His	Histidine
HWE	Hardy-Weinberg Equilibrium
IGF-I	Insulin-like Growth Factor-1 Gene
IGF-I	Insulin-like Growth Factor-1
IGF-IR	Insulin-like Growth Factor-1 Receptor
IGFBP-3	Insulin-like Growth Factor Binding Protein-3 Gene
IGFBP-3	Insulin-like Growth Factor Binding Protein-3
kb	Kilo Base Pairs
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency

mRNA	Messenger Ribonucleic Acid
µL	Microliter
ng/mL	Nanograms per Milliliter
NIEHS	National Institute of Environmental Health Sciences
PD	Prevalence Difference
Pro	Proline
SAS	Statistical Analysis System
SD	Standard Deviation
SNP	Single Nucleotide Polymorphism
T	Thymine
Thr	Threonine
UFS	Uterine Fibroid Study

CHAPTER I

REVIEW OF THE LITERATURE

Epidemiology of uterine fibroids

Uterine leiomyomata (fibroids) are hormonally-dependent benign tumors of smooth muscle origin that commonly result in pelvic pain and menstrual irregularities including heavy bleeding in addition to potential fertility problems (1). Uterine fibroids are the principal indication for hysterectomies in the United States, resulting in more than 200,000 of these surgical procedures per year (2, 3). The National Institute of Environmental Health Sciences (NIEHS) Uterine Fibroid Study estimated the risk of developing fibroids by age 50 to be over 80% among African Americans and over 60% among Caucasians (4), which is in contrast with previous estimates that focused their definition of fibroids on women with diagnoses based on clinical symptoms or hysterectomy, without accounting for asymptomatic cases.

The most consistent risk factors for uterine fibroids are age, race, and reproductive history. Fibroid prevalence increases with age up to menopause (4-7); the onset of menopause can initiate regression of uterine fibroids, which is most likely due to declines in reproductive hormone levels. Most epidemiologic studies have supported a greater prevalence and higher degree of related morbidity from uterine fibroids among African Americans compared to Caucasians (4, 5, 8). In particular, the Nurses' Health Study II

reported an earlier age of onset and higher age-adjusted incidence rates of hysterectomy or ultrasound confirmed fibroids among African Americans (30.6 per 1,000 woman-years) compared with Caucasians (8.9 per 1,000 woman-years) (5). Only one case-control study did not support an association between African American race and fibroids, but its study population was less than 10% African American (6). In addition, African American women have been reported to have more fibroid-related morbidity, a greater number and size of tumors, and an earlier age of diagnosis compared to Caucasians (9). Early menarche has been related to fibroid prevalence in three studies (8, 10, 11) in addition to the NIEHS Uterine Fibroid Study (unpublished) although a small case-control study did not support this finding (6). Causal explanation for relations between early onset of menarche and fibroids involve possible alterations in estrogen and progesterone levels in adults, but this mechanism is not established, and early age at menarche may instead act as a marker for other hormonally-related exposures that promote fibroids (12).

Parity has been inversely associated with uterine fibroids (13, 14); however, the timing of births may be a more important factor than the overall number (8) since stronger associations have been observed with births after 24 years of age (15). Baird et al. postulated that the protective association of live births after age 24 with fibroids is due to uterine remodeling subsequent to each completed pregnancy, which may eliminate small fibroids. According to this hypothesis, only births during mid-reproductive years, which is the time of peak tumor development, would decrease fibroid prevalence (15). Others have suggested that the inverse associations between fibroids and parity may be confounded by infertility or breastfeeding; however, the parity association was not reduced by exclusion of infertile women in two large cohort studies (10, 11), and

breastfeeding has not been associated with fibroids in the few studies that have evaluated it as a risk factor (6, 11). The relation of infertility with uterine fibroids has not been resolved. Some studies find associations (10, 16), but this could be due to reverse causation and possibly common causes.

Smoking has been inconsistently associated with uterine fibroids (8, 13, 14, 16), with inverse associations attributed to reduced estrogen levels in smokers versus non-smokers. However, neither the Nurses' Health Study II (17) nor the Black Women's Health study (18) found an association between smoking and fibroids. Obesity has been positively associated with fibroids, possibly as a consequence of elevated estrogen levels (8, 16, 17, 19). Although relations between physical activity and fibroids are highly plausible, epidemiologic investigations have been very limited. Specifically, physical activity has been associated with increased levels of estrogen metabolites with low receptor affinity among premenopausal women in Bentz et al. (20), but Campbell et al. did not confirm this finding (21). The NIEHS study noted an inverse association between an index of current physical activity and fibroid prevalence (22), and Frisch et al. reported reduced clinically diagnosed fibroids among former college athletes compared to non-athletes (23). Two studies have reported positive associations between alcohol intake and fibroids. In the NIEHS Uterine Fibroid Study, there was a threshold effect in which even low levels of alcohol consumption (at least 0.5 drinks per week) were associated with an increased fibroid prevalence among Caucasians, while consumption of over 2 drinks per week were positively associated with fibroids in African Americans (24). These results are somewhat consistent with the Black Women's Health Study, which suggested that

strongest associations were for consumption of at least 7 drinks per week (especially beer intake) and duration of consumption of at least 20 years (18).

Associations between fibroids and oral contraceptive use have been inconsistent, with some studies reporting inverse associations with current oral contraceptive use (8, 25) and duration of oral contraceptive use (8, 14), while another reported a positive association with ever use of oral contraceptives (16). The Black Women's Health and Nurses' Health Studies did not find an association with current oral contraceptives overall, but instead suggested that use during teenage years may be associated with increased fibroid risk (10, 11). Originally, estrogen was regarded as the main hormone involved in fibroid development, but more recently progesterone also has been considered important in growth promotion of fibroids (26, 27). A hypothesis involving indirect effects of estrogens or progesterones on fibroids through accelerating transcription of cytokines or growth factors has been proposed (28, 29).

Early research suggested that reported positive associations with hypertension and uterine fibroids were due to fibroid-mediated hypertension secondary to obstruction of the urinary tract by the growing uterine tumors (30, 31). However, two recent studies have refuted this hypothesis because they did not support hypertension occurring subsequent to fibroids and instead suggest that fibroids and hypertension may share a common pathogenic mechanism (32, 33). Shared associations with race, obesity, blood pressure, and physical activity, support a recent hypothesis that atherosclerosis and uterine fibroids share a common biologic mechanism involving hyperinsulinemia (1, 33). In addition, smooth muscle proliferation is involved in the pathogenesis of both

atheromatous plaques and fibroids (34). However, fasting insulin was not associated with increased prevalence of fibroids in the NIEHS Uterine Fibroid Study (unpublished).

Studies of genetic susceptibility to fibroids among women in the United States have been limited; however, Al-Hendy et al. reported that a catechol-o-methyltransferase (*COMT*) polymorphism (Val158Met) (35) and an estrogen receptor-α polymorphism (36) were both related to fibroids, and that the proportion of women with high-risk genotypes was greater among African Americans than Caucasians or Hispanics. However, Gooden et al. reported no association between the *COMT* Val158Met single nucleotide polymorphism (SNP) and fibroids within the NIEHS Uterine Fibroid Study (37). A relative increase in the prevalence of fibroids was noted among Taiwanese women with SNPs in the progesterone receptor, epidermal growth factor receptor and tumor-necrosis factor-alpha genes (38-40) in addition to associations with a CAG trinucleotide repeat polymorphism and a TA dinucleotide repeat polymorphism in the androgen and estrogen receptor genes respectively (41, 42). Also, uterine fibroids have been associated with a DNA repair gene (*XRCC1*) polymorphism among Koreans (43), and a cytochrome P450 (*CYP17*) polymorphism in a small South African study (44). Three small German case-control studies reported that uterine fibroids were not associated with estrogen receptor-alpha (45), *COMT* (45), or angiopoietin-2 polymorphisms (46), but were associated with a *p53* polymorphism (46) and possibly with three cytochrome P450 polymorphisms (*CYP17A* (45), *CYP2A13* (47), *CYP1A1* (47)). Null findings were reported in two Japanese studies investigating a cytochrome P450 (*CYP17*) polymorphism (48) and matrix metalloproteinase polymorphisms (49). Overall, conclusions that may be inferred from these genetic susceptibility studies are limited due to their inconsistent findings,

small sample sizes, emphasis on statistically significant findings, examination of only clinically diagnosed or hysterectomy treated fibroids, and lack of representative controls (45-47).

Insulin-like growth factor overview

Insulin-like growth factors (IGFs), formerly known as somatomedins, are a family of molecules that include two IGFs (IGF-I and IGF-II), two IGF receptors (IGF-IR and IGF-IIR), six IGF binding proteins (IGFBP-1 through IGFBP-6), and several IGFBP proteases. IGF-I and IGF-II are polypeptides that are structurally similar to insulin, and have mitogenic properties. IGF-I synthesis is regulated by growth hormone and occurs predominantly in the liver, but is also produced in other tissues. In contrast, IGF-II production is not dependent on growth hormone. IGF-I has historically been the focus of research because of its potential role in disease. Specifically, IGF-I has been implicated in many biologic processes, including stimulation of cell cycle activities, differentiation, proliferation, hormone secretion, and inhibition of apoptosis. IGF-II also has important roles in cell processes, especially in the context of fetal growth and bone development, but it has been less studied regarding its possible involvement in adult disease. Binding to the IGF-IR, which has structural resemblance to the insulin receptor, often precedes IGF-I or IGF-II involvement in cellular processes. IGF-IR has much stronger binding affinity to IGF-I than to IGF-II or insulin (50, 51).

IGFBPs attach to free IGFs in circulation and have even greater binding affinities to IGFs than the IGF-IR. Some IGFBPs reduce the level of IGFs available to act on cells, which suggests that they may inhibit IGF-mediated disease processes; however, IGFBP

binding also could prevent IGF natural degradation in the blood. In addition, IGFBPs may have independent effects such as through interactions with cell surface molecules that may stimulate apoptosis of cancer cells (50, 51). The function of IGFBPs and the preferential binding affinities for IGF-I or IGF-II vary for the six peptides. IGFBP-3, the most studied of the IGFBPs, binds approximately 90% of circulating IGF-I (52). Possible independent effects of IGFBP-3 have included preventing cell growth and inducing apoptosis of breast cancer cells (53, 54) but these findings have not been confirmed.

Insulin-like growth factor involvement with uterine fibroids

Biologic evidence has accumulated in support of the role of IGFs, especially IGF-I, in fibroid development, including *in vitro* evidence of the promotion of fibroid cell growth by IGF-I (55) and increased expression of IGF-I mRNA or peptides (56-58) as well as elevated IGF-IR levels in fibroid tumor tissue as opposed to myometrium (57-60). Wolanska, et al. reported that the relative increase in IGF-I protein detected in fibroids versus normal myometrium was positively related to tumor size (56). Van der Ven et al. reported that increased IGF-I peptide and IGF-IR mRNA levels in fibroids versus myometrium occurred in the absence of differences in IGF-I mRNA levels, which may indicate IGF-I was elevated due to increased binding by IGF-IR rather than increased production of IGF-I in tumor versus myometrial cells (58).

Treatment of fibroids with gonadotropin-releasing hormone analogue, which substantially reduces estrogen and progesterone hormone levels, has been associated with reductions in tumor size and IGF-IR, but not changes in IGF-I protein levels (61). One

study did not find IGF-I mRNA differences between fibroids and myometrium (62) but another reported a decrease in IGF-I mRNA and protein levels in both tissues after treatment with a gonadotropin-releasing hormone analogue (62). Finally, another study detected no differences with IGF-I mRNA but reported increased IGFBP-3 mRNA in fibroids versus myometrium upon gonadotropin-releasing hormone analogue treatment; however, there was elevation of IGF-I mRNA measured in untreated fibroids that were undergoing the proliferative phase of the menstrual cycle (63).

Two studies have supported the involvement of both IGF-I and IGF-II in fibroids (57, 64, 65) based on differential expression in tumors versus the myometrium. Specifically, Boehm, et al. reported elevation of both IGF-I and IGF-II mRNAs (65) in fibroids, and Vollenhoven et al. observed no difference in IGF-I mRNA but increased IGF-II mRNA and reduced IGFBP-3 mRNA in fibroids relative to myometrium (64). Gloudemans, et al. did not support IGF-I or IGF-II involvement with fibroid pathogenesis based on no IGF-I or IGF-II mRNA or protein differences for fibroids versus myometrium (66).

We have hypothesized that inverse associations between IGFBP-3 and fibroids would support IGF-I versus IGF-II involvement in tumor development, especially if increased levels of free IGF-I promote tumor growth. However, it is possible that positive associations between IGFBP-3 and fibroids would also support IGF-I involvement if IGFBP-3 binding to IGF-I increases IGF-I half-life by preserving it from destruction. In addition, effects of IGFBP-3 that are independent of IGF-I binding could affect the estimated association between IGFBP-3 and fibroids.

A recent systematic review of microarray gene expression studies (67) emphasized IGF-II involvement with fibroids by noting that elevated IGF-II mRNA expression in fibroid tumor cells versus normal myometrium is one of the most consistent findings (67-71) (72-74). Microarray studies are a useful screening tool to identify genes that may be involved in fibroid pathogenesis, but results should be interpreted with caution prior to confirmation by other types of studies (75). A recent microarray study reported elevated expression of IGF-I mRNA and protein, along with differences in two factors involved in IGF-I pathways (MAP kinase phosphatase-I and A-myb), in fibroids relative to myometrium after estrogen treatment (76).

IGF-I mRNA concentration is increased in fibroids during the follicular phase relative to the luteal phase of the menstrual cycle (63, 65), which is consistent with IGF-I regulation by estrogen rather than progesterone; however, menstrual cycle variation in IGF-I has not been detected in all studies (77). Englund et al. found evidence of hormonal regulation of IGF-I in fibroids based on positive correlations between circulating estradiol levels and IGF-I expression in fibroid tumors compared with myometrium (77). IGF-I may promote fibroid growth by elevating expression of proliferating cell nuclear antigen (PCNA) and may inhibit apoptosis by promoting Bcl-2 expression in tumor cells (78).

Previous studies have focused primarily on the role of estrogen in regulation of IGF-I in fibroid pathogenesis, but more recent studies have focused on the role of progesterone. Yamada et al. reported that treatment with progesterone but not estrogen was associated with reduced IGF-I mRNA expression in cultured fibroid cells, while neither hormone altered IGF-IR mRNA expression (79). Treatment with a gonadotropin-releasing

hormone analogue resulted in decreased expression of progesterone receptor mRNA and protein and IGF-I mRNA as compared to untreated women in the proliferative phase, but progesterone receptor and IGF-I expression was similar in fibroid and myometrium tissue (62).

In summary, evidence of higher IGF-I mRNA levels in fibroids versus myometrium, variation in the expression of IGF-I according to menstrual cycle phase, and *in vitro* stimulation of fibroid growth by IGF-I supports the role of IGF-I in fibroid development; however, IGF-II is also implicated because of consistent results from microarray studies. Studies with increased IGF-IR expression in fibroids relative to myometrium can support involvement of either IGF-I or IGF-II since both IGFs can bind with this receptor, although IGF-II has lower affinity than IGF-I. In addition, the activity of both IGFs is influenced by IGFBPs, IGF-IR, hormonal regulation and other cellular factors that can affect fibroid associations.

Circulating insulin-like growth factor levels

Genetic factors regulating IGF serum levels have been estimated to account for nearly 40% to 50% of the interindividual variation in IGF-I (80, 81) and approximately 60% of the variation in IGFBP-3 (80, 82). In adults, age is the most consistently reported nongenetic factor associated with IGF-I blood levels (52, 83-92) and to a lesser extent with the IGF-I:IGFBP-3 ratio, (86, 87, 91, 92) with lower circulating IGF-I levels (52, 83-92) and IGF-I:IGFBP-3 ratio (86, 87, 91, 92) associated with increasing age. Women have lower circulating IGF-I (84, 87, 88, 90, 92, 93) but higher IGFBP-3 concentrations (84, 87, 88, 92, 93) than men. In addition, post-menopausal women generally have lower

IGF-I levels than pre-menopausal women; however, distinguishing between age and hormonal influences on pre- and postmenopausal IGF-I levels is difficult. Studies of racial differences in circulating IGF levels among women have been limited, but a few studies have suggested that African American women have greater circulating IGF-I concentrations than Caucasians (88, 94). IGF-I levels among adolescent girls were increased in Caucasians relative to African Americans after adjustment for sexual development and body fat, but mean values were higher in African Americans prior to adjustment (95) .

Blood measurements of IGF-I and IGFBP-3 have been inconsistently associated with BMI (85, 88, 92, 96-98), smoking (84, 85, 93), alcohol consumption (88, 90, 99-101), and physical activity (85-87, 92, 97). Variations between studies may be at least partly due to analytic differences including adjustment for exogenous hormones and factors associated with endogenous hormone levels. Holmes et al., from a cross-sectional analysis of a sample from the Nurses Health Study, found an inverse association between parity and plasma IGF-I levels (85).

Oral contraceptive use was inversely associated with circulating IGF-I but positively related to IGFBP-3 levels in a cross-sectional study of premenopausal nulliparous women (94, 99). The Jernstrom et al. analysis including only Caucasian women also reported that the estradiol dose within oral contraceptives was inversely associated with circulating IGF-I but not IGFBP-3 (99), and the Jernstrom et al. analysis of the entire study population found that inverse associations between IGF-I and oral contraceptives were present among Caucasians but not African Americans (94). Two U.S. cross sectional studies from Holmes et al. and Chang et al. reported inverse relations between

hormone replacement therapy (HRT) and circulating IGF-I (85, 93). Holmes et al. measured the lowest levels of plasma IGF-I and IGFBP-3 among postmenopausal women taking estrogen only, followed by women taking estrogen plus progesterone hormone replacement therapy (HRT) (85).

Jernstrom et al. reported a positive association between age at menarche and plasma IGF-I levels among oral contraceptive users only, but no association with plasma IGFBP-3 levels (99). Probst-Hensch et al., from a cross-sectional analysis of a sample from the Singapore Chinese Health Study, found that women who reported age at menarche age was less than 17 years had lower serum IGF-I than women who reported menarche occurred at 17 years or older (87). However, an Italian cross-sectional study, with participants from a breast cancer chemoprevention trial (unaffected or early invasive breast cancer), reported that women with age at menarche below 13 years had higher circulating concentrations of both IGF-I and IGFBP-3 than women age 13 or older at menarche (91). Although the plasma IGF-I:IGFBP-3 ratio increased as nonusers of oral contraceptives, progressed from the follicular phase to the luteal phase until the start of menses in Jernstrom et al. (99), Helle et al. reported that plasma IGF-I levels were highest in the follicular phase of a Norwegian pilot sample (102), and Wang et al. reported no menstrual cycle variation of serum IGF-I and IGFBP-3 in a small sample of women from Taiwan (103).

Dietary factors have also been investigated in relation to circulating IGF levels. Specifically, dietary and supplementary calcium was positively associated with IGF-I, IGFBP-3, and IGF-I:IGFBP-3 serum levels in Probst-Hensch et al. (87), and milk consumption was related to increased plasma IGF-I levels in Morimoto et al. (cross-

sectional study of controls from The Seattle Colon Cancer Family Registry) and Holmes et al. (92, 101). Holmes et al. reported that plasma IGF-I was positively associated with dietary calcium and Vitamin D but not supplement sources of these nutrients (101). Saturated fat was inversely associated with plasma IGFBP-3 in three studies (87, 101, 104), including Probst-Hensch et al, who also reported positive relations between IGFBP-3 and dietary fiber and omega-3 polyunsaturated fat (87). Circulating IGF-I was inversely associated with fish intake in a Hawaiian cross-sectional study (105) and with consumption of carbohydrates and bread in a small Greek cross-sectional study (104). Holmes et al. also reported that plasma IGF-I was associated with energy intake among women with normal BMI, protein intake due to dairy foods, and zinc intake regardless of the source (101). Nutritional relations with IGFs were fairly inconsistent although an inverse association between saturated fat and IGFBB-3 was detected in three separate studies. Overall conclusions from these nutritional studies with IGF circulating levels can be difficult because of correlations between nutritional factors and discrepancies in associations with food sources, dietary micronutrients, and supplement sources.

Circulating insulin-like growth factor levels and cancer

Several studies have focused on IGF-I, and to a lesser extent on IGFBP-3 in association with cancer, especially breast cancer. A small U.S. case-control study reported that circulating IGF-I was associated with overall breast cancer risk (106) while two case-control studies nested within the Nurses' Health Study cohort and a New York cohort found positive relations with IGF-I but not IGFBP-3 among premenopausal women only (107, 108). A Chinese population-based case-control study reported

positive associations between breast cancer and both IGF-I and IGFBP-3 levels among women regardless of menopausal status (109), while small U.S. and Italian nested case-control studies and a case-control study with breast cancer cases pooled from three cohorts (Italy, Sweden, U.S.) reported these positive associations among premenopausal women only (110-112). Two case-control studies nested within Dutch and U.S. cohorts of postmenopausal women reported null findings for associations between breast cancer and circulating IGF-I or IGFBP-3 (113, 114). A meta-analysis of 21 studies involving circulating IGFs in association with common cancers reported that elevated IGF-I (OR=1.65; 95% CI: 1.26, 2.08) and IGFBP-3 (OR=1.51; 95% CI: 1.01, 2.27) levels were associated with premenopausal breast cancer (115).

A case-control study nested within the Northern Sweden Health and Disease Cohort reported positive associations between circulating both IGF-I and IGFBP-3 and colon cancer, but inverse associations with rectal cancer (116). Increased IGF-I and reduced IGFBP-3 in circulation were associated with colorectal cancer among women (117) and among men (118) from two case-control studies nested within the Nurses' Health Study and Physicians' Health Study cohorts respectively. However, a Chinese nested case-control study found no relation between serum IGF-I and colorectal cancer in men but instead reported positive associations with circulating IGFBP-2 and IGFBP-3 levels (119). Circulating IGF-I and IGFBP-3 were not related to endometrial cancer in a Swedish case-control study, although there was a suggestion of a positive association with IGFBP-1 among hormone replacement users (120). A case-control study of ovarian cancer nested within three cohorts (Italy, Sweden, U.S.) reported a strong, but imprecise positive association with circulating IGF-I but not IGFBP-3 (121).

In conclusion, there has been some consistency across studies suggesting a positive relation (including postmenopausal breast cancer) between circulating IGF-I and premenopausal breast cancer, but studies of other cancers have conflicting results and are too few in number. Evaluation of cancer study results for relations with IGFBP-3 are inconclusive.

Insulin-like growth factor polymorphisms

The *IGF-I* gene, which is approximately 84.6 kb, is located on chromosome 12q22-q23. It has multiple transcription variants, which explains why it has been reported to have between 4 to 5 exons (122, 123). Based on mapping information from the Entrez Gene website sponsored by the National Center for Biotechnology Information (NCBI), the *IGF-I* gene is shown with four exons. Several studies have focused on cancer development in association with *IGF-I* polymorphisms, while fewer studies have examined variation in the *IGFBP-3* gene.

The focus of epidemiologic studies within the *IGF-I* gene has been on the dinucleotide CA repeat polymorphism (position -969) located in the promoter approximately one kb upstream of the transcription site. The number of *IGF-I* dinucleotide CA repeats within individuals has been reported to typically range from 15 to maximum of 23, although as few as 11 repeats have been reported. Substantial racial variation exists in the frequency of the predominant 19-repeat allele, which is detected in 60 to 70% of Caucasians (94, 124, 125, 126, 127) compared with approximately 40% or less of African Americans (94, 124, 126, 127). Inconsistent relations between the CA repeat polymorphisms and circulating IGF-I levels have been reported. Two studies

reported that having two copies of the 19 CA repeat allele versus other genotypes was inversely associated with plasma IGF-I levels (128), although one found this relation only among oral contraceptive users (94). Two other studies also suggested a relative decrease in plasma IGF-I among oral contraceptive users with at least one 19 CA repeat allele compared with other genotypes (99, 129). Specifically, Jernstrom et al. (94) reported that racial differences in the association between oral contraceptive use and plasma IGF-I were explained by the race-stratified distribution of the *IGF-I* CA repeat polymorphism (19 repeat allele). The Rotterdam Study reported that a decline in circulating IGF-I with advancing age was only measured for persons having two copies of the 19 CA repeat (130). Lai et al. (131) measured slightly lower IGF-I levels in association with increasing number of 19 CA repeat alleles, but only among postmenopausal women. In contrast, a nested case-control study within the Nurses' Health Study cohort and a study of a population-based sample from the Rotterdam cohort reported a positive association between circulating IGF-I and the homozygous 19 CA repeat genotype compared with having no copies of this allele (132, 133), but three studies found no association between the 19 CA repeat allele and circulating IGF-I (127, 134, 135).

A small case-control study (over 50% African American) reported a positive association (OR= 2.87; 95 percent CI: 1.16, 7.06) with breast cancer among women with at least one 19 CA repeat allele relative to women with no 19 CA repeat alleles. In addition, they reported that plasma levels of IGF-I synergistically modified this association (136). The Long Island Breast Cancer Study reported a positive association with breast cancer (OR=3.31; 95% CI: 1.47, 7.48) among premenopausal women carrying at least one allele with less than 19 CA repeats relative to women with at least

one 19 CA repeat allele. There was no independent association with the 19 CA repeat allele, although there was suggested modification by hormonal contraceptives and postmenopausal BMI (125). Various *IGF-I* repeat genotype classifications were evaluated in a nested case-control study within the Nurses' Health Study cohort, but none were clearly associated with breast cancer although weak, nonsignificant effect estimates were reported for some genotypes (132). The Multiethnic Cohort Study also did not provide strong support for involvement of the *IGF-I* 19 CA repeat with breast cancer, even though nonsignificant, positive associations were reported for Latin American and African American women (127). In addition, they evaluated 29 tagging SNPs as well as 35 other SNPs to capture *IGF-I* variation, but found no evidence for involvement with breast cancer (137).

The *IGF-I* CA repeat polymorphism was not associated with colorectal adenomas (135) or colorectal cancer, except for a possible interaction with another insulin-related gene that increased cancer risk (138). However, an inverse association with colorectal cancer was reported for two copies of the 19 CA repeat allele versus all other genotypes among persons who engaged in high vigorous activity (139). A Chinese nested case control study that assessed the CA repeat polymorphism and an *IGF-I* SNP (-533 T/C) also located in the promoter, reported an inverse association between colorectal cancer among persons having two copies of the 21 CA repeat alleles versus other genotypes. In addition, having at least one copy of the variant allele (C) at the -533 C/T SNP decreased the risk of colon cancer but not rectal cancer (140).

IGFBP-3 is a highly conserved gene on chromosome 7p13-p12 with a length of approximately 9 kb and 5 exons. The *IGFBP-3* -202 A/C SNP (rs2854744) located

approximately 200 bp in front of the transcription site in the promoter has the most substantiated evidence for influence on circulating IGFBP-3. Deal et al. (141) noted that the -202 A/C SNP was the most prevalent of the five *IGFBP-3* SNPs examined in Physicians' Health Study participants, and that it was strongly correlated with plasma IGFBP-3. Six studies reported higher circulating IGFBP-3 among women with two copies of the A allele compared with women who had no A alleles and intermediate levels among -202 A/C heterozygotes (99, 131, 142-145). Three studies found no association between the -202 A/C SNP and colorectal cancer (138-140). The Multiethnic Cohort Study examined the -202 A/C SNP and a nonsynonymous SNP (G2133C, rs2854746) in relation to circulating IGFBP-3 and found that only G2133C retained an association after accounting for both SNPs in the model; therefore, G2133C was evaluated with colorectal cancer using a dominant model in which having at least one variant allele resulted in increased (OR=1.32; 95% CI: 1.07, 1.62) colorectal cancer risk (145).

The Nurses' Health Study did not detect an association between the -202 A/C SNP and breast cancer despite reporting that -202 A/C influenced circulating IGFBP-3 (142), and a German case-control study also reported no association of the *IGFBP-3* -336 A/C SNP with breast cancer (146). A Chinese population-based case-control study (143) evaluated five *IGFBP-3* SNPs including -202 A/C (variant allele = C among Chinese) and estimated haplotypes in relation to breast cancer, and they reported a positive association (OR=2.3; 95% CI:1.3, 3.9) with the haplotype including five variant alleles compared to the wildtype haplotype (143) among women under 45 years of age.

A case-control study nested within the European Prospective Investigation into Cancer and Nutrition measured five *IGF-I* SNPs and eight *IGFBP-3* SNPs in relation to breast cancer (144). The homozygous variant genotype (GG) of the *IGF-I* intron 1 (rs2162679) SNP was inversely associated with breast cancer risk but was not associated with circulating IGF-I relative to women with the homozygous wildtype genotype. For the *IGFBP-3* SNPs, -202 A/C (rs2854744) and three other SNPs in the same haplotype block were associated with circulating IGFBP-3. The -202 A/C homozygous genotype (AA) was associated with the highest IGFBP-3 levels; however, there was no consistent evidence for a relation between any of the *IGFBP*-3 SNPs and breast cancer (144).

A recent British population-based case-control study of breast cancer analyzed nine *IGF-I* SNPs and four *IGFBP-3* SNPs in association with circulating IGFs and breast cancer (147). Al-Zahrani et al. also based their conclusions regarding associations between IGF SNPs and their blood levels on the analysis of a middle-aged British cohort of men and women conducted immediately prior to the case-control study. Five *IGF-I* SNPs were associated with circulating IGF-I, and four of these SNPs, along with another *IGF-I* SNP, were associated with breast cancer. After simultaneously accounting for the effects of all of the *IGF-I* SNPs, there was only one SNP (rs1520220) associated with circulating IGF-I and with breast cancer (OR=1.41; 95% CI: 1.11, 1.79) when comparing the homozygous variant genotype (CC) to the homozygous wildtype genotype (GG). Specifically, the rs1520220 homozygous variant genotype (CC) was associated with the highest circulating IGF-I levels, followed by the heterozygous genotype (CG). Three of the four *IGFBP-3* SNPs evaluated were associated with circulating IGFBP-3, and two of these SNPs were associated with breast cancer. The -202 A/C SNP had the highest minor

allele frequency (MAF) of the four SNPs and was positively associated with circulating IGFBP-3 levels such that levels among those with the homozygous variant genotype (AA) were 19% higher than mean IGFBP-3 levels among participants without a copy of the A allele (heterozygous genotype: 9% relative increase in IGFBP-3 levels). Also, a weak protective association (OR=0.87; 95% CI: 0.77, 0.99) was estimated with breast cancer for the AA genotype relative to the homozygous wildtype genotype (CC) (147).

REFERENCES

1. Schwartz SM, Marshall LM, Baird DD. Epidemiologic contributions to understanding the etiology of uterine leiomyomata. Environ Health Perspect 2000;108 Suppl 5:821-7.

2. Farquhar CM, Steiner CA. Hysterectomy rates in the United States 1990-1997. Obstet Gynecol 2002;99:229-34.

3. Wilcox LS, Koonin LM, Pokras R, Strauss LT, Xia Z, Peterson HB. Hysterectomy in the United States, 1988-1990. Obstet Gynecol 1994;83:549-55.

4. Baird DD, Dunson DB, Hill MC, Cousins D, Schectman JM. High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence. Am J Obstet Gynecol 2003;188:100-7.

5. Marshall LM, Spiegelman D, Barbieri RL, et al. Variation in the incidence of uterine leiomyoma among premenopausal women by age and race. Obstet Gynecol 1997;90:967-73.

6. Samadi AR, Lee NC, Flanders WD, Boring JR, 3rd, Parris EB. Risk factors for self-reported uterine fibroids: a case-control study. Am J Public Health 1996;86:858-62.

7. Wise LA, Palmer JR, Stewart EA, Rosenberg L. Age-specific incidence rates for self-reported uterine leiomyomata in the Black Women's Health Study. Obstet Gynecol 2005;105:563-8.

8. Faerstein E, Szklo M, Rosenshein N. Risk factors for uterine leiomyoma: a practice-based case-control study. I. African-American heritage, reproductive history, body size, and smoking. Am J Epidemiol 2001;153:1-10.

9. Kjerulff KH, Langenberg P, Seidman JD, Stolley PD, Guzinski GM. Uterine leiomyomas. Racial differences in severity, symptoms and age at diagnosis. J Reprod Med 1996;41:483-90.

10. Marshall LM, Spiegelman D, Goldman MB, et al. A prospective study of reproductive factors and oral contraceptive use in relation to the risk of uterine leiomyomata. Fertil Steril 1998;70:432-9.

11. Wise LA, Palmer JR, Harlow BL, et al. Reproductive factors, hormonal contraception, and risk of uterine leiomyomata in African-American women: a prospective study. Am J Epidemiol 2004;159:113-23.

12. Baird DD. Invited commentary: uterine leiomyomata-we know so little but could learn so much. Am J Epidemiol 2004;159:124-6.

13. Parazzini F, La Vecchia C, Negri E, Cecchetti G, Fedele L. Epidemiologic characteristics of women with uterine fibroids: a case-control study. Obstet Gynecol 1988;72:853-7.

14. Ross RK, Pike MC, Vessey MP, Bull D, Yeates D, Casagrande JT. Risk factors for uterine fibroids: reduced risk associated with oral contraceptives. Br Med J (Clin Res Ed) 1986;293:359-62.

15. Baird DD, Dunson DB. Why is parity protective for uterine fibroids? Epidemiology 2003;14:247-50.

16. Parazzini F, Negri E, La Vecchia C, Chatenoud L, Ricci E, Guarnerio P. Reproductive factors and risk of uterine fibroids. Epidemiology 1996;7:440-2.

17. Marshall LM, Spiegelman D, Manson JE, et al. Risk of uterine leiomyomata among premenopausal women in relation to body size and cigarette smoking. Epidemiology 1998;9:511-7.

18. Wise LA, Palmer JR, Harlow BL, et al. Risk of uterine leiomyomata in relation to tobacco, alcohol and caffeine consumption in the Black Women's Health Study. Hum Reprod 2004;19:1746-54.

19. Wise LA, Palmer JR, Spiegelman D, et al. Influence of body size and body fat distribution on risk of uterine leiomyomata in U.S. black women. Epidemiology 2005;16:346-54.

20. Bentz AT, Schneider CM, Westerlind KC. The relationship between physical activity and 2-hydroxyestrone, 16alpha-hydroxyestrone, and the 2/16 ratio in premenopausal women (United States). Cancer Causes Control 2005;16:455-61.

21. Campbell KL, Westerlind KC, Harber VJ, Friedenreich CM, Courneya KS. Associations between aerobic fitness and estrogen metabolites in premenopausal women. Med Sci Sports Exerc 2005;37:585-92.

22. Baird DD, Dunson DB, Hill MC, Cousins D, Schectman JM. Association of physical activity with development of uterine leiomyoma. Am J Epidemiol 2007;165:157-63.

23. Frisch RE, Wyshak G, Albright NL, et al. Lower prevalence of breast cancer and cancers of the reproductive system among former college athletes compared to non-athletes. Br J Cancer 1985;52:885-91.

24. D'Aloisio AA, Baird DD. Variation in the association of alcohol intake with uterine fibroids by race and tumor size. Ann Epidemiol 2004;14:622.

25. Chiaffarino F, Parazzini F, La Vecchia C, Marsico S, Surace M, Ricci E. Use of oral contraceptives and uterine fibroids: results from a case-control study. Br J Obstet Gynaecol 1999;106:857-60.

26. Rein MS. Advances in uterine leiomyoma research: the progesterone hypothesis. Environ Health Perspect 2000;108 Suppl 5:791-3.

27. Rein MS, Barbieri RL, Friedman AJ. Progesterone: a critical role in the pathogenesis of uterine myomas. Am J Obstet Gynecol 1995;172:14-8.

28. Sozen I, Arici A. Interactions of cytokines, growth factors, and the extracellular matrix in the cellular biology of uterine leiomyomata. Fertil Steril 2002;78:1-12.

29. Sozen I, Arici A. Cellular biology of myomas: interaction of sex steroids with cytokines and growth factors. Obstet Gynecol Clin North Am 2006;33:41-58.

30. Hocutt JE, Jr. Uterine fibroids and hypertension. Del Med J 1979;51:697-9.

31. Everett HS, Scott RB. The possible etiologic role of gynecologic lesions in the production of hypertension. Amer J Obstet Gynecol 1942;44:1010-25.

32. Luoto R, Rutanen EM, Auvinen A. Fibroids and hypertension. A cross-sectional study of women undergoing hysterectomy. J Reprod Med 2001;46:359-64.

33. Faerstein E, Szklo M, Rosenshein NB. Risk factors for uterine leiomyoma: a practice-based case-control study. II. Atherogenic risk factors and potential sources of uterine irritation. Am J Epidemiol 2001;153:11-9.

34. Moss NS, Benditt EP. Human atherosclerotic plaque cells and leiomyoma cells. Comparison of in vitro growth characteristics. Am J Pathol 1975;78:175-90.

35. Al-Hendy A, Salama SA. Catechol-O-methyltransferase polymorphism is associated with increased uterine leiomyoma risk in different ethnic groups. J Soc Gynecol Investig 2006;13:136-44.

36. Al-Hendy A, Salama SA. Ethnic distribution of estrogen receptor-alpha polymorphism is associated with a higher prevalence of uterine leiomyomas in black Americans. Fertil Steril 2006.

37. Gooden KM, Schroeder JC, North KE, et al. Val153Met polymorphism of catechol-O-methyltransferase and prevalence of uterine leiomyomata. Reprod Sci 2007;14:117-20.

38. Hsieh YY, Chan IP, Wang HI, Chang CC, Huang CW, Lin CS. PROGINS Alu sequence insertion is associated with hyperprolactinaemia but not leiomyoma susceptibility. Clin Endocrinol (Oxf) 2005;62:492-7.

39. Hsieh YY, Chang CC, Tsai FJ, Lin CC, Tsai CH. T homozygote and allele of epidermal growth factor receptor 2073 gene polymorphism are associated with higher susceptibility to endometriosis and leiomyomas. Fertil Steril 2005;83:796-9.

40. Hsieh YY, Chang CC, Tsai FJ, Lin CC, Yeh LS, Tsai CH. Tumor necrosis factor-alpha-308 promoter and p53 codon 72 gene polymorphisms in women with leiomyomas. Fertil Steril 2004;82 Suppl 3:1177-81.

41. Hsieh YY, Chang CC, Tsai FJ, Lin CC, Yeh LS, Peng CT. Androgen receptor trinucleotide polymorphism in leiomyoma. J Assist Reprod Genet 2004;21:453-7.

42. Hsieh YY, Chang CC, Tsai FJ, et al. Estrogen receptor thymine-adenine dinucleotide repeat polymorphism is associated with susceptibility to leiomyoma. Fertil Steril 2003;79:96-9.

43. Jeon YT, Kim JW, Park NH, Song YS, Kang SB, Lee HP. DNA repair gene XRCC1 Arg399Gln polymorphism is associated with increased risk of uterine leiomyoma. Hum Reprod 2005;20:1586-9.

44. Amant F, Dorfling CM, de Brabanter J, et al. A possible role of the cytochrome P450c17alpha gene (CYP17) polymorphism in the pathobiology of uterine leiomyomas from black South African women: a pilot study. Acta Obstet Gynecol Scand 2004;83:234-9.

45. Denschlag D, Bentz EK, Hefler L, et al. Genotype distribution of estrogen receptor-alpha, catechol-O-methyltransferase, and cytochrome P450 17 gene polymorphisms in Caucasian women with uterine leiomyomas. Fertil Steril 2006;85:462-7.

46. Denschlag D, Bettendorf H, Watermann D, Keck C, Tempfer C, Pietrowski D. Polymorphism of the p53 tumor suppressor gene is associated with susceptibility to uterine leiomyoma. Fertil Steril 2005;84:162-6.

47. Herr D, Bettendorf H, Denschlag D, Keck C, Pietrowski D. Cytochrome P2A13 and P1A1 gene polymorphisms are associated with the occurrence of uterine leiomyoma. Arch Gynecol Obstet 2006.

48. Tsujino T, Ohara N, Yoshida S, et al. The CYP17 MspA1 polymorphism is not associated with an increased risk of uterine leiomyomas in a Japanese population. Gynecol Endocrinol 2006;22:87-91.

49. Takemura N, Yoshida S, Kennedy S, Deguchi M, Ohara N, Maruo T. Matrix metalloproteinase-1 and -9 promoter polymorphisms are not associated with an increased risk of uterine leiomyomas in a Japanese population. J Soc Gynecol Investig 2006;13:232-6.

50. Jones JI, Clemmons DR. Insulin-like growth factors and their binding proteins: biological actions. Endocr Rev 1995;16:3-34.

51. Hwa V, Oh Y, Rosenfeld RG. The insulin-like growth factor-binding protein (IGFBP) superfamily. Endocr Rev 1999;20:761-87.

52. Juul A, Main K, Blum WF, Lindholm J, Ranke MB, Skakkebaek NE. The ratio between serum levels of insulin-like growth factor (IGF)-I and the IGF binding proteins (IGFBP-1, 2 and 3) decreases with age in healthy adults and is increased in acromegalic patients. Clin Endocrinol (Oxf) 1994;41:85-93.

53. Gill ZP, Perks CM, Newcomb PV, Holly JM. Insulin-like growth factor-binding protein (IGFBP-3) predisposes breast cancer cells to programmed cell death in a non-IGF-dependent manner. J Biol Chem 1997;272:25602-7.

54. Schedlich LJ, Graham LD. Role of insulin-like growth factor binding protein-3 in breast cancer cell growth. Microsc Res Tech 2002;59:12-22.

55. Strawn EY, Jr., Novy MJ, Burry KA, Bethea CL. Insulin-like growth factor I promotes leiomyoma cell growth in vitro. Am J Obstet Gynecol 1995;172:1837-43; discussion 1843-4.

56. Wolanska M, Bankowski E. An accumulation of insulin-like growth factor I (IGF-I) in human myometrium and uterine leiomyomas in various stages of tumour growth. Eur Cytokine Netw 2004;15:359-63.

57. Van der Ven LT, Roholl PJ, Gloudemans T, et al. Expression of insulin-like growth factors (IGFs), their receptors and IGF binding protein-3 in normal, benign and malignant smooth muscle tissues. Br J Cancer 1997;75:1631-40.

58. Van der Ven LT, Gloudemans T, Roholl PJ, et al. Growth advantage of human leiomyoma cells compared to normal smooth-muscle cells due to enhanced sensitivity toward insulin-like growth factor I. Int J Cancer 1994;59:427-34.

59. Dixon D, He H, Haseman JK. Immunohistochemical localization of growth factors and their receptors in uterine leiomyomas and matched myometrium. Environ Health Perspect 2000;108 Suppl 5:795-802.

60. Chandrasekhar Y, Heiner J, Osuamkpe C, Nagamani M. Insulin-like growth factor I and II binding in human myometrium and leiomyomas. Am J Obstet Gynecol 1992;166:64-9.

61. Di Lieto A, Iannotti F, De Falco M, et al. Immunohistochemical detection of insulin-like growth factor type I receptor and uterine volume changes in gonadotropin-releasing hormone analog-treated uterine leiomyomas. Am J Obstet Gynecol 2003;188:702-6.

62. Wu X, Wang H, Englund K, Blanck A, Lindblom B, Sahlin L. Expression of progesterone receptors A and B and insulin-like growth factor-I in human myometrium and fibroids after treatment with a gonadotropin-releasing hormone analogue. Fertil Steril 2002;78:985-93.

63. Giudice LC, Irwin JC, Dsupin BA, et al. Insulin-like growth factor (IGF), IGF binding protein (IGFBP), and IGF receptor gene expression and IGFBP synthesis in human uterine leiomyomata. Hum Reprod 1993;8:1796-806.

64. Vollenhoven BJ, Herington AC, Healy DL. Messenger ribonucleic acid expression of the insulin-like growth factors and their binding proteins in uterine fibroids and myometrium. J Clin Endocrinol Metab 1993;76:1106-10.

65. Boehm KD, Daimon M, Gorodeski IG, Sheean LA, Utian WH, Ilan J. Expression of the insulin-like and platelet-derived growth factor genes in human uterine tissues. Mol Reprod Dev 1990;27:93-101.

66. Gloudemans T, Prinsen I, Van Unnik JA, Lips CJ, Den Otter W, Sussenbach JS. Insulin-like growth factor gene expression in human smooth muscle tumors. Cancer Res 1990;50:6689-95.

67. Arslan AA, Gold LI, Mittal K, et al. Gene expression studies provide clues to the pathogenesis of uterine leiomyoma: new evidence and a systematic review. Hum Reprod 2005;20:852-63.

68. Skubitz KM, Skubitz AP. Differential gene expression in uterine leiomyoma. J Lab Clin Med 2003;141:297-308.

69. Quade BJ, Wang TY, Sornberger K, Dal Cin P, Mutter GL, Morton CC. Molecular pathogenesis of uterine smooth muscle tumors from transcriptional profiling. Genes Chromosomes Cancer 2004;40:97-108.

70. Hoffman PJ, Milliken DB, Gregg LC, Davis RR, Gregg JP. Molecular characterization of uterine fibroids and its implication for underlying mechanisms of pathogenesis. Fertil Steril 2004;82:639-49.

71. Wang H, Mahadevappa M, Yamamoto K, et al. Distinctive proliferative phase differences in gene expression in human myometrium and leiomyomata. Fertil Steril 2003;80:266-76.

72. Catherino W, Salama A, Potlog-Nahari C, Leppert P, Tsibris J, Segars J. Gene expression studies in leiomyomata: new directions for research. Semin Reprod Med 2004;22:83-90.

73. Tsibris JC, Segars J, Coppola D, et al. Insights from gene arrays on the development and growth regulation of uterine leiomyomata. Fertil Steril 2002;78:114-21.

74. Ahn WS, Kim KW, Bae SM, et al. Targeted cellular process profiling approach for uterine leiomyoma using cDNA microarray, proteomics and gene ontology analysis. Int J Exp Pathol 2003;84:267-79.

75. Catherino WH, Segars JH. Microarray analysis in fibroids: which gene list is the correct list? Fertil Steril 2003;80:293-4.

76. Swartz CD, Afshari CA, Yu L, Hall KE, Dixon D. Estrogen-induced changes in IGF-I, Myb family and MAP kinase pathway genes in human uterine leiomyoma and normal uterine smooth muscle cell lines. Mol Hum Reprod 2005;11:441-50.

77. Englund K, Lindblom B, Carlstrom K, Gustavsson I, Sjoblom P, Blanck A. Gene expression and tissue concentrations of IGF-I in human myometrium and fibroids under different hormonal conditions. Mol Hum Reprod 2000;6:915-20.

78. Gao Z, Matsuo H, Wang Y, Nakago S, Maruo T. Up-regulation by IGF-I of proliferating cell nuclear antigen and Bcl-2 protein expression in human uterine leiomyoma cells. J Clin Endocrinol Metab 2001;86:5593-9.

79. Yamada T, Nakago S, Kurachi O, et al. Progesterone down-regulates insulin-like growth factor-I expression in cultured human uterine leiomyoma cells. Hum Reprod 2004;19:815-21.

80. Harrela M, Koistinen H, Kaprio J, et al. Genetic and environmental components of interindividual variation in circulating levels of IGF-I, IGF-II, IGFBP-1, and IGFBP-3. J Clin Invest 1996;98:2612-5.

81. Pantsulaia I, Trofimov S, Kobyliansky E, Livshits G. Genetic regulation of the variation of circulating insulin-like growth factors and leptin in human pedigrees. Metabolism 2005;54:975-81.

82. Pantsulaia I, Trofimov S, Kobyliansky E, Livshits G. Pedigree-based quantitative genetic analysis of interindividual variation in circulating levels of IGFBP-3. J Bone Miner Metab 2002;20:156-63.

83. LeRoith D. Insulin-like growth factors. N Engl J Med 1997;336:633-40.

84. Kaklamani VG, Linos A, Kaklamani E, Markaki I, Mantzoros C. Age, sex, and smoking are predictors of circulating insulin-like growth factor 1 and insulin-like growth factor-binding protein 3. J Clin Oncol 1999;17:813-7.

85. Holmes MD, Pollak MN, Hankinson SE. Lifestyle correlates of plasma insulin-like growth factor I and insulin-like growth factor binding protein 3 concentrations. Cancer Epidemiol Biomarkers Prev 2002;11:862-7.

86. Lukanova A, Toniolo P, Akhmedkhanov A, et al. A cross-sectional study of IGF-I determinants in women. Eur J Cancer Prev 2001;10:443-52.

87. Probst-Hensch NM, Wang H, Goh VH, Seow A, Lee HP, Yu MC. Determinants of circulating insulin-like growth factor I and insulin-like growth factor binding protein 3 concentrations in a cohort of Singapore men and women. Cancer Epidemiol Biomarkers Prev 2003;12:739-46.

88. DeLellis K, Rinaldi S, Kaaks RJ, Kolonel LN, Henderson B, Le Marchand L. Dietary and lifestyle correlates of plasma insulin-like growth factor-I (IGF-I) and IGF binding protein-3 (IGFBP-3): the multiethnic cohort. Cancer Epidemiol Biomarkers Prev 2004;13:1444-51.

89. Harris TB, Kiel D, Roubenoff R, et al. Association of insulin-like growth factor-I with body composition, weight history, and past health behaviors in the very old: the Framingham Heart Study. J Am Geriatr Soc 1997;45:133-9.

90. Goodman-Gruen D, Barrett-Connor E. Epidemiology of insulin-like growth factor-I in elderly men and women. The Rancho Bernardo Study. Am J Epidemiol 1997;145:970-6.

91. Johansson H, Baglietto L, Guerrieri-Gonzaga A, et al. Factors associated with circulating levels of insulin-like growth factor-I and insulin-like growth factor binding protein-3 in 740 women at risk for breast cancer. Breast Cancer Res Treat 2004;88:63-73.

92. Morimoto LM, Newcomb PA, White E, Bigler J, Potter JD. Variation in plasma insulin-like growth factor-1 and insulin-like growth factor binding protein-3: personal and lifestyle factors (United States). Cancer Causes Control 2005;16:917-27.

93. Chang S, Wu X, Yu H, Spitz MR. Plasma concentrations of insulin-like growth factors among healthy adult men and postmenopausal women: associations with body composition, lifestyle, and reproductive factors. Cancer Epidemiol Biomarkers Prev 2002;11:758-66.

94. Jernstrom H, Chu W, Vesprini D, et al. Genetic factors related to racial variation in plasma levels of insulin-like growth factor-1: implications for premenopausal breast cancer risk. Mol Genet Metab 2001;72:144-54.

95. Wong WW, Copeland KC, Hergenroeder AC, Hill RB, Stuff JE, Ellis KJ. Serum concentrations of insulin, insulin-like growth factor-I and insulin-like growth

factor binding proteins are different between white and African American girls. J Pediatr 1999;135:296-300.

96. Lukanova A, Lundin E, Zeleniuch-Jacquotte A, et al. Body mass index, circulating levels of sex-steroid hormones, IGF-I and IGF-binding protein-3: a cross-sectional study in healthy women. Eur J Endocrinol 2004;150:161-71.

97. Slattery ML, Baumgartner KB, Byers T, et al. Genetic, anthropometric, and lifestyle factors associated with IGF-1 and IGFBP-3 levels in Hispanic and non-Hispanic white women. Cancer Causes Control 2005;16:1147-57.

98. Voskuil DW, Bueno de Mesquita HB, Kaaks R, et al. Determinants of circulating insulin-like growth factor (IGF)-I and IGF binding proteins 1-3 in premenopausal women: physical activity and anthropometry (Netherlands). Cancer Causes Control 2001;12:951-8.

99. Jernstrom H, Deal C, Wilkin F, et al. Genetic and nongenetic factors associated with variation of plasma levels of insulin-like growth factor-I and insulin-like growth factor-binding protein-3 in healthy premenopausal women. Cancer Epidemiol Biomarkers Prev 2001;10:377-84.

100. Vrieling A, Voskuil DW, Bueno de Mesquita HB, et al. Dietary determinants of circulating insulin-like growth factor (IGF)-I and IGF binding proteins 1, -2 and -3 in women in the Netherlands. Cancer Causes Control 2004;15:787-96.

101. Holmes MD, Pollak MN, Willett WC, Hankinson SE. Dietary correlates of plasma insulin-like growth factor I and insulin-like growth factor binding protein 3 concentrations. Cancer Epidemiol Biomarkers Prev 2002;11:852-61.

102. Helle SI, Anker GB, Meadows KA, Holly JM, Lonning PE. Alterations in the insulin-like growth factor system during the menstrual cycle in normal women. Maturitas 1998;28:259-65.

103. Wang HS, Lee JD, Soong YK. Serum levels of insulin-like growth factor I and insulin-like growth factor-binding protein-1 and -3 in women with regular menstrual cycles. Fertil Steril 1995;63:1204-9.

104. Kaklamani VG, Linos A, Kaklamani E, Markaki I, Koumantaki Y, Mantzoros CS. Dietary fat and carbohydrates are independently associated with circulating insulin-like growth factor 1 and insulin-like growth factor-binding protein 3 concentrations in healthy adults. J Clin Oncol 1999;17:3291-8.

105. Maskarinec G, Takata Y, Kaaks R. The relation between nutritional factors and insulin-like growth factor-I in premenopausal women of different ethnicity. Eur J Nutr 2005;44:105-13.

106. Li BD, Khosravi MJ, Berkel HJ, et al. Free insulin-like growth factor-I and breast cancer risk. Int J Cancer 2001;91:736-9.

107. Toniolo P, Bruning PF, Akhmedkhanov A, et al. Serum insulin-like growth factor-I and breast cancer. Int J Cancer 2000;88:828-32.

108. Hankinson SE, Willett WC, Colditz GA, et al. Circulating concentrations of insulin-like growth factor-I and risk of breast cancer. Lancet 1998;351:1393-6.

109. Yu H, Jin F, Shu XO, et al. Insulin-like growth factors and breast cancer risk in Chinese women. Cancer Epidemiol Biomarkers Prev 2002;11:705-12.

110. Krajcik RA, Borofsky ND, Massardo S, Orentreich N. Insulin-like growth factor I (IGF-I), IGF-binding proteins, and breast cancer. Cancer Epidemiol Biomarkers Prev 2002;11:1566-73.

111. Rinaldi S, Toniolo P, Muti P, et al. IGF-I, IGFBP-3 and breast cancer in young women: a pooled re-analysis of three prospective studies. Eur J Cancer Prev 2005;14:493-6.

112. Muti P, Quattrin T, Grant BJ, et al. Fasting glucose is a risk factor for breast cancer: a prospective study. Cancer Epidemiol Biomarkers Prev 2002;11:1361-8.

113. Keinan-Boker L, Bueno De Mesquita HB, Kaaks R, et al. Circulating levels of insulin-like growth factor I, its binding proteins -1,-2, -3, C-peptide and risk of postmenopausal breast cancer. Int J Cancer 2003;106:90-5.

114. Rollison DE, Newschaffer CJ, Tao Y, Pollak M, Helzlsouer KJ. Premenopausal levels of circulating insulin-like growth factor I and the risk of postmenopausal breast cancer. Int J Cancer 2006;118:1279-84.

115. Renehan AG, Zwahlen M, Minder C, O'Dwyer ST, Shalet SM, Egger M. Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis. Lancet 2004;363:1346-53.

116. Palmqvist R, Hallmans G, Rinaldi S, et al. Plasma insulin-like growth factor 1, insulin-like growth factor binding protein 3, and risk of colorectal cancer: a prospective study in northern Sweden. Gut 2002;50:642-6.

117. Giovannucci E, Pollak MN, Platz EA, et al. A prospective study of plasma insulin-like growth factor-1 and binding protein-3 and risk of colorectal neoplasia in women. Cancer Epidemiol Biomarkers Prev 2000;9:345-9.

118. Ma J, Pollak MN, Giovannucci E, et al. Prospective study of colorectal cancer risk in men and plasma levels of insulin-like growth factor (IGF)-I and IGF-binding protein-3. J Natl Cancer Inst 1999;91:620-5.

119. Probst-Hensch NM, Yuan JM, Stanczyk FZ, Gao YT, Ross RK, Yu MC. IGF-1, IGF-2 and IGFBP-3 in prediagnostic serum: association with colorectal cancer in a cohort of Chinese men in Shanghai. Br J Cancer 2001;85:1695-9.

120. Weiderpass E, Brismar K, Bellocco R, Vainio H, Kaaks R. Serum levels of insulin-like growth factor-I, IGF-binding protein 1 and 3, and insulin and endometrial cancer risk. Br J Cancer 2003;89:1697-704.

121. Lukanova A, Lundin E, Toniolo P, et al. Circulating levels of insulin-like growth factor-I and risk of ovarian cancer. Int J Cancer 2002;101:549-54.

122. Rotwein P. Structure, evolution, expression and regulation of insulin-like growth factors I and II. Growth Factors 1991;5:3-18.

123. Jansen E, Steenbergh PH, LeRoith D, Roberts CT, Jr., Sussenbach JS. Identification of multiple transcription start sites in the human insulin-like growth factor-I gene. Mol Cell Endocrinol 1991;78:115-25.

124. Takacs I, Koller DL, Peacock M, et al. Sibling pair linkage and association studies between bone mineral density and the insulin-like growth factor I gene locus. J Clin Endocrinol Metab 1999;84:4467-71.

125. Cleveland RJ, Gammon MD, Edmiston SN, et al. IGF1 CA repeat polymorphisms, lifestyle factors and breast cancer risk in the Long Island Breast Cancer Study Project. Carcinogenesis 2006;27:758-65.

126. Schildkraut JM, Demark-Wahnefried W, Wenham RM, et al. IGF1 (CA)19 repeat and IGFBP3 -202 A/C genotypes and the risk of prostate cancer in Black and White men. Cancer Epidemiol Biomarkers Prev 2005;14:403-8.

127. DeLellis K, Ingles S, Kolonel L, et al. IGF1 genotype, mean plasma level and breast cancer risk in the Hawaii/Los Angeles multiethnic cohort. Br J Cancer 2003;88:277-82.

128. Rosen CJ, Kurland ES, Vereault D, et al. Association between serum insulin growth factor-I (IGF-I) and a simple sequence repeat in IGF-I gene: implications for genetic studies of bone mineral density. J Clin Endocrinol Metab 1998;83:2286-90.

129. Jernstrom H, Sandberg T, Bageman E, Borg A, Olsson H. Insulin-like growth factor-1 (IGF1) genotype predicts breast volume after pregnancy and hormonal contraception and is associated with circulating IGF-1 levels: implications for risk of early-onset breast cancer in young women from hereditary breast cancer families. Br J Cancer 2005;92:857-66.

130. Rietveld I, Janssen JA, Hofman A, Pols HA, van Duijn CM, Lamberts SW. A polymorphism in the IGF-I gene influences the age-related decline in circulating total IGF-I levels. Eur J Endocrinol 2003;148:171-5.

131. Lai JH, Vesprini D, Zhang W, Yaffe MJ, Pollak M, Narod SA. A polymorphic locus in the promoter region of the IGFBP3 gene is related to mammographic breast density. Cancer Epidemiol Biomarkers Prev 2004;13:573-82.

132. Missmer SA, Haiman CA, Hunter DJ, et al. A sequence repeat in the insulin-like growth factor-1 gene and risk of breast cancer. Int J Cancer 2002;100:332-6.

133. Vaessen N, Heutink P, Janssen JA, et al. A polymorphism in the gene for IGF-I: functional properties and risk for type 2 diabetes and myocardial infarction. Diabetes 2001;50:637-42.

134. Wen W, Gao YT, Shu XO, et al. Insulin-like growth factor-I gene polymorphism and breast cancer risk in Chinese women. Int J Cancer 2005;113:307-11.

135. Giovannucci E, Haiman CA, Platz EA, Hankinson SE, Pollak MN, Hunter DJ. Dinucleotide repeat in the insulin-like growth factor-I gene is not related to risk of colorectal adenoma. Cancer Epidemiol Biomarkers Prev 2002;11:1509-10.

136. Yu H, Li BD, Smith M, Shi R, Berkel HJ, Kato I. Polymorphic CA repeats in the IGF-I gene and breast cancer. Breast Cancer Res Treat 2001;70:117-22.

137. Setiawan VW, Cheng I, Stram DO, et al. Igf-I genetic variation and breast cancer: the multiethnic cohort. Cancer Epidemiol Biomarkers Prev 2006;15:172-4.

138. Slattery ML, Samowitz W, Curtin K, et al. Associations among IRS1, IRS2, IGF1, and IGFBP3 genetic polymorphisms and colorectal cancer. Cancer Epidemiol Biomarkers Prev 2004;13:1206-14.

139. Slattery ML, Murtaugh M, Caan B, Ma KN, Neuhausen S, Samowitz W. Energy balance, insulin-related genes and risk of colon and rectal cancer. Int J Cancer 2005;115:148-54.

140. Wong HL, Delellis K, Probst-Hensch N, et al. A new single nucleotide polymorphism in the insulin-like growth factor I regulatory region associates with colorectal cancer risk in singapore chinese. Cancer Epidemiol Biomarkers Prev 2005;14:144-51.

141. Deal C, Ma J, Wilkin F, et al. Novel promoter polymorphism in insulin-like growth factor-binding protein-3: correlation with serum levels and interaction with known regulators. J Clin Endocrinol Metab 2001;86:1274-80.

142. Schernhammer ES, Hankinson SE, Hunter DJ, Blouin MJ, Pollak MN. Polymorphic variation at the -202 locus in IGFBP3: Influence on serum levels of insulin-like growth factors, interaction with plasma retinol and vitamin D and breast cancer risk. Int J Cancer 2003;107:60-4.

143. Ren Z, Cai Q, Shu XO, et al. Genetic polymorphisms in the IGFBP3 gene: association with breast cancer risk and blood IGFBP-3 protein levels among Chinese women. Cancer Epidemiol Biomarkers Prev 2004;13:1290-5.

144. Canzian F, McKay JD, Cleveland RJ, et al. Polymorphisms of genes coding for insulin-like growth factor 1 and its major binding proteins, circulating levels of IGF-I and IGFBP-3 and breast cancer risk: results from the EPIC study. Br J Cancer 2006;94:299-307.

145. Le Marchand L, Kolonel LN, Henderson BE, Wilkens LR. Association of an exon 1 polymorphism in the IGFBP3 gene with circulating IGFBP-3 levels and colorectal cancer risk: the multiethnic cohort study. Cancer Epidemiol Biomarkers Prev 2005;14:1319-21.

146. Rohrbacher M, Risch A, Kropp S, Chang-Claude J. The A(-336C) insulin-like growth factor binding protein-3 promoter polymorphism is not a modulator of breast cancer risk in Caucasian women. Cancer Epidemiol Biomarkers Prev 2005;14:289-90.

147. Al-Zahrani A, Sandhu MS, Luben RN, et al. IGF1 and IGFBP3 tagging polymorphisms are associated with circulating levels of IGF1, IGFBP3 and risk of breast cancer. Hum Mol Genet 2006;15:1-10.

CHAPTER II

STATEMENT OF SPECIFIC AIMS

Study questions

The purpose of this dissertation is to examine whether variation in two genes, *IGF-I* and *IGFBP-3*, as studied by specific SNPs and estimated diplotypes (paired haplotypes) is associated with: 1) circulating IGF-I and IGFBP-3 levels and 2) prevalence of uterine leiomyomata (fibroids). Because of biological interrelations between IGF-I and IGFBP-3, variants in each gene (*IGF-I* and *IGFBP-3*) were evaluated in association with circulating levels of both proteins. We hypothesized that *IGF-I* and *IGFBP-3* variants would be associated with prevalence of fibroids based on evidence of IGF-I involvement with fibroid pathogenesis from gene expression studies. Specifically, we conducted two separate analyses to evaluated relations with prevalence of fibroids defined by 1) any size and 2) at least 2 cm in diameter), which the latter may be more clinically relevant. We hypothesized that IGF-I, which acts as a potent mitogen that is structurally similar to insulin, would promote fibroid growth rather than their incidence. Therefore, *IGF-I* and *IGFBP-3* variant associations would be stronger with prevalence of larger fibroids rather than with fibroids of any size. Given that there are substantial allele frequency differences by race, analyses of *IGF-I* and *IGFBP-3* variants will be performed separately

within African Americans and Caucasians. The results of this study could help identify factors that contribute to the racial disparity in the prevalence and size of these hormonally-dependent benign tumors.

Primary aims

1) To describe the prevalence of SNPs and estimated diplotypes for *IGF-I* and *IGFBP-3* among African Americans and Caucasians.

2) To estimate associations between *IGF-I* and *IGFBP-3* gene variants (diplotypes and individual polymorphisms) and plasma IGF-I and IGFBP-3 levels among African Americans and Caucasian women.

3) To estimate association in genetic polymorphisms and estimated diplotypes for *IGF-I* and *IGFBP-3* in relation to the prevalence of uterine fibroids (any size and 2+ cm in diameter) among African American and Caucasian women.

CHAPTER III

METHODS

Study population

From 1996 to 1999, a random sample of 2,384 George Washington University health plan members who were 35 to 49 years of age was contacted for potential enrollment into the NIEHS Uterine Fibroid Study (1-3). The study was approved by Institutional Review Boards at NIEHS and George Washington University. Approximately 50% of the health plan population was thought to be African American. Of the approximately 30,000 health plan members, most were enrolled through their employers, except for about 1,000 Medicaid patients. The parent Uterine Fibroid Study was conducted to ascertain the prevalence of uterine fibroids within premenopausal women and to estimate the age-specific incidence. A second aim was to investigate the etiologic factors for fibroids, including genetic variants that could increase the likelihood of their development. Because these benign tumors can develop and continue to grow until menopause, the parent study population was restricted to a narrow age range including the late premenopausal years, in order to approximate the lifetime risk of developing uterine fibroids (1).

Almost 90% of the random sample (2,102 women) was contacted by telephone and consented to eligibility screening for the parent study. The study criteria were satisfied by 1,786 women. Approximately 70% of excluded women were ineligible because they

no longer attended the Washington DC clinic where fibroid screening would be performed. Others were deemed ineligible because they were misidentified as female, had no further access to a telephone, were not within the required age range, or did not speak English. Approximately 20% of eligible women refused participation in the study, resulting in 1,430 participants (2).

The prevalence of uterine fibroids and the distribution of IGF gene variants differ between racial groups; therefore, to facilitate race-specific analyses, we restricted the current study population to African American and Caucasian women. We also excluded postmenopausal women from the current investigation because they did not have the clinic visit to draw blood and screen for fibroids since these tumors can regress following menopause. Race and menopausal status criteria were met by 1,146 women from the parent study. Based on availability of DNA samples for genotyping, we further limited the study population to 984 women (582 African Americans and 402 Caucasians). The study consent form specified use of specimens for genetic polymorphism analyses.

Preliminary data

The distribution of study characteristics stratified by race is displayed in Table 3.1. The prevalence of fibroids was 72% among African Americans and 50% among Caucasians. A greater proportion of African Americans (24%) had large fibroids (≥4 cm) in comparison to Caucasians (11%). African Americans (11%) were more likely than Caucasians (4%) to experience menarche at less than 11 years of age. The majority of women did not have births after 24 years of age (51% of African Americans, 63% of Caucasians). African Americans (60%) reported a greater use of oral contraceptives

during teenage years than Caucasians (33%), but few women in either group were taking oral contraceptives when they entered the parent study. There were more African Americans classified as overweight or obese than Caucasians, and there were fewer African Americans with high levels of physical activity than Caucasians. In addition, a greater proportion of African Americans (30%) reported current smoking than Caucasians (8%). However, fewer African Americans (42%) consumed alcohol on a regular basis (at least 0.5 drinks per week in the past year) compared to Caucasians (78%).

Data collection

Uterine fibroids

Since women can have asymptomatic fibroids, a primary goal of the NIEHS Uterine Fibroid Study was to provide a better estimate of fibroid prevalence among premenopausal women than previous studies based on clinical diagnosis or treatment by hysterectomy alone. Therefore, transabdominal and transvaginal ultrasound examinations were performed to screen for fibroids. Ultrasound examinations were considered the gold standard for measurement of uterine fibroids (4) since very high sensitivity (0.99) and specificity (0.91) have been reported in comparison to pathologic verification in hysterectomy specimens (5). Criteria used for ultrasound identification of fibroids in this study were based on Muram, et al. (6) with modifications to reflect technological changes that have decreased the limits of detection (7), so that tumors of 0.5 cm diameter can be detected.

Both transabdominal and transvaginal ultrasounds were performed at a clinic visit that occurred within three months of study entry. Trained sonographers endorsed by the American Registry of Diagnostic Medical Sonographers performed these examinations under the supervision of a radiologist who verified their assessments (1). Transabdominal ultrasounds for evaluating the superior portion of the uterus were performed using 3.5 to 5.0 MHz probes, and transvaginal examinations were conducted using 5.0 to 7.0 MHz probes (1) with ATL HDI 9 (ATL, Bothell, WA), Acuson 128 XP (Siemens, Issaquah, WA), or Diasonics DRF 400 (GE, Milwaukee, WI) ultrasound systems. Information collected by sonographers included the size and location of the two largest fibroids and the number of tumors within the uterus (1).

Study sonograms were not performed for women who had a recent ultrasound at the clinic that conducted study sonograms. Instead, medical records were abstracted to obtain fibroid information for these women, which represent approximately 20% (N = 204) of the current study sample (African American and Caucasian premenopausal women with extracted DNA from blood samples). In addition, 19 women in our study sample were classified with regard to the size and presence of fibroids based on self-report of previous diagnoses only. Fibroid status was not assessed in seven African Americans and eight Caucasians in our study sample; however, these women were genotyped and included in current study for analyses with circulating IGF-I and IGFBP-3 levels.

Sample collection and assays

Approximately 50 cc of blood were obtained from the premenopausal women in the parent study who attended the clinic visit. Blood samples were collected through

venipuncture after an 8-hour fast and processed for plasma. Plasma aliquots were stored at -80°C until analysis.

Plasma was analyzed for IGF-I and IGFBP-3 as nanograms per milliliter (ng/mL) at NIEHS. Quality control for both IGF-I and IGFBP-3 measurements was performed by verifying that standard curve parameters and controls were reproduced for each assay.

IGF-I was measured by radioimmunoassay using a kit with instructions from Nichols Institute Diagnostics (San Juan Capistrano, CA). The first step involves extraction of IGF-I from binding proteins through octadecasilyl-silica cartridges (C18 Sep-Pak, Waters & Associates, Milford, MA). Prior to extraction, approximately 200 μL of EDTA plasma is divided into 4 tubes of 50 μL of EDTA plasma, and then each tube is acidified with 950 μL of 0.5 N HCL. The C18 Sep-Pak column is pretreated with 5 mL isopropyl alcohol followed by 5 mL methanol and 10 mL 4% glacial acetic acid/HPLC H_20. The acidified EDTA plasma is then loaded onto the prewashed column, and it is sequentially washed with 10 mL 4% glacial acetic acid/HPLC H_20 and 2 mL methanol to extract IGF-I. The extracted IGF-I samples are dried in nitrogen gas within a 37°C H_20 bath, and the radioimmunoassay is then performed. This extraction technique is very reproducible and is rapidly completed.

For the radioimmunoassay, standards for IGF-I range from 0.3 to 4.8 ng/mL, and two human EDTA plasma controls with specified IGF-I concentrations were provided by the manufacturer for validation. IGF-I labeled with ^{125}I and antiserum for IGF-I from rabbits were used for detection in the assays. Cross-reactivity of antiserum to IGF-I with IGF-II has been reported to be <0.5%. Intra-assay and inter-assay variances for this procedure

are 2.3% and 11.4% respectively using samples averaging approximately 1 ng/mL of IGF-I. The reported detection limit of this technique is 0.06 ng/mL.

IGFBP-3 was measured by immunoradiometric assay using a kit with instructions from Diagnostic Systems Laboratories, Inc. (Webster, TX). Prior to the assay, a 1:100 dilution of 10 μL EDTA plasma with a 1 mL 0 ng/mL IGFBP-3 standard consisting of bovine serum albumin (BSA) buffer with sodium azide is performed. IGFBP-3 standards ranging from 2.0 to 100.0 ng/mL (in BSA buffer) and controls consisting of manufacturer-specified IGFBP-3 concentrations (in BSA buffer preserved with sodium azide) were used for validation of the immunoradiometric assay. The kit contains tubes with immobilized antiserum for IGFBP-3 polyclonal immunoglobulin from goats to bind IGFBP-3. Goat antiserum to IGFBP-3 polyclonal immunoglobulin in BSA buffer with sodium azide labeled with ^{125}I was used to detect IGFBP-3 on the inside walls of the tubes. The IGFBP-3 intra-assay and inter-assay variances ranged from 1.8 to 3.9% and 0.6 to 1.9% respectively, with a reported detection limit of 0.05 ng/mL. Less than 3 ng/mL of IGFBP-3 was reported using other IGF binding proteins as standards.

Whole blood samples were sent to BioServe Biotechnologies (Laurel, MD) for extraction of genomic DNA. Initially, a phenol:chloroform procedure was utilized for DNA extraction. However, a more efficient and safe modified salt precipitation kit (GenQuik Protocol) was adopted when it became available, and was used for the majority of samples.

Genetic polymorphisms

We selected single nucleotide polymorphisms (SNPs) in *IGF-I* and *IGFBP-3* for evaluation based on previous research, functional significance, or haplotype-tagging

properties (Tables 3.2-3.3; Figures 3.1-3.2). Because there are substantial differences in allele frequencies and in the pattern of linkage disequilibrium among populations, haplotype-tagging SNPs were selected separately for African Americans and Caucasians according to Genome Variation Server (GVS) software sponsored by the Seattle SNPs Program for Genomic Applications (PGA) (8). We selected the Seattle SNPs database, which sequences genes to search for SNPs, as the reference population for *IGF-I*; however, we used the HapMap population for *IGFBP-3* because it has not been evaluated by Seattle SNPs (8). We expanded coverage for selection of haplotype-tagging SNPs to include those within 5 kilobases (kb) of the 5' and 3' ends of each gene, and we used a value of 0.8 for the pairwise correlation coefficient (r^2) for identifying haplotype-tagging SNPs. Only haplotype-tagging SNPs with greater than 5% minor allele frequency (MAF) among African American or Caucasian reference populations were chosen for genotyping. According to the GVS software, there were 29 haplotype-tagging SNPs for *IGF-I* and 12 haplotype-tagging SNPs for *IGFBP-3*. In addition, we selected four SNPs *a priori* based on functional significance, including one nonsynonymous *IGF-I* SNP (rs17884626), one synonymous *IGF-I* SNP (rs3729846), and two nonsynonymous *IGFBP-3* SNPs (rs2854746, rs9282734) and an *IGFBP-3* promoter SNP (-202 A/C, rs2854744) *a priori* based on previous studies of associations with circulating IGF-I and IGFBP-3 levels and health outcomes (9-16).

Genotyping of SNPs was performed using the TaqMan genotyping approach under the supervision of Dr. Jason Luo, Director of the Mammalian Genotyping Core at the Lineberger Comprehensive Cancer Center (Chapel Hill, NC). The TaqMan genotyping procedure is advantageous because it is highly accurate, rapid and reliable, and utilizes

only small amounts of DNA (17-19). This method can only measure one SNP at a time and may not be as cost-effective as high-throughput methods when genotyping large numbers of SNPs (19, 20); however, the number of selected SNPs in our study is small enough to make TaqMan genotyping a cost-efficient method. Specifically, the TaqMan procedure utilizes two allele-specific olignonucleotide probes that match the two possible alleles of a SNP. Each probe includes a distinct fluorescent reporter dye on its 5' end and a nonfluorescent quencher on its 3' end that suppresses the reporter dye signal in the absence of amplification. During the polymerase chain reaction (PCR), the allele-specific probe that corresponds to the target sequence will be disintegrated by Taq DNA polymerase, and the fluorescent signal from the reporter dye will strengthen with each cycle of DNA amplification due to the absence of the quencher. In contrast, the mismatched allele-specific probe will be dislodged by DNA polymerase without being cleaved so that its fluorescent signal remains suppressed by the quencher (18, 19).

Allele-specific oligonucleotide probes for 39 selected SNPs were purchased from Applied Biosystems (ABI; Foster City, CA) "TaqMan® Validated and Coding SNP or Pre-Designed SNP Genotyping Assays" (previously known as Assays-On Demand, AOD), and ABI attempted to develop custom assays for the 6 remaining SNPs through their "Custom TaqMan® SNP Genotyping Assays" service (previously known as Assays-By-Design, ABD). Two *IGF-I* haplotype tagging SNPs were excluded from analyses, including one for which a custom assay could not be developed, and one with a pre-designed assay that did not meet ABI technical specifications.

We evaluated whether the race-specific distributions of SNP genotypes in our study population were consistent with Hardy-Weinberg equilibrium (HWE) based on the exact

test statistic with one degree of freedom (21) estimated in Haploview software (22). We genotyped an alternate *IGF-I* haplotype-tagging SNP (rs9308315) to substitute for a SNP (rs4764883), that seemed inconsistent with HWE among African Americans in our study population. Therefore, we genotyped 30 *IGF-I* and 15 *IGFBP-3* SNPs in total, including 40 haplotype-tagging SNPs and 5 SNPs selected *a priori*. Tables 3.4 and 3.5 list the p values from exact tests evaluating HWE and the observed genotype frequencies for *IGF-I* and *IGFBP-3* SNPs respectively.

PCR amplification was performed on an ABI GeneAmp® PCR System 9700 thermal cycler with dual 384-well-blocks, and then endpoint plates were read using the ABI 7900HT system. VIC and 6-FAM reporter dyes were used as the fluorescent signals to distinguish wild type and variant alleles. Alleles were called automatically through Sequence Detection System (SDS) 2.3 software, and output was reviewed by experienced operators. DNA concentration of samples from 984 women was validated using a NanoDrop® ND-1000 Spectrophotometer prior to diluting them to a concentration of 5 ng/ul (using DNA grade sterile water). All of the samples were placed in eleven 96-well microtiter plates and then were aliquoted into three 384-well PCR plates for analysis. Each of the 11 microtiter plates contained four randomly assigned controls represented by two blank samples and two samples consisting of a known DNA standard (Control DNA CEPH Individual 1347-02, ABI). Quality control measures also included blinded genotyping of 28 duplicate samples representing 22 women, which produced concordant results for all samples. The overall call rate was 98.8%, and only 5 women had less than 50% of complete allele calls for the 45 SNPs assayed.

Covariates

Data on potential covariates were obtained through telephone interview, self-administered mail questionnaires, or the initial clinic visit. Age was calculated according to date of study entry using birth date information provided at telephone interview. The following age categories were created for analyses: 35 to 39, 40 to 44, and ≥45 years. Highest level of education was assessed from categorized responses in self-administered mail questionnaires.

Complete reproductive history including detailed information related to pregnancies, menstrual cycles, and contraceptive use was provided at telephone interview. Previous research from the parent study identified an inverse association between number of births and prevalence of fibroids, specifically for births occurring after age 24. The hypothesized mechanism was apoptosis of early lesions during postpartum remodeling of the uterus (23). The number of births after age 24 was classified into the following categories: 0, 1, 2, and ≥3 births. Oral contraceptive use was assessed at study entry and during teenage years based on two studies that found positive associations between early initiation of oral contraceptive use and fibroids (24, 25).

Current body mass index (BMI) in kg/m^2 was calculated from self-reported height at telephone interview and weight at clinic visit. BMI was categorized as normal (<25), overweight (25 to <30), and obese (≥30).

Information on current physical activity was obtained at the telephone interview through separate questions on the duration of vigorous and moderate recreational activities. In addition, women reported the time spent walking and performing household chores. Responses to all of these questions were used to classify overall physical activity

into an index calculated at equivalent to hours of vigorous physical activity. Current physical activity index was categorized according to its overall distribution.

Smoking status was categorized as nonsmoker, former smoker, and current smoker based on information given in the telephone interview. Women were considered nonsmokers if they responded "no" when asked if they ever smoked an average of at least one cigarette a day for six months or more. The remaining women were classified as either former smokers, if they did not currently smoke at least one cigarette per day, or as current smokers.

Racial classification was based on self-report from mail questionnaires having the following categories: White, not Hispanic; White, Hispanic; Black, not Hispanic; Black, Hispanic; Asian/Pacific Islander; American Indian/Eskimo/Aleut; and Other. Women who reported "Other" were given the option of describing their race-ethnicity. Because race represents an important variable in all potential analyses of factors related to fibroids, additional efforts were made to obtain racial information from women who did not return mail questionnaires (N=65, 6.6% of the study sample for the current investigation), including follow-up telephone calls and medical records abstraction (only if calls could not be completed). Women were categorized as African American if they indicated their race as "black" regardless of whether they also reported other race-ethnicity groups. They were considered Caucasian if they specified their race as "white" and did not also record their race as "black". Women who were not classified as either African American or Caucasian will be excluded from the present investigation.

Alcohol intake was estimated from two self-administered questionnaires. The first questionnaire inquired about usual intake within the past 12 months and at 30 years of

age, without regard to the type of alcohol consumed. Specifically, women selected the most appropriate categories for how often and how many alcohol drinks they typically consumed. The second source of data was from a modified Block food frequency questionnaire (FFQ) (26, 27) that assessed alcohol intake during the past 12 months. Women chose their usual serving size and the collective category reflecting the typical number of drinks and frequency for beer, wine, and liquor separately. We computed the drinks per week from each questionnaire by multiplying the midpoints of the frequency and amount categories for the two time points (current and 30 years of age) from the first questionnaire and the serving size and combined frequency and amount category from the FFQ. There was a strong correlation (Spearman rank correlation coefficient = 0.90) between the two reporting sources for estimated drinks per week within the past year. Because of concerns with underreporting of alcohol intake (28), we used the maximum value of self-reported drinks per week within the past twelve months to reflect current intake. Women were classified as current drinkers (≥0.5 drinks per week) or nondrinkers (<0.5 drinks per week) based on previous analyses of parent study data that supported a threshold effect of alcohol intake in which even low levels of consumption are associated with an increased prevalence of fibroids (29).

Analytical Approach

Diplotype estimation

We estimated race-specific diplotypes for groups of related SNPs within each gene, as described below. We excluded one SNP (rs4764883), which appeared inconsistent with HWE among African Americans, from African American diplotype analyses since

including SNPs with a deviation from HWE can reduce reliability of diplotype estimation. In addition, SNPs were excluded from race-specific diplotype analyses if their MAF in our study population was below 5% for haplotype-tagging SNPs or below 3% for *a priori* SNPs within the racial group being evaluated. Women missing genotype data for more than 50% of the SNPs included in diplotype analyses within a gene were excluded from diplotype estimation for that gene (1 Caucasian and 3 African Americans for *IGFBP-3* analyses, 3 Caucasians and 2 African Americans for *IGF-I* analyses). To determine which SNPs in each gene could be combined for diplotype estimation, we examined race-specific linkage disequilibrium (LD) patterns using Haploview software (22). We identified LD blocks consisting of individual SNPs (with MAF at least 5%) in strong LD based on the Gabriel, et al. definition (95% of pairwise SNP comparisons with one-sided 95% confidence intervals for the D prime statistic within 0.7 to 0.98) (30). We also used the Tagger approach (31) in Haploview to identify pairs of redundant SNPs in strong LD (pairwise r^2 values of at least 0.8), and excluded one member of each redundant pair from diplotype estimation unless both SNPs were selected *a priori*. Race-specific pairwise r^2 values are available for *IGFBP-3* and *IGF-I* SNPs in Tables A.1-A.4.

Race-specific diplotypes representing defined groups of SNPs within each gene were estimated using PHASE version 2.1, a software program that employs a Bayesian method to determine the phase of ambiguous haplotypes (32, 33). Specifically, this approach involves a Markov chain-Monte Carlo (MCMC) algorithm to allocate the most likely diplotype for each person, with the prior assumption that frequently observed haplotypes with less ambiguity due to homozygotes are more probable for assignment. This

software program also provides an estimate of the posterior probability, which expresses the uncertainty associated with each diplotype assignment.

Analytic methods to account for diplotype uncertainty in models of gene-disease associations have been debated in the literature, but they have focused primarily on case-control studies (34, 35). One of these methods assigns individuals to their most probable diplotype while eliminating persons with posterior probabilities under a pre-specified value. However, this approach can bias effect estimates towards the null if the criterion is set too low (i.e. posterior probability of 70%) or inflate variance estimates when the criterion is set too high (i.e. posterior probability of 90%) due to the increased number of excluded observations (34). Excluding persons with posterior probabilities less than 90% also may bias estimates due to the systematic removal of observations with a sizeable number of heterozygous genotypes (34). We considered more complex methods to account for diplotype uncertainty in our study, but there were still potential issues with bias and estimating appropriate variances. In addition, more complex methods often imposed analysis restrictions such as types of estimates for evaluating diplotype associations with outcomes.

We chose to assign the women in our study population to their most probable diplotype since this method offered the most flexibility with our analyses. We eliminated women with uncertain diplotypes using a high criterion for the posterior probability (90%) to lessen misclassification bias with assigned diplotypes. To reduce the number of women excluded from diplotype analyses, we examined LD patterns as previously described using Haploview software (22) to determine which combinations of SNPs would estimate diplotypes with a high degree of certainty (posterior probability $\geq$ 90%)

for the greatest proportion of our study population. Specifically, we created race-specific diplotype groups by combining individual SNPs that were not included in a block with an adjacent block, and combining adjacent blocks with each other, when doing so resulted in estimated diplotypes with at least 90% certainty (posterior probability) for at least 90% of observations. Otherwise, diplotype groups represented individual SNPs not included in a block, or SNPs within a single block only. We assigned women to their most probable diplotype for each group, or we classified them as missing for the diplotype group if their most probable diplotype had a posterior probability below 90%.

Overview of analyses and confounding

All statistical analyses were stratified by race and conducted using SAS V9.1 (SAS Institute Inc., Cary, NC). We evaluated *IGF-I* and *IGFBP-3* variants (individual SNPs and estimated diplotypes) as the exposures of interest in association with: 1) circulating IGF-I and IGFBP levels and 2) prevalence of fibroids (any size and at least 2 cm in diameter).

A priori investigation of potential confounders using a directed acyclic graph (DAG) found that race was the only confounder for the association between *IGF-I* and *IGFBP-3* variants and their circulating protein levels and for the association between *IGF-I* and *IGFBP-3* variants and fibroids (Figure 3.3). Rather than adjusting for race, we stratified all analyses by race (African American or Caucasian) and reported race-specific estimates. Population stratification could still bias our race-specific estimates if outcome distribution (plasma levels or fibroid prevalences) and genotype frequencies differ among ethnic subgroups within the African American or Caucasian populations in our study. There is stronger potential for admixture within African Americans based on their

extensive genetic heterogeneity compared to Caucasians and findings from the literature (36-38). However, a breast cancer study applied the approach provided by the Wacholder, et al. evaluation of Caucasians (38) and found that ignoring race in analyses did not bias results in a population well represented by both African Americans and Caucasians (39). The findings of two recent simulation studies, including one that used African American data, suggest that population stratification results in negligible bias when there is substantial ethnic heterogeneity within racial groups, but that bias could increase when the number of ethnic subpopulations is small (36, 40). Although we do not have additional information on the ethnic background of Caucasians and African Americans in our study population, we expect that conducting separate analyses by race may be sufficient to limit bias from population stratification in our study.

Paper 1 analyses

We evaluated whether *IGF-I* and *IGFBP-3* variants were associated with IGF-I and IGFBP-3 plasma levels using the race-specific ordinary linear regression models for gene variants. Because of biological interrelations between IGF-I and IGFBP-3 plasma, variants in each gene were analyzed for associations with circulating levels of both proteins.

For the analysis of individual SNPs, we first considered a codominant (general) model of inheritance in which heterozygous and homozygous variant genotypes were separately evaluated relative to the wildtype genotype. Codominant models make the fewest assumptions with regard to the inheritance pattern but were not feasible for less common variants. Therefore, when there were 10 or fewer women with the homozygous variant genotype, we assumed a dominant model that compared a single index category,

including homozygous variant and heterozygous genotypes, to the homozygous wildtype genotype. We compared race-specific associations with IGF-I or IGFBP-3 levels for individual SNPs with similar models of inheritance for both racial groups. Specifically, we combined data from both racial groups and estimated racial differences in mean plasma levels (i.e., differences of mean differences) associated with variant vs. referent genotypes. We fit linear regression models that included multiplicative gene by race interaction terms along with separate parameters for race and genotypes. When there were differences by race in homozygous wildtype (referent) genotypes for SNPs, we used the referent genotype for Caucasians in regression models including both racial groups.

To estimate diplotype associations with IGF-I or IGFBP-3 plasma levels, we used separate race-specific models for each diplotype group, with the most frequent diplotype as the reference category. Imprecision and bias (away from the null) with regression estimates were especially problematic for associations with diplotypes based on few observations. We did not estimate associations with individual rare diplotypes (assigned to 5 or fewer women), but instead combined them into one "rare diplotype" category for each group. However, due to the heterogeneity of the "rare diplotype" category, we were not able to interpret associations with this category. Despite the creation of "rare diplotype" categories, some of the diplotype groups had multiple diplotype categories that were not considered rare but were based on small numbers of observations.

We used a shrinkage (empirical-Bayes) method of information-weighted averaging (41) to improve the validity and precision of diplotype regression estimates. Specifically, we assumed a prior mean of 0, since we did not have prior information to group diplotypes according to the anticipated direction or strength of potential associations with

plasma levels, and specified a prior variance corresponding to +/- one standard deviation $(2*\text{standard deviation}/3.92)^2$ of the mean plasma level of IGF-I (prior variances: African Americans, 1,419; Caucasians, 901) and IGFBP-3 (prior variances: African Americans, 186,819; Caucasians, 174,161) in the study population. We applied the shrinkage estimator for each diplotype and reported posterior medians (50^{th} percentile of the posterior probability distribution) and 95% posterior limits (2.5^{th} and 97.5^{th} percentiles of the posterior probability distribution). With this method, imprecise regression estimates based on fewer observations were shifted further toward the prior mean than more precise regression estimates. Despite the increase in bias (toward the prior mean) with posterior medians, the reduction in the overall mean square error based on a greater reduction in variance of estimates is an advantage with this approach (42).

Paper 2 analyses

We estimated race-specific prevalence differences (PDs) to evaluate associations between *IGF-I* and *IGFBP-3* gene variants and prevalent fibroids. We performed two separate analyses to 1) estimate associations with gene variants and prevalence of any uterine fibroids versus no uterine fibroids, and 2) estimate associations with prevalence of fibroids at least 2 cm in diameter versus a combined category including no fibroids and fibroids < 2cm. We used the latter analysis, which excludes women with smaller tumors from the fibroid definition, to explore the hypothesis that IGF-I stimulates fibroid growth rather than incidence. We estimated 95% confidence intervals for crude prevalence differences (PD) using SAS V9.1, which calculates standard errors from the square root of the sum of the fibroid prevalence variances based on formula for independent proportions.

For the analysis of individual SNPs, we first considered a codominant (general) model of inheritance in which heterozygous and homozygous variant genotypes were separately evaluated relative to the homozygous wildtype genotype. However, codominant models were feasible only for common variants. Therefore, for less common variants, we used a dominant model in which homozygous variant and heterozygous genotypes were combined and compared to those with the homozygous wildtype genotype. When there were 5 or fewer women in a racial group with heterozygous and homozygous variant genotypes combined, we did not estimate associations with individual SNPs since validity of estimates would be questionable.

We estimated associations between *IGF-I* or *IGFBP-3* diplotypes and fibroids using separate race-specific models for each diplotype group, with the most common diplotype as the reference category. Due to convergence issues with binomial regression models, we used Poisson regression with the robust variance option in PROC GENMOD to estimate prevalence differences with 95% confidence intervals (43). Diplotypes assigned to 5 or fewer women were combined into a "rare diplotype" category; however, because of the heterogeneity of diplotypes within these categories, their estimated associations with fibroids could not be interpreted.

Imprecision and bias (away from the null) were a limitation for interpreting estimated associations with diplotypes based on smaller numbers of observations. Therefore, we used a shrinkage (empirical-Bayes) method of information-weighted averaging (41) to improve the validity and precision of regression estimates. We assumed a prior mean of 0 since we did not have information to group diplotypes according to the direction or strength of their potential associations with fibroids. We

specified a prior variance ($(0.5/3.92)^2 \approx 0.016$), which assumed with 95% probability that regression estimates would vary within a range of ± 0.25. For each diplotype, we applied the shrinkage estimator and reported posterior medians (50^{th} percentile of the posterior probability distribution) and 95% posterior limits (2.5^{th} and 97.5^{th} percentiles of the posterior probability distribution). Imprecise estimates based on few observations were shrunk further toward the prior mean than more precise regression estimates.

Power calculations

Table 3.6 depicts statistical power to estimate a range of statistically significant prevalence differences ($\alpha = .05$) for dichotomous genotypes (SNPs) based on values generated by the PS program (44). In this scenario, the genotype frequency reflects heterozygous and homozygous variant genotypes grouped together in order to maximize power. Estimated values for baseline fibroid prevalence (any size) in calculations were based on known values in the study population of 73% and 51% among African Americans and Caucasians, respectively.

For heterozygous and homozygous variant genotypes with frequencies of at least 20%, estimated power is over 80% for prevalence differences as low as 0.15 for African Americans and 0.20 for Caucasians. There will be limited power to estimate weaker associations (i.e. PD of 0.10 or less) or fibroid relations with rare gene variants. Power estimates were not calculated for diplotype analyses using information-weighted averaging (41), but this approach may enhance statistical power over conventional regression methods.

REFERENCES

1. Baird DD, Dunson DB, Hill MC, Cousins D, Schectman JM. High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence. Am J Obstet Gynecol 2003;188:100-7.

2. Baird DD, Dunson DB, Hill MC, Cousins D, Schectman JM. Association of physical activity with development of uterine leiomyoma. Am J Epidemiol 2007;165:157-63.

3. Baird DD, Kesner JS, Dunson DB. Luteinizing hormone in premenopausal women may stimulate uterine leiomyomata development. J Soc Gynecol Investig 2006;13:130-5.

4. Schwartz SM, Marshall LM, Baird DD. Epidemiologic contributions to understanding the etiology of uterine leiomyomata. Environ Health Perspect 2000;108 Suppl 5:821-7.

5. Dueholm M, Lundorf E, Hansen ES, Ledertoug S, Olesen F. Accuracy of magnetic resonance imaging and transvaginal ultrasonography in the diagnosis, mapping, and measurement of uterine myomas. Am J Obstet Gynecol 2002;186:409-15.

6. Muram D, Gillieson M, Walters JH. Myomas of the uterus in pregnancy: ultrasonographic follow-up. Am J Obstet Gynecol 1980;138:16-9.

7. Hurley V. Imaging techniques for fibroid detection. Baillieres Clin Obstet Gynaecol 1998;12:213-24.

8. Seattle SNPs Program for Genomic Applications (PGA). Genome Variation Server (GVS) Version 1.04, 2006:http://gvs.gs.washington.edu/GVS/.

9. Canzian F, McKay JD, Cleveland RJ, et al. Polymorphisms of genes coding for insulin-like growth factor 1 and its major binding proteins, circulating levels of IGF-I and IGFBP-3 and breast cancer risk: results from the EPIC study. Br J Cancer 2006;94:299-307.

10. Deal C, Ma J, Wilkin F, et al. Novel promoter polymorphism in insulin-like growth factor-binding protein-3: correlation with serum levels and interaction with known regulators. J Clin Endocrinol Metab 2001;86:1274-80.

11. Al-Zahrani A, Sandhu MS, Luben RN, et al. IGF1 and IGFBP3 tagging polymorphisms are associated with circulating levels of IGF1, IGFBP3 and risk of breast cancer. Hum Mol Genet 2006;15:1-10.

12. Jernstrom H, Deal C, Wilkin F, et al. Genetic and nongenetic factors associated with variation of plasma levels of insulin-like growth factor-I and insulin-like growth factor-binding protein-3 in healthy premenopausal women. Cancer Epidemiol Biomarkers Prev 2001;10:377-84.

13. Lai JH, Vesprini D, Zhang W, Yaffe MJ, Pollak M, Narod SA. A polymorphic locus in the promoter region of the IGFBP3 gene is related to mammographic breast density. Cancer Epidemiol Biomarkers Prev 2004;13:573-82.

14. Schernhammer ES, Hankinson SE, Hunter DJ, Blouin MJ, Pollak MN. Polymorphic variation at the -202 locus in IGFBP3: Influence on serum levels of insulin-like growth factors, interaction with plasma retinol and vitamin D and breast cancer risk. Int J Cancer 2003;107:60-4.

15. Cheng I, Henderson KD, Haiman CA, et al. Genetic Determinants of Circulating IGF-I, IGFBP-1 and IGFBP-3 Levels in a Multiethnic Population. J Clin Endocrinol Metab 2007.

16. Ren Z, Cai Q, Shu XO, et al. Genetic polymorphisms in the IGFBP3 gene: association with breast cancer risk and blood IGFBP-3 protein levels among Chinese women. Cancer Epidemiol Biomarkers Prev 2004;13:1290-5.

17. Behrens M, Lange R. A highly reproducible and economically competitive SNP analysis of several well characterized human mutations. Clin Lab 2004;50:305-16.

18. Livak KJ. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. Genet Anal 1999;14:143-9.

19. McGuigan FE, Ralston SH. Single nucleotide polymorphism detection: allelic discrimination using TaqMan. Psychiatr Genet 2002;12:133-6.

20. Mamotte CD. Genotyping of single nucleotide substitutions. Clin Biochem Rev 2006;27:63-75.

21. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet 2005;76:887-93.

22. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.

23. Baird DD, Dunson DB. Why is parity protective for uterine fibroids? Epidemiology 2003;14:247-50.

24. Marshall LM, Spiegelman D, Goldman MB, et al. A prospective study of reproductive factors and oral contraceptive use in relation to the risk of uterine leiomyomata. Fertil Steril 1998;70:432-9.

25. Wise LA, Palmer JR, Harlow BL, et al. Reproductive factors, hormonal contraception, and risk of uterine leiomyomata in African-American women: a prospective study. Am J Epidemiol 2004;159:113-23.

26. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. Am J Epidemiol 1986;124:453-69.

27. Block G, Hartman AM, Naughton D. A reduced dietary questionnaire: development and validation. Epidemiology 1990;1:58-64.

28. Ekholm O. Influence of the recall period on self-reported alcohol intake. Eur J Clin Nutr 2004;58:60-3.

29. D'Aloisio AA, Baird DD. Variation in the association of alcohol intake with uterine fibroids by race and tumor size. Ann Epidemiol 2004;14:622.

30. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225-9.

31. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet 2005;37:1217-23.

32. Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 2003;73:1162-9.

33. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001;68:978-89.

34. Kraft P, Cox DG, Paynter RA, Hunter D, De Vivo I. Accounting for haplotype uncertainty in matched association studies: a comparison of simple and flexible techniques. Genet Epidemiol 2005;28:261-72.

35. Cordell HJ. Estimation and testing of genotype and haplotype effects in case-control studies: comparison of weighted regression and multiple imputation procedures. Genet Epidemiol 2006;30:259-75.

36. Wang Y, Localio R, Rebbeck TR. Evaluating bias due to population stratification in case-control association studies of admixed populations. Genet Epidemiol 2004;27:14-20.

37. Ardlie KG, Lunetta KL, Seielstad M. Testing for population subdivision and association in four case-control studies. Am J Hum Genet 2002;71:304-11.

38. Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst 2000;92:1151-8.

39. Millikan RC. Re: Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst 2001;93:156-8.

40. Wang Y, Localio R, Rebbeck TR. Evaluating bias due to population stratification in epidemiologic studies of gene-gene or gene-environment interactions. Cancer Epidemiol Biomarkers Prev 2006;15:124-32.

41. Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. Int J Epidemiol 2006;35:765-75.

42. Greenland S, Poole C. Empirical-Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. Arch Environ Health 1994;49:9-16.

43. Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. Am J Epidemiol 2005;162:199-200.

44. Dupont WD, Plummer WD. PS power and sample size program available for free on the Internet. Control Clin Trials 1997;18:274.

Table 3.1. Characteristics of premenopausal women with available DNA by race from NIEHS Uterine Fibroid Study

Characteristic	African American (N=582) N (%)	Caucasian (N=402) N (%)
Age (years)		
35-39	219 (37.63)	137 (34.08)
40-44	205 (35.22)	135 (33.58)
45+	158 (27.15)	130 (32.34)
Education		
High school	121 (20.8)*	12 (3.0)
Some postsecondary	265 (45.5)	33 (8.2)
College degree	123 (21.1)	133 (33.1)
Graduate degree	68 (11.7)	217 (54.0)
Missing	5 (0.9)	7 (1.7)
Age at menarche (years)		
<11	65 (11.17)	16 (3.98)
11	94 (16.15)	62 (15.42)
12	159 (27.32)	111 (27.61)
13	138 (23.71)	132 (32.84)
14	55 (9.45)	44 (10.95)
>14	68 (11.68)	35 (8.7)
Missing	3 (0.52)	2 (0.5)
Oral contraceptive use		
Current†	28 (4.8)	37 (9.2)
Teenage years	350 (60.1)	131 (32.6)
Births (25+ years of age)		
0	295 (50.69)	255 (63.4)
1	181 (31.10)	57 (14.2)
2	85 (14.60)	80 (19.9)
3+	21 (3.61)	10 (2.5)
Body mass index		
Under- normal weight (<25)	147 (25.26)	236 (58.7)
Overweight (25-<30)	178 (30.58)	95 (23.6)
Obese (30+)	257 (44.16)	71 (17.7)
Physical activity (index)‡		
0 to <2	223 (38.3)	114 (28.4)
2 to 4	163 (28.0)	127 (31.6)
>4 to <6	96 (16.5)	74 (18.4)
6+	97 (16.7)	86 (21.4)
Missing	3 (0.5)	1 (0.25)
Smoking status		
Current smoker	172 (29.6)	31 (7.7)
Former smoker	133 (22.9)	139 (34.6)
Never smoked	277 (47.6)	232 (57.7)
Alcohol intake (past year)		
<0.5 drinks/week	300 (51.5)	60 (14.9)
≥0.5 drinks/week	245 (42.1)	314 (78.1)
Missing	37 (6.4)	28 (7.0)

Uterine fibroids		
None	154 (26.46)	194 (48.26)
Any	421 (72.34)	200 (49.75)
<2 cm	94 (16.15)	69 (17.16)
2 - <4 cm	187 (32.13)	87 (21.64)
≥4 cm	140 (24.05)	44 (10.95)
Missing	7 (1.20)	8 (1.99)
	Mean (SD)	Mean (SD)
Plasma IGF-I (ng/mL)†	165 (74.0)	172 (58.7)
Plasma IGFBP-3 (ng/mL)†	4085 (859.7)	4524 (822.3)

* Includes 11 women with less than high school education.

† N missing: oral contraceptives: 4 African Americans, 3 Caucasians; plasma IGF-I and IGFBP-3: 8 African Americans, 1 Caucasian.

‡ Calculated at equivalent to hours of vigorous physical activity.

Table 3.2. *IGF-I* single nucleotide polymorphisms (SNPs) selected for genotyping in the NIEHS Uterine Fibroid Study*

Reference SNP ID	Alternate Name	Chromosome Position	Location	Alleles	Amino Acid Change	TaqMan Assay
rs35767		101399699	promoter	A, G	N/A	Validated
rs5742612		101398994	promoter	A, G	N/A	Pre-designed
rs5742614		101397367	intron 1	C, G	N/A	Pre-designed
rs17032634†		101395474	intron 1	A, G	N/A	Pre-designed
rs3729846	Thr52Thr	101393615	exon 2	C, T	None	Validated
rs12821878		101391797	intron 2	A, G	N/A	Validated
rs10860869		101389182	intron 2	A, T	N/A	Pre-designed
rs1019731		101388555	intron 2	A, C	N/A	Validated
rs7956547		101382946	intron 2	C, T	N/A	Pre-designed
rs5742626		101382039	intron 2	C, T	N/A	Pre-designed
rs17880975		101376814	intron 2	A, G	N/A	Custom
rs2033178		101371206	intron 2	A, G	N/A	Pre-designed
rs17884646		101369945	intron 2	C, T	N/A	Custom
rs5742657		101358998	intron 2	A, G	N/A	Pre-designed
rs5742663		101348120	intron 2	G, T	N/A	Pre-designed
rs11829586		101344517	intron 2	A, G	N/A	Pre-designed
rs4764884		101343729	intron 2	C, T	N/A	Pre-designed
rs5742683		101337847	intron 2	A, G	N/A	Pre-designed
rs17884626	Ala115Thr	101337476	exon 3	C, T	Alanine to Threonine	Validated
rs5009837		101334399	intron 3	C, T	N/A	Pre-designed
rs17727841		101333760	intron 3	C, G	N/A	Pre-designed
rs4764883		101330435	intron 3	C, T	N/A	Pre-designed
rs9308315		101328023	intron 3		N/A	Pre-designed
rs978458		101326369	intron 3	C, T	N/A	Validated
rs12316064†		101324933	intron 3	C, T	N/A	Custom†
rs5742692		101323728	intron 3	A, G	N/A	Pre-designed
rs11111262		101322307	intron 3	A, G	N/A	Custom
rs1520220		101320652	intron 3	C, G	N/A	Validated
rs3730204		101319644	exon 4; 3' untranslated	C, T	N/A	Custom
rs6214		101317699	exon 4; 3' untranslated	C, T	N/A	Validated

rs6219	101314322	exon 4;	C, T	N/A	Validated
rs2946834	101311944	3' untranslated downstream of IGF-I	A, G	N/A	Validated

* Alleles represent DNA strand from TaqMan assays. rs9308315 was genotyped as an alternate SNP to replace rs4764883, which appeared inconsistent with Hardy Weinberg Equilibrium ($\alpha = 0.01$) among African Americans.

† SNPs dropped from analyses due to TaqMan assay failure with design.

Table 3.3. *IGFBP-3* single nucleotide polymorphisms (SNPs) selected for genotyping in the NIEHS Uterine Fibroid Study*

Reference SNP ID	Alternate Name	Chromosome Position	Location	Alleles	Amino Acid Change	TaqMan Assay
rs903889		45931520	promoter	G, T	N/A	Pre-designed
rs924140		45929639	promoter	C, T	N/A	Pre-designed
rs2854744	-202 A/C	45927600	promoter	G, T	N/A	Pre-designed
rs2854746	G2133C (Ala32Gly)	45927170	exon 1	C, G	Alanine to Glycine	Pre-designed
rs2471551		45923580	intron 1	C, G	N/A	Pre-designed
rs9282734	His158Pro	45923494	exon 2	A, C	Histidine to Proline	Custom
rs2453837		45922575	intron 3	A, G	N/A	Pre-designed
rs6953668		45922400	intron 3	A, G	N/A	Pre-designed
rs3110697		45921554	intron 3	A, G	N/A	Pre-designed
rs2453840†		45920337	intron 4	A, C	N/A	Custom
rs2453839		45920098	intron 4	C, T	N/A	Pre-designed
rs6670		45918779	exon 5; 3' untranslated	A, T	N/A	Pre-designed
rs13223993		45917755	downstream of IGFBP-3	A, G	N/A	Pre-designed
rs2270628		45916095	downstream of IGFBP-3	C, T	N/A	Pre-designed
rs12671457		45913451	downstream of IGFBP-3	A, C	N/A	Pre-designed

* Alleles represent DNA strand from TaqMan assays, except alleles on opposite DNA strand shown for rs2854744 (-202 A/C).

† Possible triallelic polymorphism but has not been confirmed by sequencing groups.

Table 3.4. Genotype distribution of *IGF-I* single nucleotide polymorphisms (SNPs) among premenopausal Caucasian and African American women in the NIEHS Uterine Fibroid Study

SNP Reference ID	Caucasians*				African Americans*			
	MAF (%)	Genotypes	N	HWE p value†	MAF (%)	Genotypes	N	HWE p value†
rs35767	16.0	AA	10	1.000	40.8	AA	102	0.322
		AG	107			AG	266	
		GG	280			GG	208	
rs5742612	3.9	GG	1	0.909	3.3	GG	2	0.240
		AG	29			AG	34	
		AA	369			AA	543	
rs5742614	1.8	CC	0	1.000	11.5	CC	8	1.000
		CG	14			CG	117	
		GG	386			GG	451	
rs3729846	0	CT	0	--	0.8	CT	9	1.000
		CC	400			CC	565	
rs12821878	22.7	AA	21	0.986	5.8	AA	3	0.598
		AG	139			AG	61	
		GG	238			GG	513	
rs10860869	29.1	TT	31	0.596	32.3	TT	61	0.931
		AT	170			AT	251	
		AA	197			AA	266	
rs1019731	12.3	AA	5	0.858	3.4	AA	0	1.000
		AC	88			AC	39	
		CC	305			CC	536	
rs7956547	27.3	CC	29	0.971	26.2	CC	44	0.363
		CT	160			CT	213	
		TT	210			TT	318	
rs5742626	0.5	CC	0	1.000	7.2	CC	4	0.680
		CT	4			CT	75	
		TT	395			TT	498	
rs17880975	0.1	AG	1	1.000	2.3	AG	26	1.000
		GG	396			GG	548	

rs2033178	6.2	AA	1	1.000	4.7	AA	4	0.062
		AG	47			AG	46	
		GG	349			GG	521	
rs17884646	0	TT	398	--	0	TT	578	--
rs5742657	2.2	GG	1	0.354	12.1	GG	10	0.625
		AG	16			AG	119	
		AA	384			AA	447	
rs5742663	0.1	GG	0	1.000	11.6	GG	9	0.699
		GT	1			GT	115	
		TT	397			TT	451	
rs11829586	2.3	AA	0	1.000	11.5	AA	10	0.411
		AG	18			AG	112	
		GG	381			GG	452	
rs4764884	25.7	TT	25	0.880	23.7	TT	35	0.570
		CT	154			CT	202	
		CC	218			CC	338	
rs5742683	0.1	GG	0	1.000	10.6	GG	8	0.610
		AG	1			AG	106	
		AA	397			AA	461	
rs17884626	0	CT	0	--	1.0	CT	11	1.000
		CC	400			CC	567	
rs5009837	30.3	TT	36	1.000	41.3	TT	107	0.155
		CT	168			CT	262	
		CC	192			CC	207	
rs17727841	19.1	CC	15	0.997	4.7	CC	3	0.242
		CG	123			CG	48	
		GG	262			GG	526	
rs4764883	30.9	CC	37	1.000	45.8	TT	135	0.007‡
		CT	168			CT	249	
		TT	186			CC	183	
rs9308315	28.2	AA	28	0.490	49.8	TT	154	0.052
		AT	167			AT	262	
		TT	201			AA	156	

rs978458	28.0	TT	28	0.567	37.5	TT	85	0.466
		CT	165			CT	260	
		CC	202			CC	229	
rs5742692	2.1	GG	0	1.000	11.7	GG	11	0.270
		AG	17			AG	112	
		AA	383			AA	451	
rs11111262	9.9	AA	3	0.881	2.7	AA	2	0.118
		AG	73			AG	27	
		GG	322			GG	542	
rs1520220	19.1	GG	13	0.805	34.7	GG	74	0.443
		CG	125			CG	252	
		CC	258			CC	250	
rs3730204	2.3	CT	18	1.000	0.3	CT	3	1.000
		TT	379			TT	570	
rs6214	39.8	TT	67	0.495	45.2	CC	127	0.107
		CT	184			CT	264	
		CC	148			TT	182	
rs6219	10.0	TT	2	0.429	9.1	TT	9	0.072
		CT	76			CT	87	
		CC	322			CC	481	
rs2946834	33.0	AA	40	0.588	49.3	AA	143	0.665
		AG	181			AG	282	
		GG	175			GG	151	

NOTE: MAF, minor allele frequency; HWE, Hardy Weinberg Equilibrium.

* Caucasians: N = 402; African Americans: N = 582.

† Based on exact test statistic ($\alpha = 0.01$).

‡ Appears inconsistent with HWE ($\alpha = 0.01$).

Table 3.5. Genotype distribution of *IGFBP-3* single nucleotide polymorphisms (SNPs) among premenopausal Caucasian and African American women in the NIEHS Uterine Fibroid Study

SNP Reference ID	Caucasians*				African Americans*			
	MAF (%)	Genotypes	N	HWE p value†	MAF (%)	Genotypes	N	HWE p value†
rs903889	22.0	GG	22	0.510	9.4	GG	10	0.040
		GT	132			GT	88	
		TT	246			TT	478	
rs924140	46.8	TT	82	0.335	38.9	CC	95	0.188
		CT	210			CT	258	
		CC	108			TT	223	
rs2854744	46.3	AA	79	0.397	42.2	CC	107	0.332
		AC	204			AC	265	
		CC	108			AA	196	
rs2854746	41.7	CC	65	0.427	32.1	CC	53	0.284
		CG	203			CG	263	
		GG	131			GG	259	
rs2471551	19.4	CC	17	0.587	20.3	CC	27	0.425
		CG	120			CG	178	
		GG	260			GG	367	
rs9282734	0.4	CC	0	1.000	3.3	CC	2	0.242
		AC	3			AC	34	
		AA	394			AA	540	
rs2453837	0	GG	398	--	0	GG	576	--
rs6953668	0.6	AA	0	1.000	5.0	AA	3	0.335
		AG	5			AG	52	
		GG	395			GG	521	
rs3110697	41.4	AA	63	0.347	36.1	AA	80	0.391
		AG	203			AG	255	
		GG	131			GG	240	
rs2453840	18.3	AA	12	0.840	9.2	AA	7	0.394
		AC	121			AC	92	
		CC	264			CC	477	

rs2453839	20.1	CC	14	0.677	40.9	CC	94	0.784
		CT	131			CT	282	
		TT	251			TT	198	
rs6670	21.2	AA	12	0.101	12.6	AA	9	1.000
		AT	145			AT	127	
		TT	241			TT	439	
rs13223993	21.4	AA	19	0.867	43.6	AA	117	0.185
		AG	131			AG	265	
		GG	245			GG	190	
rs2270628	21.0	TT	15	0.542	36.2	TT	84	0.137
		CT	138			CT	248	
		CC	247			CC	242	
rs12671457	16.5	CC	8	0.495	5.1	CC	3	0.339
		AC	111			AC	52	
		AA	267			AA	518	

NOTE: MAF, minor allele frequency; HWE, Hardy Weinberg Equilibrium.
* Caucasians: N = 402; African Americans: N = 582.
† Based on exact test statistic ($\alpha = 0.01$).

Table 3.6. Estimated power to detect main effects prevalence differences (PD) by race with SNPs or diplotypes ($\alpha = 0.05$)

	Frequency of Genetic Variant*					
PD	0.05	0.10	0.20	0.30	0.40	0.50
	African Americans (n=582)†					
0.10	19.7	31.8	49.9	60.8	66.3	68.0
0.15	37.2	60.0	83.3	91.7	94.6	95.4
0.20	58.3	84.1	97.5	99.4	99.8	99.8
0.30	91.6	99.6	>99.9	>99.9	>99.9	>99.9
	Caucasians (n=402)†					
0.10	17.8	27.0	41.0	49.8	54.4	55.7
0.15	31.8	50.0	71.9	82.1	86.5	87.7
0.20	48.8	72.7	91.7	96.7	98.2	98.5
0.30	80.9	96.8	99.9	>99.9	>99.9	>99.9

*Refers to combined frequency of heterozygous and homozygous variant genotypes.
†Given baseline prevalence of 0.40 (African Americans), 0.30 (Caucasians).

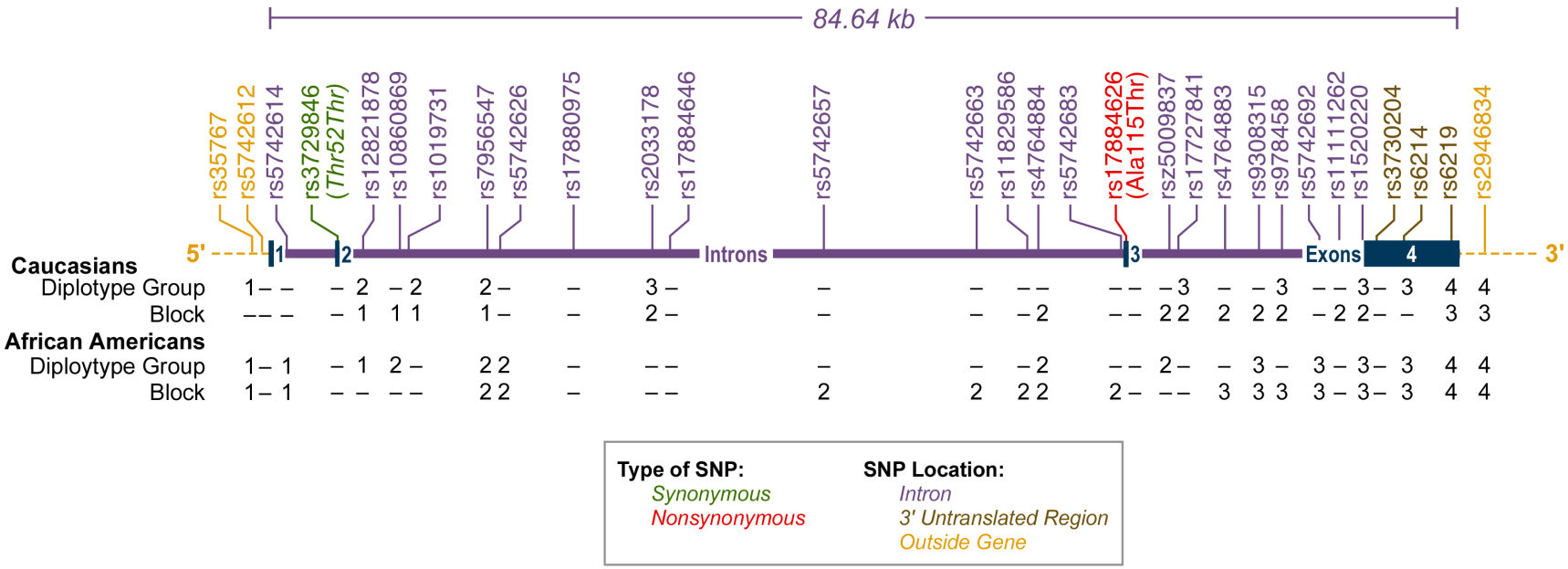

Figure 3.1. Depiction of *IGF-I* single nucleotide polymorphisms (SNPs) evaluated among premenopausal African American and Caucasian participants from the National Institute of Environmental Health Sciences (NIEHS) Uterine Fibroid Study. Relative distances of SNPs were based on mapping information from the Entrez Gene database sponsored by the National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM). SNPs selected *a priori* include: rs3729846, synonymous SNP; rs17884626, nonsynonymous SNP. Race-specific diplotype groups and linkage disequilibrium (LD) blocks are listed for each SNP. Group refers to combination of SNPs for diplotype estimation. LD blocks consist of SNPs with minor allele frequency (MAF) ≥ 5% in strong LD (95% of pairwise SNP comparisons with one-sided 95% confidence intervals for the D prime statistic within 0.7-0.98). Values for diplotype group and position (--) were not listed for SNPs excluded from diplotype estimation based on either: 1) Tagger algorithm (pairwise $r^2 \geq 0.8$); 2) MAF < 5% (excluded from LD blocks (--)) or MAF < 3% for SNPs selected *a priori*.

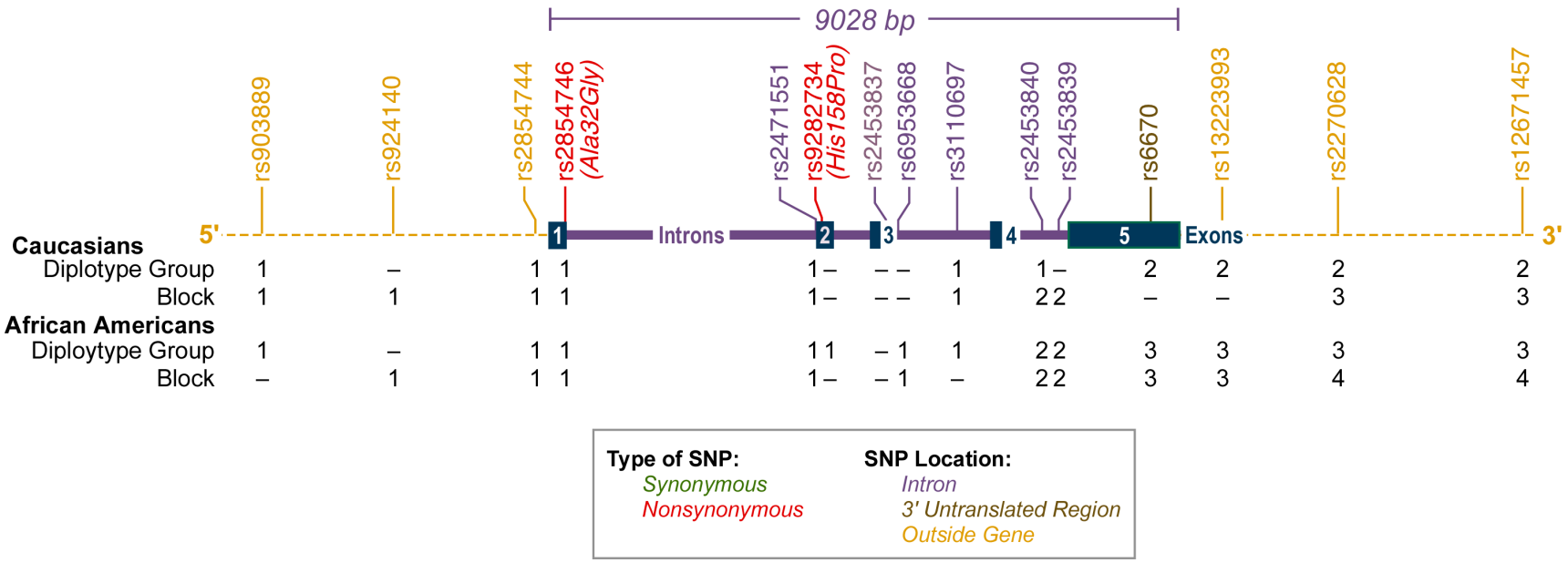

Figure 3.2. Depiction of *IGFBP-3* single nucleotide polymorphisms (SNPs) evaluated among premenopausal African American and Caucasian participants from the National Institute of Environmental Health Sciences (NIEHS) Uterine Fibroid Study. Relative distances of SNPs were based on mapping information from the Entrez Gene database sponsored by the National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM). SNPs selected *a priori* include: rs2854744 (-202 A/C); rs2854746 and rs9282734, nonsynonymous SNPs. Race-specific diplotype groups and linkage disequilibrium (LD) blocks are listed for each SNP. Group refers to combination of SNPs for diplotype estimation. LD blocks consist of SNPs with minor allele frequency (MAF) ≥ 5% in strong LD (95% of pairwise SNP comparisons with one-sided 95% confidence intervals for the D prime statistic within 0.7-0.98). Values for diplotype group and position (--) were not listed for SNPs excluded from diplotype estimation based on either: 1) Tagger algorithm (pairwise $r^2 \geq 0.8$); 2) MAF < 5% (excluded from LD blocks (--)) or MAF < 3% for SNPs selected *a priori*.

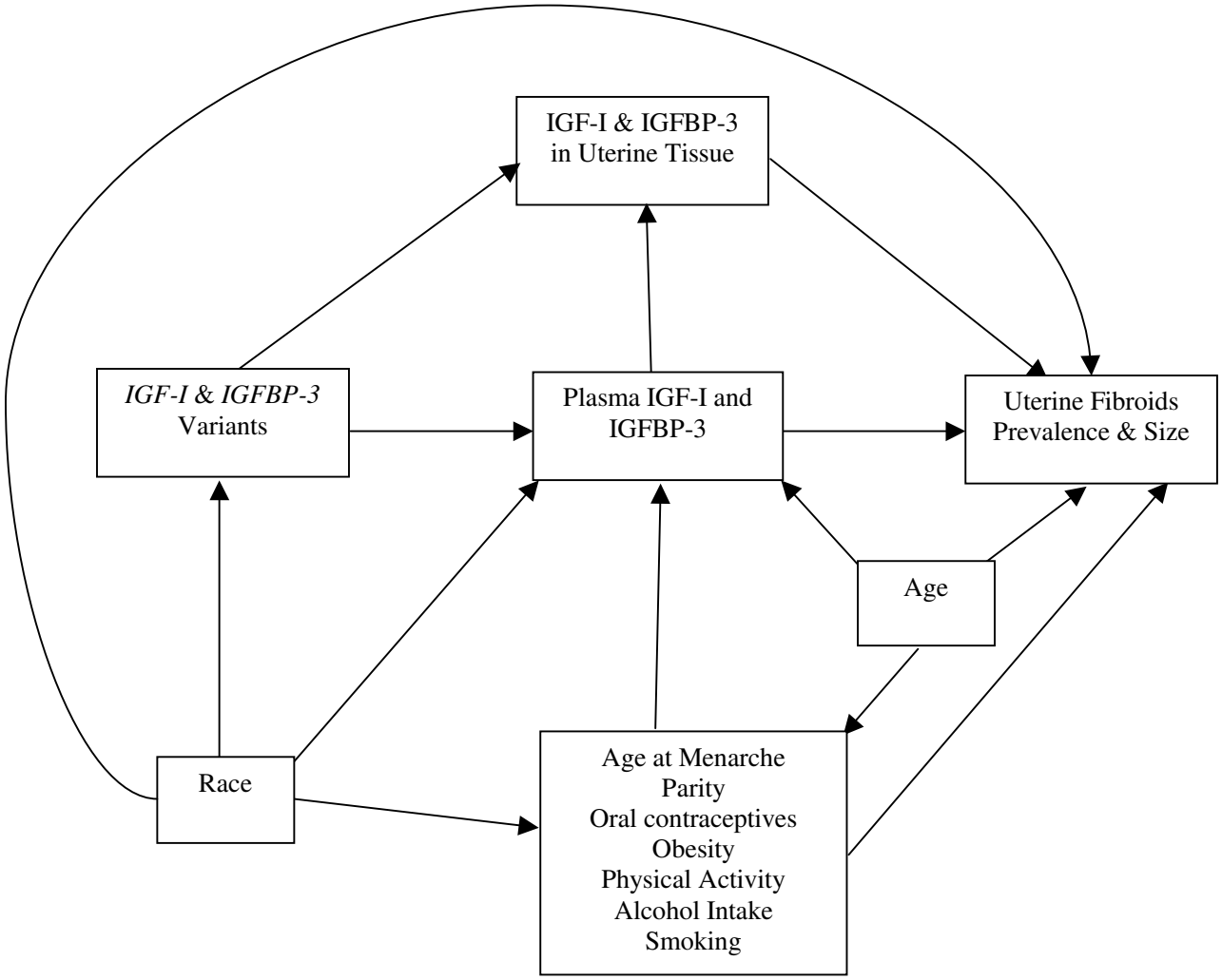

Figure 3.3. Hypothetical causal diagram.

CHAPTER IV

PAPER 1: *IGF-I* AND *IGFBP-3* POLYMORPHISMS IN RELATION TO CIRCULATING LEVELS AMONG AFRICAN AMERICAN AND CAUCASIAN WOMEN

Abstract

Circulating insulin-like growth factor-one (IGF-I) and IGF binding protein-3 (IGFBP-3) have been studied in relation to common diseases. Although twin- and family-based studies suggest that genetic variation contributes to circulating IGF-I and IGFBP-3 levels, prior analyses of associations with multiple *IGF-I* and *IGFBP-3* polymorphisms have been limited, especially among African Americans. We evaluated 30 *IGF-I* and 15 *IGFBP-3* single nucleotide polymorphisms (SNPs) and estimated diplotypes in association with plasma IGF-I and IGFBP-3 levels among 984 premenopausal African American and Caucasian women. We noted associations between *IGF-I* and *IGFBP-3* SNPs and plasma IGF-I and IGFBP-3 among both races. In both racial groups, *IGFBP-3* rs2854746 (Ala32Gly) was positively associated with plasma IGFBP-3, (CC versus GG: Caucasians: 631 ng/ml, 95% CI: 398, 864; African Americans: 897 ng/ml, 95% CI: 656, 1138). In comparison to referent diplotypes with the rs2854746 CG genotype, *IGFBP-3* diplotypes with the GG genotype had lower mean plasma IGFBP-3 levels while *IGFBP-3* diplotypes with the CC genotype had higher mean plasma IGFBP-3 levels among both races. The nearby *IGFBP-3* promoter SNP, rs2854744, which was in strong LD ($r^2 > 0.8$) with rs2854746 in Caucasians but not

African Americans, was also associated with plasma IGFBP-3 in both races. Eight other *IGFBP-3* SNPs were associated with plasma IGFBP-3, and associations were generally consistent between races. Twelve *IGF-I* SNPs were associated with plasma IGF-I, but these associations were generally discordant between races. There was no consistent pattern between *IGF-I* SNP and diplotype associations with plasma IGF-I. Associations were also noted between *IGFBP-3* SNPs and plasma IGF-I and between *IGF-I* SNPs and plasma IGFBP-3 although many were relatively imprecise, and there was no consistent pattern with their respective diplotype associations.

Introduction

Insulin-like growth factor-one (IGF-I), a peptide with structural similarities to insulin, has been the focus of prior research because of its potential role in disease. Specifically, IGF-I has been implicated in many biologic processes, including cell cycle regulation, differentiation, proliferation, hormone secretion, and apoptosis. IGF-binding proteins (IGFBPs) help regulate the activity of IGFs by influencing their bioavailablity and degradation (1, 2). IGFBPs may also have independent effects through interactions with cell surface molecules (1, 2). IGFBP-3, which binds approximately 90% of circulating IGF-I, has been the most studied of the IGFBPs (3). Independent of effects on IGFs, IGFBP-3 may inhibit growth and increase apoptosis in breast cancer cells (4, 5), but these findings have not been confirmed.

Circulating IGF-I levels, and to a lesser extent circulating IGFBP-3 levels, have been studied in association with cardiovascular disease, diabetes, and cancer (6-9). Prior studies have reported positive associations between circulating IGF-I and premenopausal

breast cancer (10), but two recent studies reported positive IGF-I associations with postmenopausal breast cancer only (11, 12). Associations between pre- or post-menopausal breast cancer and circulating IGFBP3 associations have been inconsistent (10).

Estimates from twin- or family-based studies suggest that genetic factors may account for up to 50% of the inter-individual variation in plasma IGF-I levels (13, 14) and up to 60% of the variation in plasma IGFBP-3 levels (13, 15). In adults, age is the most consistently reported nongenetic factor associated with IGF-I blood levels (3, 16-25), with lower circulating IGF-I levels (3, 16-25) associated with increasing age. Women have lower circulating IGF-I (17, 20, 21, 23, 25, 26) but higher IGFBP-3 levels (17, 20, 21, 25, 26) than men, and two studies have suggested that African American women have higher circulating IGF-I levels than Caucasian women (21, 27).

Our research goal was to investigate relations between *IGF-I* and *IGFBP-3* polymorphisms and circulating IGF-I and IGFBP-3 levels among African American and Caucasian women. Prior analysis of dense single nucleotide polymorphisms (SNP) and IGF-I and IGFBP-3 levels among African Americans has been limited to the Multiethnic Cohort Study, which included only about 150 African Americans (28). We selected 45 SNPs in *IGF-I* and *IGFBP-3* and examined whether these SNPs and their estimated diplotypes (paired haplotypes) were associated with IGF-I and IGFBP-3 plasma levels in premenopausal African American and Caucasian women that participated in the National Institute for Environmental Health Sciences (NIEHS) Uterine Fibroid Study.

Methods

Study population

The study population consisted of 984 premenopausal women (582 African Americans and 402 Caucasians) with available DNA samples for genotyping from the National Institute for Environmental Health Sciences (NIEHS) Uterine Fibroid Study (UFS). The parent study was designed to estimate the prevalence of uterine leiomyomata (fibroids) among African American and Caucasian women and evaluate potential etiologic factors for fibroids. Details of the parent study were previously described (29, 30). Briefly, a random sample of 2,384 George Washington University female health plan members, aged 35 to 49, was obtained for potential enrollment into the parent study (29, 30). The study was approved by Institutional Review Boards at NIEHS and George Washington University, and the consent form specified use of specimens for genetic polymorphism analyses.

UFS eligibility criteria were met by 1,786 of the 2,102 women that consented to eligibility screening. Most ineligible women were excluded because they no longer attended the health plan clinic where the parent study was based (71%) or they had been misidentified as a 35-49 year-old female (16%). Enrollment occurred from 1996 through 1999. Approximately 20% of eligible women refused participation, resulting in a total of 1,430 participants in the parent study (30). Demographic characteristics, reproductive history, smoking status, and alcohol use were assessed from telephone interviews and self-administered questionnaires. Body weight was measured at the clinic visit.

We restricted the current study population to women that self-identified as African American or Caucasian (n = 1,323) to facilitate race-specific analyses, and excluded

postmenopausal women because they did not attend the UFS clinic visit for ultrasound screening and blood collection (n = 178). Race and menopausal status criteria for the current study were met by 1,145 women, and DNA was extracted for 984 of the 1,003 women with collected blood samples.

Sample collection and assays

Fasting blood samples were collected through venipuncture and processed for plasma, which was stored at -80°C. Plasma IGF-I was measured in nanograms per milliliter (ng/mL) at NIEHS using a double-antibody radioimmunoassay by extraction method (Nichols Institute Diagnostics, San Juan Capistrano, CA), with a reported detection limit of 0.06 ng/mL. Plasma IGFBP-3 was measured in ng/mL at NIEHS by a double-antibody immunoradiometric assay from Diagnostic Systems Laboratories, Inc. (Webster, TX), with a reported detection limit of 0.05 ng/mL The mean inter-assay coefficients of variation on replicate quality control samples was 8.8% for IGF-I and 4.2% for IGFBP-3.

Genomic DNA was extracted from whole blood using a phenol:chloroform procedure, but a safer and more efficient modified salt precipitation protocol was adopted when it became available (GenQuik Protocol).

Genetic polymorphisms

Race-specific haplotype-tagging SNPs in *IGF-I* and *IGFBP-3* were selected using Genome Variation Server (GVS) software (sponsored by the Seattle SNPs Program for Genomic Applications (PGA)) (31). We used the Seattle SNPs database as the reference population for *IGF-I*, while for *IGFBP-3*, which had not been evaluated by Seattle SNPs, we used the HapMap database (31). We expanded coverage to include 5 kilobases (kb)

outside the 5' and 3' ends of each gene, and specified a value of 0.8 for the pairwise correlation coefficient (r^2) to identify haplotype-tagging SNPs to capture variation across the gene. Haplotype-tagging SNPs with a minor allele frequency (MAF) greater than 5% among women in African American or Caucasian reference populations were selected for genotyping. Overall, the GVS software identified 29 haplotype-tagging SNPs for *IGF-I* and 12 haplotype-tagging SNPs for *IGFBP-3*. In addition, we selected four SNPs *a priori* based on functional significance, including one nonsynonymous *IGF-I* SNP (rs17884626), one synonymous *IGF-I* SNP (rs3729846), and two nonsynonymous *IGFBP-3* SNPs (rs2854746, rs9282734) and an *IGFBP-3* promoter SNP (-202 A/C, rs2854744) *a priori* based on previous studies of associations with circulating IGF-I and IGFBP-3 levels and health outcomes (28, 32-38).

Genotyping was performed using the TaqMan genotyping approach (39-41) at an outside laboratory (Mammalian Genotyping Core, Lineberger Comprehensive Cancer Center; Chapel Hill, NC). Allele-specific oligonucleotide probes for 39 selected SNPs were purchased from Applied Biosystems (ABI; Foster City, CA) "TaqMan® Validated and Coding SNP or Pre-Designed SNP Genotyping Assays". ABI attempted to develop custom assays for the 6 remaining SNPs through their "Custom TaqMan® SNP Genotyping Assays" service. Two *IGF-I* haplotype tagging SNPs were dropped from analyses, including one for which a custom assay could not be developed, and one with a pre-designed assay that did not meet ABI technical specifications. In addition, we genotyped an alternate *IGF-I* haplotype-tagging SNP to substitute one that appeared inconsistent with Hardy-Weinberg equilibrium among our African American study

population. Thus, we genotyped 30 *IGF-I* and 15 *IGFBP-3* SNPs, including 40 haplotype-tagging SNPs and 5 SNPs selected *a priori*.

PCR amplification was performed on an ABI GeneAmp® PCR System 9700 thermal cycler with dual 384-well-blocks, and endpoint plates were read using the ABI 7900HT system. Fluorescent VIC and 6-FAM reporter dyes were used to distinguish wild type and variant alleles. Alleles were called automatically through Sequence Detection System (SDS) 2.3 software, but output also was reviewed by experienced operators. The DNA concentration of all samples was validated using a NanoDrop® ND-1000 Spectrophotometer prior to dilution to 5 ng/ul (using DNA grade sterile water). All samples were placed into eleven 96-well microtiter plates, with each plate containing two blank samples and two known DNA standard (Control DNA CEPH Individual 1347-02, ABI) samples. Samples were subsequently aliquoted into three 384-well PCR plates for analysis. Quality control measures also included blinded genotyping of 28 duplicate samples representing 22 women, which produced concordant results for all samples. The overall call rate was 98.8%, and only 5 women had less than 50% of complete allele calls for the 45 SNPs assayed. We confirmed that SNP genotype frequencies were consistent with Hardy-Weinberg equilibrium (HWE) within each racial group using the exact test statistic with one degree of freedom (alpha = 0.01) (42).

Diplotype estimation

We estimated associations with race-specific diplotypes (paired haplotypes), that were imputed for groups of related SNPs within each gene, as described below. SNPs were excluded from race-specific diplotype analyses if their MAF in our study population was below 5% for haplotype-tagging SNPs or below 3% for *a priori* SNPs within the

racial group being evaluated. Women missing genotype data for more than 50% of the SNPs considered for diplotype analyses within a gene were excluded from diplotype estimation for that gene (1 Caucasian and 3 African Americans for *IGFBP-3* analyses, 3 Caucasians and 2 African Americans for *IGF-I* analyses). We examined race-specific linkage disequilibrium (LD) patterns using Haploview software (43) to identify SNPs in each gene that could be combined for estimating diplotypes. First, we identified blocks consisting of individual SNPs (with MAF at least 5%) in strong LD (95% of pairwise SNP comparisons with one-sided 95% confidence intervals for the D prime statistic within 0.7 to 0.98) (44). Next, we used the Tagger approach (45) in Haploview to identify pairs of redundant SNPs in strong LD (pairwise r^2 values of at least 0.8), and excluded one member of each redundant pair from diplotype estimation unless both SNPs were selected *a priori*. Race-specific pairwise r^2 values are available for *IGFBP-3* and *IGF-I* SNPs in Tables A.1-A.4.

Race-specific diplotypes representing defined groups of SNPs in each gene were then estimated using PHASE version 2.1 (46, 47), which allocates the most likely diplotypes for each person, with the prior assumption that frequently observed haplotypes with less ambiguity due to homozygosity are more probable. PHASE also provides a posterior probability estimate that expresses the uncertainty associated with each diplotype assignment. To reduce the number of race-specific diplotype groups evaluated for each gene, we combined individual SNPs that were not included in a block with an adjacent block, and combined adjacent blocks with each other, when doing so resulted in diplotypes that were estimated with at least 90% certainty (posterior probability) for at least 90% of observations. Otherwise, diplotype groups encompassed individual SNPs

not included in a block, or SNPs within a single block only. Women were assigned their most probable diplotype for each group, or were classified as missing for the diplotype group if their most probable diplotype had a posterior probability below 90%.

Statistical analysis

All statistical analyses were stratified by race and conducted using SAS V9.1 (SAS Institute Inc., Cary, NC). We used ordinary linear regression to estimate associations between *IGF-I* and *IGFBP-3* gene variants (individual SNPs and race-specific diplotypes) and IGF-I and IGFBP-3 plasma levels. Because of biological interrelations between IGF-I and IGFBP-3, variants in each gene were analyzed for associations with circulating levels of both proteins. Estimated associations are unadjusted since there are no known factors other than race that would predict both plasma levels and gene variants.

For individual SNP analyses, we first considered a codominant model of inheritance in which heterozygous and homozygous variant genotypes were separately evaluated relative to the homozygous wildtype (referent) genotype (most common for specific race). Codominant models make the fewest assumptions with regard to inheritance pattern, but were feasible only for common variants. Therefore, when there were 10 or fewer women with the homozygous variant genotype, we assumed a dominant model that compared a single index category, including homozygous variant and heterozygous genotypes, to the homozygous wildtype genotype. To compare race-specific associations with plasma IGF-I or IGFBP-3 levels, we combined data from both racial groups and estimated racial differences in mean plasma levels (i.e., differences of mean differences) associated with variant vs. referent genotypes. Specifically, we fit linear regression

models that included multiplicative gene by race interaction terms along with separate parameters for race and genotypes (Tables A.5-A.6).

To estimate diplotype associations with IGF-I or IGFBP-3 plasma levels, we used separate race-specific models for each diplotype group, with the most common diplotype as the reference category. Diplotypes assigned to 5 or fewer women were combined into one "rare diplotype" category. To enhance the validity and precision of regression estimates, we used an empirical-Bayes method of information-weighted averaging (48). Specifically, we assumed a prior mean of 0, since we did not have prior information to group diplotypes according to the anticipated direction or strength of potential associations with plasma levels. We specified a prior variance corresponding to +/- one standard deviation $(2*\text{standard deviation}/3.92)^2$ of the mean plasma level of IGF-I (prior variances: African Americans, 1,419; Caucasians, 901) and IGFBP-3 (prior variances: African Americans, 186,819; Caucasians, 174,161) in the study population. This method shrinks regression estimates toward the prior mean such that imprecise estimates based on smaller numbers of observations move further toward the prior mean than more precise regression estimates. We applied the shrinkage estimator for each diplotype and report posterior medians (50^{th} percentile of the posterior probability distribution) and 95% posterior limits (2.5^{th} and 97.5^{th} percentiles of the posterior probability distribution). Regression estimates and 95% confidence limits estimated directly by linear regression are available in Tables A.7-A.14.

Results

Participant characteristics

Mean plasma IGF-I and IGFBP-3 levels, as well as characteristics of the study population are displayed in Table 4.1 by race. African Americans were less likely than Caucasians to have a college or graduate degree (33% vs. 87%), to report regular alcohol consumption (at least 0.5 drinks per week in the past year, 42% vs. 78%), or to report being nulliparous at UFS enrollment (21% vs. 59%). African Americans were more likely than Caucasians to be overweight or obese (75% vs. 41%) and to report current smoking (30% vs. 8%). Few women in either racial group were currently taking oral contraceptives.

IGFBP-3 SNPs

For plasma IGFBP-3, we emphasize associations where mean plasma levels estimated for the index genotypes are approximately 200 ng/mL higher or lower than the mean level estimated for the referent genotype (i.e., roughly +/- 5% of the estimated mean level for the referent group, which ranged from about 3798 to 4693 ng/mL). However, we do not consider imprecise associations with SNPs having ten or fewer observations with heterozygous and homozygous variant genotypes combined.

Ten *IGFBP-3* SNP variants (rs903889, rs924140, rs2854744, rs2854746, rs2471551, rs3110697, rs2453840, rs2453839, rs2270628, rs12671457) were associated with plasma IGFBP-3 levels in at least one racial group when compared with homozygous wildtype genotypes. Among Caucasians, plasma IGFBP-3 levels were elevated in association with variant genotypes for three SNPs (rs924140, rs2854744, rs2854746) relative to reference genotypes (approximately 600 ng/mL higher for homozygous variants, almost

400 ng/mL higher for heterozygotes) (Table 4.2). Pairwise r^2 values for all three SNPs were at least 0.8, indicating strong LD. These three SNPs were also associated with plasma IGFBP-3 levels among African Americans. Specifically, plasma levels were about 900 ng/mL higher in association with the CC genotype for the nonsynonymous rs2854746 SNP compared to the GG genotype, and about 460 ng/mL higher in association with the CG genotype (Table 4.3). Plasma IGFBP-3 levels in African Americans were higher in association with the rs924140 TT and rs2854744 AA genotypes corresponding to the homozygous variant genotypes in Caucasians, though estimated associations were inverse because these genotypes were the homozygous wildtype (referent) genotypes among African Americans. As in Caucasians, rs924140 and rs2854744 were in strong LD ($r^2 = 0.82$); however, rs2854746 was not in strong LD with either of these SNPs (r^2 range 0.30 to 0.34) among African Americans.

In both racial groups, rs3110697 variant genotypes were inversely associated with plasma IGFBP-3 relative to the referent genotypes (about 430 ng/mL lower in Caucasians and 550 ng/mL lower in African Americans for the AA genotype, with smaller differences associated with the AG genotype). Among both racial groups, r^2 values for rs3110697 with rs924140 and rs2854744 ranged from 0.55 to 0.65, indicating moderate LD, but rs3110697 was not in LD with rs2854746 ($r^2 < 0.5$). Plasma IGFBP-3 was also inversely associated with rs2471551 variants among African Americans (460 ng/mL lower for the CC genotype and 260 ng/mL lower for the CG genotype relative to the GG genotype).

Relatively imprecise estimates based on fewer than 25 women suggest inverse associations between plasma IGFBP-3 and homozygous variant genotypes for five SNPs

(rs903889, rs2471551, rs2453840, rs2453839, rs2270628) among Caucasians. Among African Americans, homozygous and heterozygous variants combined for two less common SNPs were positively (rs2453840) and inversely (rs12671457) associated with plasma IGFBP-3, with differences of about 200 ng/mL relative to homozygous wildtype genotypes. None of the other *IGFBP-3* SNPs with plasma IGFBP-3 associations were in LD with each other, except for rs2453840 and rs2453839, which were in strong LD among Caucasians ($r^2 = 0.87$).

Based on combined race models with interaction terms, race-specific SNP associations with plasma IGFBP-3 showed little evidence of discordance by race (Table A.5). Two possible exceptions were rs2854746, which was more strongly associated with plasma IGFBP-3 among African Americans than Caucasians, and rs2270628, which had estimated race-specific associations with the homozygous variant that were imprecise and in opposite direction. However, estimated differences of mean differences between racial groups, especially for homozygous variant genotypes, were relatively imprecise.

For plasma IGF-I, we emphasize associations where mean plasma levels estimated for the index genotypes are approximately 16-18 ng/mL higher or lower than the mean level estimated for the referent genotype (i.e., about +/- 10% of the estimated mean level for the referent group, which ranged from 158 to 176 ng/mL). Homozygous variants for two *IGFBP-3* SNPs (rs903889, rs6670) among Caucasians and four *IGFBP-3* SNPs (rs924140, rs2471551, rs3110697, rs2270628) among African Americans were associated with plasma IGF-I relative to the homozygous wildtype genotypes. However, three of the six SNP associations (rs903889, rs6670, rs2475551) with plasma IGF-I were relatively imprecise due to fewer than 30 homozygous variant observations.

IGFBP-3 Diplotypes

Among Caucasians, there were three LD blocks representing 10 of 12 *IGFBP-3* SNPs for diplotype analyses, with two SNPs outside LD blocks. After excluding two redundant *IGFBP-3* SNPs ($r^2 \geq 0.8$), two groups were created for diplotype estimation by combining LD blocks or SNPs outside LD blocks as described previously. Among the two Caucasian *IGFBP-3* groups, there were 48 unique diplotypes, with 20 of these diplotypes classified as rare (5 or fewer women). Diplotypes for each group were estimated for 93% to 98% of the 401 Caucasians for *IGFBP-3* diplotype analyses.

Among African Americans, there were four LD blocks representing 11 of 14 *IGFBP-3* SNPs for diplotype analyses, with three SNPs outside LD blocks. After excluding one redundant *IGFBP-3* SNP ($r^2 \geq 0.8$), three groups were created for diplotype estimation. Among the three African American *IGFBP-3* groups, there were 71 unique diplotypes, which included 33 rare diplotypes. Diplotypes for each group were estimated for 94% to 99% of the 579 African Americans for *IGFBP-3* diplotype analyses.

Similar to individual SNP analyses, we emphasized associations where the index diplotypes were associated with plasma IGFBP-3 differences of at least 200 ng/mL from the mean IGFBP-3 level for the referent diplotype. Six Caucasian *IGFBP-3* group 1 diplotypes (GCGGAC/GCGGAC, TAGGGC/GCGGAC, TCGCAC/GCGGAC, TCGCAC/TCGCAC, TCGGGA/GCGGAC, TCGGGA/TCGCAC) were inversely associated with plasma IGFBP-3 relative to the TACGGC/GCGGAC (referent) diplotype (Figure 4.1). All six diplotypes included the GG genotype for rs2854746 (3rd diplotype position), and five of the diplotypes included the CC genotype for rs2854744 (2nd diplotype position). The only two Caucasian *IGFBP-3* group 1 diplotypes

(TACGGA/TACGGC, TACGGC/TACGGC) having the CC genotype for rs2854746 were positively associated with plasma IGFBP-3.

Eight of the ten African American *IGFBP-3* group 1 diplotypes (TACGAGG/GCGGAGA, TAGGAGA/TAGGAGG, TAGGAGG/GCGGAGA, TAGGAGG/TAGGAGG, TCGCAGA/GCGGCAA, TCGCAGA/TAGGAGA, TCGCAGA/TAGGAGG, TCGCAGA/TCGCAGA, TCGCAGA/TCGGAGG, TCGGAGA/TACGAGG) that were inversely associated with plasma IGFBP-3 relative to the TAGGAGG/TACGAGG diplotype (Figure 4.2) included the GG genotype for rs2854746 (3^{rd} diplotype position). Three of the ten diplotypes also included the CC genotype for rs2854744 (2^{nd} position), and one was the only diplotype that included the CC genotype for rs2471551 (4^{th} diplotype position). The African American *IGFBP-3* group 1 diplotype TACGAGG/TACGAGG, the only diplotype with the CC genotype for rs2854746, was positively associated with plasma IGFBP-3. The African American *IGFBP-3* group 2 diplotype CC/CC was inversely associated with plasma IGFBP-3 compared to the CT/CC (referent) diplotype.

Similar to individual SNP analyses, we generally emphasized associations where the index diplotypes were associated with plasma IGF-I differences of at least 10% (16-18 ng/mL) from the mean IGF-I level for the referent diplotype. One Caucasian *IGFBP-3* group 1 diplotype (TCGGGA/TCGCAC) was inversely associated with plasma IGF-I compared to the TACGGC/GCGGAC diplotype. Among African Americans, eight *IGFBP-3* Group 1 diplotypes (TAGGAGA/TACGAGG, TAGGAGA/TAGGAGG, TAGGAGG/GCGGAGA, TCGCAGA/GCGGCAA, TCGCAGA/TACGAGG, TCGCAGA/TCGCAGA, TCGCAGA/TCGGAGG, TCGGAGG/TACGAGG) were

inversely associated with plasma IGF-I, and two *IGFBP-3* Group 1 diplotypes (TACGAGG/GCGGAGA, TCGGAGG/TCGGAGG) were positively associated with plasma IGF-I relative to the TAGGAGG/TACGAGG diplotype. Seven of the eight African American *IGFBP-3* Group 1 diplotypes that were inversely associated with plasma IGF-I included AG or AA genotypes for rs3110697 (7^{th} diplotype position), which were inversely associated with plasma IGF-I in SNP analyses. Both African American *IGFBP-3* group 3 diplotypes (TATA/TGTC, TATA/TATA) that included the rs2270628 TT genotype (3^{rd} diplotype position) were positively associated with plasma IGF-I relative to the TATA/TGCA diplotype.

IGF-I SNPs

Among Caucasians, homozygous variants for two common *IGF-I* SNPs (rs1520220, rs6214) and variants for five rare (MAF < 5%) *IGF-I* SNPs (rs5742612, rs5742614, rs5742657, rs5742692, rs3730204) were associated with plasma IGF-I, with estimated mean differences of at least 10% (17 ng/mL) relative to the homozygous wildtype genotypes (Table 4.4). However, with the exception of the positive association between rs6214 and plasma IGF-I, associations were rather imprecise due to 30 or fewer observations with variant genotypes. None of the *IGF-I* SNPs noted above were in LD ($r^2 < 0.4$) except for strong LD between rs5742657 and rs5742692 ($r^2 = 0.94$).

Among Caucasians, nine *IGF-I* SNP variants were inversely associated (rs4764884, rs5009837, rs4764883, rs9308315, rs978458, rs1520220, rs3730204, rs6219, rs2946834) with plasma IGFBP-3, while one rare (MAF < 5%) *IGF-I* SNP variant (rs5742612) was positively associated with plasma IGFBP-3 (estimated differences of at least 200 ng/mL for variant genotypes relative to homozygous wildtype genotypes), although five SNP

associations (rs5742612, rs4764884, rs9308315, rs978458, rs1520220) were relatively imprecise due to 30 or fewer observations with variant genotypes. Five (rs4764884, rs5009837, rs4764883, rs9308315, rs978458) of the ten *IGF-I* SNPs associated with plasma IGFBP-3 were in strong LD with each other ($r^2 \geq 0.8$), and rs1520220 was in moderate LD with rs978458 and rs9308315 (r^2 approximately 0.60) but had slightly smaller r^2 values of about 0.50 with rs4764884, rs5009837, and rs4764883.

Among African Americans, three rare (MAF < 5%) *IGF-I* (rs2033178, rs17727841, rs11111262) variants were associated with plasma IGF-I relative to the homozygous wildtype genotypes, although the rs11111262 association was relatively imprecise due to fewer than 30 variant observations (Table 4.5). In addition, variants for two more common *IGF-I* SNPs (rs6219, rs2946834) were positively associated with plasma IGF-I. Of the five *IGF-I* SNPs noted above, only rs17727841 and rs11111262 were in moderate LD with each other ($r^2 = 0.56$). The rs1086869 TT genotype was positively associated with plasma IGFBP-3 relative to the AA genotype, but there were no other *IGF-I* SNP variants associated with meaningful differences (≥ 200 ng/mL) in plasma IGFBP-3.

Based on models that included an interaction term with race, mean plasma IGF-I levels estimated for several of the *IGF-I* homozygous variants varied between racial groups by at least 16-18 ng/ml, which may indicate modification by race. However, estimated differences of mean differences between African Americans and Caucasians were imprecise (Table A.6)

IGF-I Diplotypes

Among Caucasians, there were three LD blocks representing 15 of 17 *IGF-I* SNPs for diplotype analyses, with two SNPs outside LD blocks. After excluding six redundant

IGF-I SNPs ($r^2 \geq 0.8$), four groups were created for diplotype estimation by combining LD blocks or SNPs outside LD blocks as described previously. Among the four Caucasian *IGF-I* groups, there were 40 unique diplotypes, with 13 of these diplotypes considered rare (5 or fewer observations). Diplotypes for each group were estimated for 97% to 99% of the 399 Caucasians for *IGF-I* diplotype analyses.

Among African Americans, there were four LD blocks representing 16 of 20 *IGF-I* SNPs for diplotype analyses, with four SNPs outside LD blocks. After excluding six redundant *IGF-I* SNPs ($r^2 \geq 0.8$), four groups were created for diplotype estimation. Among the four African American *IGF-I* groups, there were 73 unique diplotypes, with 25 of these diplotypes considered rare. Diplotypes for each group were estimated for 95% to 99% of the 580 African Americans for *IGF-I* diplotype analyses.

Five Caucasian *IGF-I* group 3 diplotypes (GCTCT/GCTGC, GGCCC/AGTGC, GGCCT/AGTGC, GGCCT/GGCCT, GGCCT/GGTGT) were positively associated with plasma IGF-I, with differences of at least 10% (17 ng/mL) higher than estimated mean levels for the referent GGCCC/GGCCT diplotype (Figure 4.3). However, the associations with the GCTCT/GCTGC and GGCCT/GGTGT diplotypes were relatively imprecise. The *IGF-I* group 3 GGCCT/GGCCT diplotype was positively associated with plasma IGFBP-3 (mean difference of at least 200 ng/mL relative to referent diplotype). *IGF-I* group 4 diplotypes (CA/TA and rare (TA/TA), CA/CA) that included the AA genotype for rs2946834 (2^{nd} diplotype position) were inversely associated with plasma IGFBP-3 relative to the mean level for the referent CG/CG diplotype. The *IGF-I* group 4 diplotype category (CA/TA and TA/TA diplotypes) also was inversely associated with plasma IGF-I.

The African American *IGF-I* group 1 GGG/GCG diplotype was positively associated with plasma IGF-I relative to the AGG/GGG diplotype, and an imprecise inverse association based on fewer than ten observations was noted between the GCG/GCG diplotype and plasma IGF-I (Figure 4.4). Four *IGF-I* group 2 diplotypes (ATCCC/TCTTT, ATTCC/TCTCC, ATTCC/TCTCT, TTTCT/TCTTT) and two *IGF-I* group 3 diplotypes (AAGC/AACT, TACC/TACC) were associated with plasma IGF-I relative to the referent diplotypes although estimates were imprecise since there were 15 or fewer observations with index diplotypes. In addition, the *IGF-I* group 4 CA/TA diplotype was positively associated with plasma IGF-I relative to the referent CA/CG diplotype. Two *IGF-I* group 2 diplotypes (ATTCC/TCTCT, TTTCT/TCTTT) and four *IGF-I* group 3 diplotypes (AACC/AGGT, AACC/TACC, AAGC/AACT, TACC/TACC) were associated with plasma IGFBP-3, relative to the referent diplotypes, but estimates were imprecise due to 15 or fewer observations with index diplotypes. None of the associations between plasma IGFBP-3 or IGF-I levels and African American *IGF-I* diplotypes were explained by patterns with SNP associations.

Discussion

Overall, we noted several associations between individual *IGF-I* and *IGFBP-3* SNPs and their respective protein levels. In addition, we detected associations between *IGFBP-3* SNPs and plasma IGF-I levels and associations between *IGF-I* SNPs and IGFBP-3 plasma levels, predominantly among Caucasians. The strongest evidence of association between genetic variation and circulating protein levels was between the nonsynonymous *IGFBP-3* SNP, rs2854746, and plasma IGFBP-3. In both races, the CC genotype was

associated with higher estimated mean plasma levels than the GG genotype, with intermediate mean levels for those with the CG genotype. In addition, *IGFBP-3* diplotypes with the rs2854746 GG genotype had lower mean plasma IGFBP-3 levels, and *IGFBP-3* diplotypes with the CC genotype had higher mean plasma IGFBP-3 levels relative to referent diplotypes with the CG genotype in both racial groups.

Among both racial groups, estimated increases in plasma IGFBP-3 levels were similar in association with AA versus CC genotypes for s2854744, though estimated associations with diplotypes that included this variant were not as consistent as for rs2854746. Associations between rs2854744 and plasma IGFBP-3 in Caucasians might be explained by strong LD between rs2854744 and rs2854746 ($r^2 = 0.82$); however, these two SNPs were not in LD among African Americans ($r^2 = 0.34$). Alternatively, estimated plasma IGFBP-3 associations with rs2854744 might be explained by effects of rs924140, which was excluded from Caucasian and African American diplotypes, but was in LD with rs2854744 among women in both racial groups (Caucasians: $r^2 = 0.99$; African Americans: $r^2 = 0.82$). However, no functional effects of rs924140 have been reported in the literature.

Biologic evidence supports a causal relation of rs2854746 with plasma IGFBP-3 levels, since this SNP results in an amino acid change from alanine to glycine. Evidence from protein sequence analysis reported in the UniProtKB/Swiss-Prot database (49) suggests that the amino acid coded by the rs2854746 SNP is within the region responsible for IGF binding. The Multiethnic Cohort Study (28) and a small study of controls from a population-based cancer case-control study (50) both reported associations between rs2854746 and plasma IGFBP-3 in the same direction to our study findings.

Other epidemiologic studies have not evaluated rs2854746, but several have examined rs2854744 and plasma IGFBP-3 predominantly among Caucasians. Seven studies reported higher circulating IGFBP-3 among women with two copies of the A allele compared with women who had no A alleles, and intermediate levels among those with AC genotype (32-38), including a study of Physicians' Health Study participants by Deal et al. (33), who noted that rs2854744 (-202 A/C SNP) had the strongest correlation with plasma IGFBP-3 of the five evaluated *IGFBP-3* promoter SNPs. In addition, Deal et al. reported that promoter activity was increased *in vitro* in association with the rs2854744 A allele (33), indicating greater IGFBP-3 protein production. However, strong LD with rs2854746 may be a possible causal explanation for associations between rs2854744 and plasma IGFBP-3 among Caucasians. As in our study population, the Multiethnic Cohort Study reported strong LD (r^2 value ≥ 0.8) for rs2854744 and rs2854746 among Caucasians but not African Americans (28). In addition, they noted that only rs2854746 was associated with plasma IGFBP-3 levels after accounting for both SNPs in their analytic model (142)(28), though sparse data limited inferences across the five race/ethnicity groups included in the cohort. Similar to our study, the Multiethnic Cohort Study also reported decreased plasma IGFBP-3 levels in association with rs3110697 AA vs. GG genotypes among both races, and that rs3110697 was not in strong LD with rs2854744 or rs2854746 (28).

We also reported consistent inverse associations between *IGFBP-3* rs2471551 variants and plasma IGFBP-3 among both racial groups. This SNP has potential functional relevance as it is located at a splice site (intron 1 within less than 20 kb of exon 2). In addition, Canzian et al. (22) reported that rs2471551 (homozygous variant versus

homozygous wildtype genotypes) was inversely associated with circulating IGFBP-3 for a predominantly Caucasian case-control study nested within the European Prospective Investigation into Cancer and Nutrition.

The focus of many prior epidemiologic studies of *IGF-I* has been the dinucleotide CA repeat polymorphism (position -969) located in the promoter approximately one kb upstream of the transcription site; however, relations between CA repeat polymorphisms and circulating IGF-I levels have been inconsistent (27, 35-36, 51-58). Methodological differences in the categorization of repeat genotypes and the potential for substantial misclassification during genotyping (59) make it difficult to compare results across studies. Although we did not evaluate this repeat polymorphism, we evaluated *IGF-I* SNPs within five kb of the 5' and 3' ends of the gene.

Our study estimated higher mean plasma IGF-I levels in association with the rs6214 TT genotype relative to the CC genotype among Caucasians and found no clear associations between rs35767 and plasma IGF-I or IGFBP-3 among either racial group. In contrast, Canzian, et al. measured five *IGF-I* SNPs in a predominantly Caucasian study population and reported associations for rs35767 with circulating IGF-I and IGFBP-3 but no associations with rs6214 (32).

We also estimated an inverse association between plasma IGF-I and the rs1520220 homozygous variant relative to the homozygous wildtype genotype among Caucasians, but noted only a weak positive association among African Americans. We also noted a positive association between rs2946834 AA versus GG genotypes and plasma IGF-I among African Americans, but a weak inverse association among Caucasians although race-specific associations may not be meaningfully different based on imprecision with

estimated difference of mean differences. Al-Zahrani et al. also reported no association between rs6214 and plasma IGF-I, but in contrast with our study, they noted a positive association with homozygous variant genotypes relative to homozygous wildtype genotypes for rs2946834 and rs1520220 (34). Although both SNPs were tightly correlated in this predominantly Caucasian study population, a positive association remained with rs1520220 only after adjustment for rs2946834 (34).

Our study had a very high call rate for genotyping. We evaluated *IGF-I* or *IGFBP-3* variants in association with plasma levels of both proteins due to biological interrelations between IGF-I and IGFBP-3. Previous studies have evaluated *IGF-I* or *IGFBP-3* variants in association with their respective protein levels only, except two studies examined associations with both proteins including the Canzian et al. evaluation of several *IGF-I* and *IGFBP-3* variants within a Caucasian study population (32) and one study that only evaluated the *IGF-I* dinucleotide CA repeat polymorphism and *IGFBP-3* rs2854746 (50).

Diplotype estimation may have been biased by the removal of women that had a lower degree of certainty (posterior probability < 90%) for assigned diplotypes, though this bias should be negligible given that 1% to 7% of women were excluded for race-specific analyses of each diplotype group. Rare diplotypes were assigned with lower certainty, as the PHASE software assumes that frequently observed haplotypes with less ambiguity are more probable. We did not attempt to estimate associations with rare diplotypes (5 or fewer observations in our study population) within each group individually, but instead combined them into a single category. Rare diplotype categories were associated with plasma IGF-I and IGFBP-3 levels in some instances, but these

associations could not be interpreted due to the heterogeneity of the diplotypes included in these categories.

We used a method of information-weighted averaging that intentionally biased estimates towards the null, since we assumed a null value for our prior mean; however, this approach increased the precision of estimates, particularly diplotypes assigned to small numbers of women. We selected a prior mean of zero since we had no prior information to indicate functional similarities between diplotypes and direction of associations with plasma IGF-I and IGFBP-3 levels relative to referent diplotypes. Despite the increase in bias for posterior medians using this approach, a reduction in the overall mean square error based on a greater decrease in variance of estimates has been shown with simulation studies and an occupational cohort study (60).

Possible selection bias exists if women excluded from the study because they lacked DNA for genotyping (161 total, 14% of the eligible study population) differed from the remaining study population with respect to their genotypes or plasma IGF-I or IGFBP-3 levels. A slightly greater proportion of the African Americans (108 total, 16%) in the eligible study population were excluded than Caucasians (53 total, 12%). The parent study obtained only one measurement of plasma IGF-I and IGFBP-3 from study participants, which may not reflect intra-individual variation; however, the Nurses' Health Study reported high correlations (> 0.8) for plasma IGF-I and IGFBP-3 measurements across premenopausal women over time (61). Age and sex are strong predictors of circulating IGF-I and IGFBP-3, but these factors were unlikely to influence our results since our study population was restricted to women within a narrow 15-year

age range (35 to 49 years) and adjusting for age did not affect our results (data not shown).

We restricted our study to Caucasians and African Americans and stratified all analyses by race; however, population stratification within each racial group is still a potential limitation of our study (62-65). Several of the reported associations may reflect random error or bias; however, at least one (between the nonsynonymous *IGFBP-3* SNP (rs2854746) and plasma IGFBP-3) showed evidence of a dose-response effect that was consistent for the individual SNP and diplotype analyses among both racial groups. It will be important to validate this finding, as well as other results, in additional study populations.

A major strength of our study was the large number of African American participants, which allowed us to expand beyond previous research that has focused almost exclusively on relations between *IGF-I* and *IGFBP-3* SNPs and their respective protein levels in Caucasians. Because African Americans have more genetic heterogeneity than Caucasians, the frequency of etiologically relevant SNPs may differ, and may at least partly explain racial disparities in the burden of cancer and cardiovascular disease. Therefore, assessing IGF-I and IGFBP-3 variants that predict circulating IGF-I and IGFBP-3 is important to improve our understanding of the potential biologic role of IGF-I and IGFBP-3 in the etiology of common diseases.

REFERENCES

1. Jones, J. I., and Clemmons, D. R. Insulin-like growth factors and their binding proteins: biological actions. Endocr Rev, *16:* 3-34, 1995.

2. Hwa, V., Oh, Y., and Rosenfeld, R. G. The insulin-like growth factor-binding protein (IGFBP) superfamily. Endocr Rev, *20:* 761-87, 1999.

3. Juul, A., Main, K., Blum, W. F., Lindholm, J., Ranke, M. B., and Skakkebaek, N. E. The ratio between serum levels of insulin-like growth factor (IGF)-I and the IGF binding proteins (IGFBP-1, 2 and 3) decreases with age in healthy adults and is increased in acromegalic patients. Clin Endocrinol (Oxf), *41:* 85-93, 1994.

4. Gill, Z. P., Perks, C. M., Newcomb, P. V., and Holly, J. M. Insulin-like growth factor-binding protein (IGFBP-3) predisposes breast cancer cells to programmed cell death in a non-IGF-dependent manner. J Biol Chem, *272:* 25602-7, 1997.

5. Schedlich, L. J., and Graham, L. D. Role of insulin-like growth factor binding protein-3 in breast cancer cell growth. Microsc Res Tech, *59:* 12-22, 2002.

6. Giovannucci, E. Insulin-like growth factor-I and binding protein-3 and risk of cancer. Horm Res, *51 Suppl 3:* 34-41, 1999.

7. Renehan, A. G., Zwahlen, M., Minder, C., O'Dwyer, S. T., Shalet, S. M., and Egger, M. Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis. Lancet, *363:* 1346-53, 2004.

8. Sandhu, M. S. Insulin-like growth factor-I and risk of type 2 diabetes and coronary heart disease: molecular epidemiology. Endocr Dev, *9:* 44-54, 2005.

9. Kaplan, R. C., Strickler, H. D., Rohan, T. E., Muzumdar, R., and Brown, D. L. Insulin-like growth factors and coronary heart disease. Cardiol Rev, *13:* 35-9, 2005.

10. Renehan, A. G., Harvie, M., and Howell, A. Insulin-like growth factor (IGF)-I, IGF binding protein-3, and breast cancer risk: eight years on. Endocr Relat Cancer, *13:* 273-8, 2006.

11. Rinaldi, S., Peeters, P. H., Berrino, F., Dossus, L., Biessy, C., Olsen, A., Tjonneland, A., Overvad, K., Clavel-Chapelon, F., Boutron-Ruault, M. C., Tehard, B., Nagel, G., Linseisen, J., Boeing, H., Lahmann, P. H., Trichopoulou, A., Trichopoulos, D., Koliva, M., Palli, D., Panico, S., Tumino, R., Sacerdote, C., van Gils, C. H., van Noord, P., Grobbee, D. E., Bueno-de-Mesquita, H. B., Gonzalez, C. A., Agudo, A., Chirlaque, M. D., Barricarte, A., Larranaga, N., Quiros, J. R., Bingham, S., Khaw, K. T., Key, T., Allen, N. E., Lukanova, A., Slimani, N., Saracci, R., Riboli, E., and Kaaks, R. IGF-I, IGFBP-3 and breast

cancer risk in women: The European Prospective Investigation into Cancer and Nutrition (EPIC). Endocr Relat Cancer, *13:* 593-605, 2006.

12. Baglietto, L., English, D. R., Hopper, J. L., Morris, H. A., Tilley, W. D., and Giles, G. G. Circulating insulin-like growth factor-I and binding protein-3 and the risk of breast cancer. Cancer Epidemiol Biomarkers Prev, *16:* 763-8, 2007.

13. Harrela, M., Koistinen, H., Kaprio, J., Lehtovirta, M., Tuomilehto, J., Eriksson, J., Toivanen, L., Koskenvuo, M., Leinonen, P., Koistinen, R., and Seppala, M. Genetic and environmental components of interindividual variation in circulating levels of IGF-I, IGF-II, IGFBP-1, and IGFBP-3. J Clin Invest, *98:* 2612-5, 1996.

14. Pantsulaia, I., Trofimov, S., Kobyliansky, E., and Livshits, G. Genetic regulation of the variation of circulating insulin-like growth factors and leptin in human pedigrees. Metabolism, *54:* 975-81, 2005.

15. Pantsulaia, I., Trofimov, S., Kobyliansky, E., and Livshits, G. Pedigree-based quantitative genetic analysis of interindividual variation in circulating levels of IGFBP-3. J Bone Miner Metab, *20:* 156-63, 2002.

16. LeRoith, D. Insulin-like growth factors. N Engl J Med, *336:* 633-40, 1997.

17. Kaklamani, V. G., Linos, A., Kaklamani, E., Markaki, I., and Mantzoros, C. Age, sex, and smoking are predictors of circulating insulin-like growth factor 1 and insulin-like growth factor-binding protein 3. J Clin Oncol, *17:* 813-7, 1999.

18. Holmes, M. D., Pollak, M. N., and Hankinson, S. E. Lifestyle correlates of plasma insulin-like growth factor I and insulin-like growth factor binding protein 3 concentrations. Cancer Epidemiol Biomarkers Prev, *11:* 862-7, 2002.

19. Lukanova, A., Toniolo, P., Akhmedkhanov, A., Hunt, K., Rinaldi, S., Zeleniuch-Jacquotte, A., Haley, N. J., Riboli, E., Stattin, P., Lundin, E., and Kaaks, R. A cross-sectional study of IGF-I determinants in women. Eur J Cancer Prev, *10:* 443-52, 2001.

20. Probst-Hensch, N. M., Wang, H., Goh, V. H., Seow, A., Lee, H. P., and Yu, M. C. Determinants of circulating insulin-like growth factor I and insulin-like growth factor binding protein 3 concentrations in a cohort of Singapore men and women. Cancer Epidemiol Biomarkers Prev, *12:* 739-46, 2003.

21. DeLellis, K., Rinaldi, S., Kaaks, R. J., Kolonel, L. N., Henderson, B., and Le Marchand, L. Dietary and lifestyle correlates of plasma insulin-like growth factor-I (IGF-I) and IGF binding protein-3 (IGFBP-3): the multiethnic cohort. Cancer Epidemiol Biomarkers Prev, *13:* 1444-51, 2004.

22. Harris, T. B., Kiel, D., Roubenoff, R., Langlois, J., Hannan, M., Havlik, R., and Wilson, P. Association of insulin-like growth factor-I with body composition, weight history, and past health behaviors in the very old: the Framingham Heart Study. J Am Geriatr Soc, *45:* 133-9, 1997.

23. Goodman-Gruen, D., and Barrett-Connor, E. Epidemiology of insulin-like growth factor-I in elderly men and women. The Rancho Bernardo Study. Am J Epidemiol, *145:* 970-6, 1997.

24. Johansson, H., Baglietto, L., Guerrieri-Gonzaga, A., Bonanni, B., Mariette, F., Macis, D., Serrano, D., Sandri, M. T., and Decensi, A. Factors associated with circulating levels of insulin-like growth factor-I and insulin-like growth factor binding protein-3 in 740 women at risk for breast cancer. Breast Cancer Res Treat, *88:* 63-73, 2004.

25. Morimoto, L. M., Newcomb, P. A., White, E., Bigler, J., and Potter, J. D. Variation in plasma insulin-like growth factor-1 and insulin-like growth factor binding protein-3: personal and lifestyle factors (United States). Cancer Causes Control, *16:* 917-27, 2005.

26. Chang, S., Wu, X., Yu, H., and Spitz, M. R. Plasma concentrations of insulin-like growth factors among healthy adult men and postmenopausal women: associations with body composition, lifestyle, and reproductive factors. Cancer Epidemiol Biomarkers Prev, *11:* 758-66, 2002.

27. Jernstrom, H., Chu, W., Vesprini, D., Tao, Y., Majeed, N., Deal, C., Pollak, M., and Narod, S. A. Genetic factors related to racial variation in plasma levels of insulin-like growth factor-1: implications for premenopausal breast cancer risk. Mol Genet Metab, *72:* 144-54, 2001.

28. Cheng, I., Henderson, K. D., Haiman, C. A., Kolonel, L. N., Henderson, B. E., Freedman, M. L., and Le Marchand, L. Genetic Determinants of Circulating IGF-I, IGFBP-1 and IGFBP-3 Levels in a Multiethnic Population. J Clin Endocrinol Metab, 2007.

29. Baird, D. D., Dunson, D. B., Hill, M. C., Cousins, D., and Schectman, J. M. High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence. Am J Obstet Gynecol, *188:* 100-7, 2003.

30. Wegienka, G., Baird, D. D., Hertz-Picciotto, I., Harlow, S. D., Steege, J. F., Hill, M. C., Schectman, J. M., and Hartmann, K. E. Self-reported heavy bleeding associated with uterine leiomyomata. Obstet Gynecol, *101:* 431-7, 2003.

31. Seattle SNPs Program for Genomic Applications (PGA). Genome Variation Server (GVS) Version 1.04. pp. http://gvs.gs.washington.edu/GVS/, 2006.

32. Canzian, F., McKay, J. D., Cleveland, R. J., Dossus, L., Biessy, C., Rinaldi, S., Landi, S., Boillot, C., Monnier, S., Chajes, V., Clavel-Chapelon, F., Tehard, B., Chang-Claude, J., Linseisen, J., Lahmann, P. H., Pischon, T., Trichopoulos, D., Trichopoulou, A., Zilis, D., Palli, D., Tumino, R., Vineis, P., Berrino, F., Bueno-de-Mesquita, H. B., van Gils, C. H., Peeters, P. H., Pera, G., Ardanaz, E., Chirlaque, M. D., Quiros, J. R., Larranaga, N., Martinez-Garcia, C., Allen, N. E., Key, T. J., Bingham, S. A., Khaw, K. T., Slimani, N., Norat, T., Riboli, E., and Kaaks, R. Polymorphisms of genes coding for insulin-like growth factor 1 and its major binding proteins, circulating levels of IGF-I and IGFBP-3 and breast cancer risk: results from the EPIC study. Br J Cancer, *94:* 299-307, 2006.

33. Deal, C., Ma, J., Wilkin, F., Paquette, J., Rozen, F., Ge, B., Hudson, T., Stampfer, M., and Pollak, M. Novel promoter polymorphism in insulin-like growth factor-binding protein-3: correlation with serum levels and interaction with known regulators. J Clin Endocrinol Metab, *86:* 1274-80, 2001.

34. Al-Zahrani, A., Sandhu, M. S., Luben, R. N., Thompson, D., Baynes, C., Pooley, K. A., Luccarini, C., Munday, H., Perkins, B., Smith, P., Pharoah, P. D., Wareham, N. J., Easton, D. F., Ponder, B. A., and Dunning, A. M. IGF1 and IGFBP3 tagging polymorphisms are associated with circulating levels of IGF1, IGFBP3 and risk of breast cancer. Hum Mol Genet, *15:* 1-10, 2006.

35. Jernstrom, H., Deal, C., Wilkin, F., Chu, W., Tao, Y., Majeed, N., Hudson, T., Narod, S. A., and Pollak, M. Genetic and nongenetic factors associated with variation of plasma levels of insulin-like growth factor-I and insulin-like growth factor-binding protein-3 in healthy premenopausal women. Cancer Epidemiol Biomarkers Prev, *10:* 377-84, 2001.

36. Lai, J. H., Vesprini, D., Zhang, W., Yaffe, M. J., Pollak, M., and Narod, S. A. A polymorphic locus in the promoter region of the IGFBP3 gene is related to mammographic breast density. Cancer Epidemiol Biomarkers Prev, *13:* 573-82, 2004.

37. Schernhammer, E. S., Hankinson, S. E., Hunter, D. J., Blouin, M. J., and Pollak, M. N. Polymorphic variation at the -202 locus in IGFBP3: Influence on serum levels of insulin-like growth factors, interaction with plasma retinol and vitamin D and breast cancer risk. Int J Cancer, *107:* 60-4, 2003.

38. Ren, Z., Cai, Q., Shu, X. O., Cai, H., Li, C., Yu, H., Gao, Y. T., and Zheng, W. Genetic polymorphisms in the IGFBP3 gene: association with breast cancer risk and blood IGFBP-3 protein levels among Chinese women. Cancer Epidemiol Biomarkers Prev, *13:* 1290-5, 2004.

39. Behrens, M., and Lange, R. A highly reproducible and economically competitive SNP analysis of several well characterized human mutations. Clin Lab, *50:* 305-16, 2004.

40. Livak, K. J. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. Genet Anal, *14:* 143-9, 1999.

41. McGuigan, F. E., and Ralston, S. H. Single nucleotide polymorphism detection: allelic discrimination using TaqMan. Psychiatr Genet, *12:* 133-6, 2002.

42. Wigginton, J. E., Cutler, D. J., and Abecasis, G. R. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet, *76:* 887-93, 2005.

43. Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics, *21:* 263-5, 2005.

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

45. de Bakker, P. I., Yelensky, R., Pe'er, I., Gabriel, S. B., Daly, M. J., and Altshuler, D. Efficiency and power in genetic association studies. Nat Genet, *37:* 1217-23, 2005.

46. Stephens, M., and Donnelly, P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet, *73:* 1162-9, 2003.

47. Stephens, M., Smith, N. J., and Donnelly, P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet, *68:* 978-89, 2001.

48. Greenland, S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. Int J Epidemiol, *35:* 765-75, 2006.

49. Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N., and Yeh, L. S. The Universal Protein Resource (UniProt). Nucleic Acids Res, *33:* D154-9, 2005.

50. Morimoto, L. M., Newcomb, P. A., White, E., Bigler, J., and Potter, J. D. Variation in plasma insulin-like growth factor-1 and insulin-like growth factor binding protein-3: genetic factors. Cancer Epidemiol Biomarkers Prev, *14:* 1394-401, 2005.

51. Rosen, C. J., Kurland, E. S., Vereault, D., Adler, R. A., Rackoff, P. J., Craig, W. Y., Witte, S., Rogers, J., and Bilezikian, J. P. Association between serum insulin growth factor-I (IGF-I) and a simple sequence repeat in IGF-I gene: implications

for genetic studies of bone mineral density. J Clin Endocrinol Metab, *83:* 2286-90, 1998.

52. Jernstrom, H., Sandberg, T., Bageman, E., Borg, A., and Olsson, H. Insulin-like growth factor-1 (IGF1) genotype predicts breast volume after pregnancy and hormonal contraception and is associated with circulating IGF-1 levels: implications for risk of early-onset breast cancer in young women from hereditary breast cancer families. Br J Cancer, *92:* 857-66, 2005.

53. Rietveld, I., Janssen, J. A., Hofman, A., Pols, H. A., van Duijn, C. M., and Lamberts, S. W. A polymorphism in the IGF-I gene influences the age-related decline in circulating total IGF-I levels. Eur J Endocrinol, *148:* 171-5, 2003.

54. Missmer, S. A., Haiman, C. A., Hunter, D. J., Willett, W. C., Colditz, G. A., Speizer, F. E., Pollak, M. N., and Hankinson, S. E. A sequence repeat in the insulin-like growth factor-1 gene and risk of breast cancer. Int J Cancer, *100:* 332-6, 2002.

55. Vaessen, N., Heutink, P., Janssen, J. A., Witteman, J. C., Testers, L., Hofman, A., Lamberts, S. W., Oostra, B. A., Pols, H. A., and van Duijn, C. M. A polymorphism in the gene for IGF-I: functional properties and risk for type 2 diabetes and myocardial infarction. Diabetes, *50:* 637-42, 2001.

56. DeLellis, K., Ingles, S., Kolonel, L., McKean-Cowdin, R., Henderson, B., Stanczyk, F., and Probst-Hensch, N. M. IGF1 genotype, mean plasma level and breast cancer risk in the Hawaii/Los Angeles multiethnic cohort. Br J Cancer, *88:* 277-82, 2003.

57. Wen, W., Gao, Y. T., Shu, X. O., Yu, H., Cai, Q., Smith, J. R., and Zheng, W. Insulin-like growth factor-I gene polymorphism and breast cancer risk in Chinese women. Int J Cancer, *113:* 307-11, 2005.

58. Giovannucci, E., Haiman, C. A., Platz, E. A., Hankinson, S. E., Pollak, M. N., and Hunter, D. J. Dinucleotide repeat in the insulin-like growth factor-I gene is not related to risk of colorectal adenoma. Cancer Epidemiol Biomarkers Prev, *11:* 1509-10, 2002.

59. Tran, N., Bharaj, B. S., Diamandis, E. P., Smith, M., Li, B. D., and Yu, H. Short tandem repeat polymorphism and cancer risk: influence of laboratory analysis on epidemiologic findings. Cancer Epidemiol Biomarkers Prev, *13:* 2133-40, 2004.

60. Greenland, S., and Poole, C. Empirical-Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. Arch Environ Health, *49:* 9-16, 1994.

61. Missmer, S. A., Spiegelman, D., Bertone-Johnson, E. R., Barbieri, R. L., Pollak, M. N., and Hankinson, S. E. Reproducibility of plasma steroid hormones, prolactin, and insulin-like growth factor levels among premenopausal women over a 2- to 3-year period. Cancer Epidemiol Biomarkers Prev, *15:* 972-8, 2006.

62. Wang, Y., Localio, R., and Rebbeck, T. R. Evaluating bias due to population stratification in case-control association studies of admixed populations. Genet Epidemiol, *27:* 14-20, 2004.

63. Wang, Y., Localio, R., and Rebbeck, T. R. Evaluating bias due to population stratification in epidemiologic studies of gene-gene or gene-environment interactions. Cancer Epidemiol Biomarkers Prev, *15:* 124-32, 2006.

64. Ardlie, K. G., Lunetta, K. L., and Seielstad, M. Testing for population subdivision and association in four case-control studies. Am J Hum Genet, *71:* 304-11, 2002.

65. Wacholder, S., Rothman, N., and Caporaso, N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst, *92:* 1151-8, 2000.

Table 4.1. Characteristics of premenopausal women with genotype information by race from NIEHS Uterine Fibroid Study

Characteristic	African Americans (N=582) N (%)	Caucasians (N=402) N (%)
Age (years)		
35-39	219 (37.6)	137 (34.1)
40-44	205 (35.2)	135 (33.6)
45+	158 (27.1)	130 (32.3)
Education		
High school	121 (20.8)*	12 (3.0)
Some postsecondary	265 (45.5)	33 (8.2)
College degree	123 (21.1)	133 (33.1)
Graduate degree	68 (11.7)	217 (54.0)
Missing	5 (0.9)	7 (1.7)
Current oral contraceptives use†	28 (4.8)	37 (9.2)
Parity		
0	120 (20.6)	236 (58.7)
1	134 (23.0)	54 (13.4)
2	194 (33.3)	91 (22.6)
3+	134 (23.0)	21 (5.2)
Body mass index		
Under- normal weight (<25)	147 (25.3)	236 (58.7)
Overweight (25-<30)	178 (30.6)	95 (23.6)
Obese (30+)	257 (44.2)	71 (17.7)
Smoking status		
Current smoker	172 (29.6)	31 (7.7)
Former smoker	133 (22.9)	139 (34.6)
Never smoked	277 (47.6)	232 (57.7)
Alcohol intake (past year)		
<0.5 drinks/week	300 (51.5)	60 (14.9)
≥0.5 drinks/week	245 (42.1)	314 (78.1)
Missing	37 (6.4)	28 (7.0)
	Mean (SD)	Mean (SD)
Plasma IGF-I (ng/mL)†	165 (74.0)	172 (58.7)
Plasma IGFBP-3 (ng/mL)†	4085 (859.7)	4524 (822.3)

* Includes 11 women with less than high school education.

† N missing: oral contraceptives: 4 African Americans, 3 Caucasians; IGF-I and IGFBP-3: 8 African Americans, 1 Caucasian.

Table 4.2. Unadjusted linear regression of *IGFBP-3* single nucleotide polymorphisms (SNPs) on plasma IGFBP-3 and IGF-I levels among Caucasians

Block*	Diplotype Group*	Diplotype Position*	SNP Reference ID	MAF (%)†	Genotype	N‡	Plasma IGFBP-3 (ng/mL)		Plasma IGF-I (ng/mL)	
							β§	95% CI§	β§	95% CI§
1	1	1	rs903889	22.0	GG	22	-398	-754, -42	21	-4, 47
					GT	131	-67	-240, 106	3	-10, 15
					TT	246	4572	4470, 4674	170	163, 178
1	--	--	rs924140	46.8	TT	82	581	352, 810	5	-12, 22
					CT	210	375	190, 560	6	-7, 20
					CC	107	4210	4059, 4361	168	157, 179
1	1	2	rs2854744‖	46.3	AA	79	592	364, 819	4	-13, 21
					AC	204	353	170, 536	6	-8, 19
					CC	107	4221	4073, 4370	169	158, 180
1	1	3	rs2854746‖	41.7	CC	65	631	398, 864	7	-10, 24
					CG	203	379	206, 551	5	-8, 18
					GG	130	4226	4091, 4360	169	158, 179
1	1	4	rs2471551	19.4	CC	17	-345	-745, 55	-9	-37, 20
					CG	119	-119	-296, 57	-4	-17, 8
					GG	260	4583	4484, 4682	174	167, 182
--	--	--	rs9282734‖	0.4	AC	3	698	-233, 1629	46	-20, 113
					AA	393	4520	4439, 4601	172	166, 178
--	--	--	rs2453837	0	GG	397	4524	4444, 4605	172	167, 178
--	--	--	rs6953668	0.6	AG	5	394	-330, 1119	33	-18, 85
					GG	394	4522	4441, 4603	172	166, 178
1	1	5	rs3110697	41.4	AA	62	-425	-670, -181	6	-11, 24
					AG	203	-188	-366, -10	-2	-15, 11
					GG	131	4693	4554, 4832	172	162, 182
2	1	6	rs2453840	18.3	AA	12	-535	-1008, -62	-3	-37, 31
					AC	121	-45	-221, 131	-6	-18, 7
					CC	263	4560	4461, 4658	174	167, 182

2	--	--	rs2453839	20.1	CC	14	-325	-760, 111	-5	-36, 26
					CT	131	-114	-285, 57	-10	-22, 2
					TT	250	4574	4474, 4675	176	169, 183
Outside	2	1	rs6670	21.2	AA	12	74	-402, 551	-22	-56, 12
					AT	145	79	-90, 249	-2	-14, 10
					TT	240	4494	4390, 4598	174	166, 181
Outside	2	2	rs13223993	21.4	AA	19	39	-343, 421	-4	-31, 24
					AG	131	42	-131, 216	4	-8, 16
					GG	244	4522	4419, 4624	171	164, 179
3	2	3	rs2270628	21.0	TT	15	-316	-743, 112	-15	-45, 16
					CT	137	-15	-186, 156	3	-9, 15
					CC	247	4544	4441, 4646	172	165, 179
3	2	4	rs12671457	16.5	AC, CC	118	-119	-296, 59	-5	-17, 8
					AA	267	4558	4460, 4657	174	167, 181

NOTE: MAF, minor allele frequency.

* Blocks consist of SNPs with MAF ≥ 5% in strong linkage disequilibrium (LD) (95% of pairwise SNP comparisons with one-sided 95% confidence intervals for the D prime statistic within 0.7-0.98). Group refers to SNP combination for diplotype estimation. Position refers to SNP order in each group for diplotype estimation; the position for each group begins with "1". Values for diplotype group and position (--) were not listed for SNPs excluded from diplotype estimation based on either: 1) Tagger algorithm ($r^2 \geq 0.8$); 2) MAF < 5% (excluded from blocks (--)). rs2453840 (block 2) was combined with block 1 SNPs for diplotype estimation. rs6670 and rs13223993 were combined with block 3 SNPs for diplotype estimation.

† Based on total N = 402.

‡ Total N = 401, excluded 1 woman missing plasma IGFBP-3 and IGF-I levels.

§ β's (95% CIs) in last row of each SNP represent the mean plasma IGFBP-3 or IGF-I levels for homozygous wildtype genotypes (reference) estimated from linear regression intercepts. β's for heterozygous and homozygous variant genotypes represent differences in the mean plasma IGFBP-3 or IGF-I levels for the reference genotypes estimated from linear regression.

‖ SNPs selected *a priori*. rs2854744 is also known as -202 A/C; rs2854746 and rs9282734, nonsynonymous SNPs.

Table 4.3. Unadjusted linear regression of *IGFBP-3* single nucleotide polymorphisms (SNPs) on plasma IGFBP-3 and IGF-I levels among African Americans

Block*	Diplotype Group*	Diplotype Position*	SNP Reference ID	MAF (%)†	Genotype	N‡	Plasma IGFBP-3 (ng/mL)		Plasma IGF-I (ng/mL)	
							β§	95% CI§	β§	95% CI§
Outside	1	1	rs903889	9.4	GT, GG	97	-131	-318, 55	5	-12, 21
					TT	471	4108	4031, 4185	164	157, 171
1	--	--	rs924140	38.9	CC	95	-560	-761, -360	-17	-35, 1
					CT	251	-276	-427, -125	-8	-21, 5
					TT	222	4304	4194, 4414	171	161, 181
1	1	2	rs2854744‖	42.2	CC	107	-534	-728, -339	-13	-31, 4
					AC	258	-258	-412, -105	-8	-21, 6
					AA	195	4305	4190, 4421	171	161, 181
1	1	3	rs2854746‖	32.1	CC	52	897	656, 1138	8	-14, 30
					CG	259	458	318, 598	7	-6, 20
					GG	256	3798	3699, 3897	161	152, 170
1	1	4	rs2471551	20.3	CC	27	-465	-796, -133	-24	-53, 5
					CG	173	-258	-412, -105	-8	-21, 5
					GG	364	4193	4106, 4280	169	161, 176
--	1	5	rs9282734‖	3.3	AC, CC	36	24	-263, 311	-13	-38, 12
					AA	532	4079	4007, 4151	166	159, 172
--	--	--	rs2453837	0	GG	568	4083	4012, 4154	165	159, 171
1	1	6	rs6953668	5.0	AG, AA	55	-124	-362, 115	-16	-37, 4
					GG	513	4094	4020, 4169	166	160, 173
Outside	1	7	rs3110697	36.1	AA	78	-551	-764, -338	-19	-37, 0
					AG	251	-365	-513, -217	-13	-26, 0
					GG	238	4325	4219, 4431	173	164, 182
2	2	1	rs2453840	9.2	AC, AA	98	209	24, 394	6	-10, 22
					CC	470	4045	3969, 4122	164	157, 170

2	2	2	rs2453839	40.9	CC	93	-167	-380, 45	2	-16, 20
					CT	279	-28	-186, 129	-3	-17, 11
					TT	194	4130	4009, 4251	166	156, 176
3	3	1	rs6670	12.6	AT, AA	133	-44	-210, 121	-9	-23, 6
					TT	434	4100	4020, 4180	167	160, 174
3	3	2	rs13223993	43.6	AA	116	-17	-216, 182	-7	-24, 10
					AG	264	-147	-308, 14	-6	-20, 8
					GG	184	4163	4039, 4286	169	158, 180
4	3	3	rs2270628	36.2	TT	83	84	-131, 300	17	-2, 35
					CT	247	-64	-218, 90	-6	-19, 8
					CC	236	4101	3991, 4211	165	155, 174
4	3	4	rs12671457	5.1	AC, CC	55	-218	-456, 21	12	-9, 32
					AA	510	4106	4032, 4181	163	157, 170

NOTE: MAF, minor allele frequency.

* Blocks consist of SNPs with MAF ≥ 5% in strong linkage disequilibrium (LD) (95% of pairwise SNP comparisons with one-sided 95% confidence intervals for the D prime statistic within 0.7-0.98). Group refers to SNP combination for diplotype estimation. Position refers to SNP order in each group for diplotype estimation; the position for each group begins with "1". Values for diplotype group and position (--) were not listed for SNPs excluded from diplotype estimation based on either: 1) Tagger algorithm ($r^2 \geq 0.8$); 2) MAF < 5% (excluded from blocks (--)) or MAF < 3% for SNPs selected *a priori*. rs903889 and rs3110697 were combined with block 1 SNPs for diplotype estimation. SNPs in blocks 3 and 4 were combined together for diplotype estimation.

† Based on total N = 582.

‡ Total N = 574, excluded 8 women missing plasma IGFBP-3 and IGF-I levels.

§ β's (95% CIs) in last row of each SNP represent the mean plasma IGFBP-3 or IGF-I levels for homozygous wildtype genotypes (reference) estimated from linear regression intercepts. β's for heterozygous and homozygous variant genotypes represent differences in the mean plasma IGFBP-3 or IGF-I levels for the reference genotypes estimated from linear regression.

‖ SNPs selected *a priori*. rs2854744 is also known as -202 A/C; rs2854746 and rs9282734, nonsynonymous SNPs.

Table 4.4. Unadjusted linear regression of *IGF-I* single nucleotide polymorphisms (SNPs) on plasma IGFBP-3 and IGF-I levels among Caucasians

Block*	Diplotype Group*	Diplotype Position*	SNP Reference ID	MAF (%)†	Genotype	N‡	Plasma IGFBP-3 (ng/mL)		Plasma IGF-I (ng/mL)	
							β§	95% CI §	β§	95% CI§
Outside	1	1	rs35767	16.0	AG, AA	116	-20	-197, 157	5	-7, 18
					GG	280	4525	4429, 4620	171	164, 177
--	--	--	rs5742612	3.9	AG, GG	30	240	-65, 546	21	-1, 42
					AA	368	4506	4422, 4590	171	165, 177
--	--	--	rs5742614	1.8	CG	14	227	-211, 665	18	-14, 49
					GG	385	4518	4436, 4600	172	166, 178
--	--	--	rs3729846‖	0	CC	399	4528	4447, 4608	172	166, 178
1	2	1	rs12821878	22.7	AA	21	235	-130, 601	0	-26, 27
					AG	139	52	-119, 223	5	-8, 17
					GG	237	4494	4390, 4598	171	163, 178
1	--	--	rs10860869	29.1	TT	31	-188	-499, 123	-5	-28, 17
					AT	170	29	-140, 198	6	-6, 18
					AA	196	4532	4417, 4647	170	162, 179
1	2	2	rs1019731	12.3	AC, AA	93	98	-93, 288	10	-4, 23
					CC	304	4509	4417, 4601	170	164, 177
1	2	3	rs7956547	27.3	CC	29	-167	-486, 152	-5	-28, 18
					CT	160	-3	-173, 166	2	-10, 14
					TT	209	4541	4429, 4652	172	164, 180
--	--	--	rs5742626	0.5	CT	4	-255	-1058, 548	-3	-61, 54
					TT	394	4522	4442, 4603	172	166, 178
--	--	--	rs17880975	0.1	AG	1	-607	-2223, 1008	-78	-193, 37
					GG	395	4531	4450, 4613	173	167, 178
2	3	1	rs2033178	6.2	AG, AA	47	-139	-389, 111	5	-13, 22
					GG	349	4537	4451, 4623	172	166, 178
--	--	--	rs17884646	0	TT	397	4523	4443, 4603	172	166, 178
--	--	--	rs5742657	2.2	AG, GG	17	29	-370, 428	18	-11, 46
					AA	383	4524	4442, 4607	171	166, 177

--	--	--	rs5742663	0.1	GT	1	-519	-2131, 1093	12	-103, 127
					TT	396	4532	4451, 4613	172	167, 178
--	--	--	rs11829586	2.3	AG	18	107	-282, 496	16	-12, 43
					GG	380	4524	4441, 4606	172	166, 178
2	--	--	rs4764884	25.7	TT	25	-251	-587, 85	-16	-40, 8
					CT	153	-128	-296, 40	0	-12, 12
					CC	218	4588	4481, 4696	173	165, 181
--	--	--	rs5742683	0.1	AG	1	-512	-2116, 1092	12	-102, 127
					AA	396	4525	4445, 4606	172	166, 178
--	--	--	rs17884626ll	0	CC	399	4526	4445, 4607	172	167, 178
2	--	--	rs5009837	30.3	TT	36	-223	-513, 67	-6	-27, 15
					CT	167	-52	-221, 117	6	-6, 18
					CC	192	4564	4449, 4680	170	162, 178
2	3	2	rs17727841	19.1	CC	15	-273	-700, 155	-4	-35, 27
					CG	123	-78	-254, 98	-4	-17, 8
					GG	261	4559	4459, 4658	174	167, 181
2	--	--	rs4764883	30.9	CC	37	-237	-525, 50	-8	-29, 12
					CT	167	-60	-230, 110	4	-8, 17
					TT	186	4577	4460, 4694	171	163, 179
2	--	--	rs9308315	28.2	AA	28	-315	-636, 5	-16	-39, 7
					AT	166	-107	-274, 60	3	-9, 15
					TT	201	4590	4478, 4703	172	164, 180
2	3	3	rs978458	28.0	TT	28	-313	-634, 8	-13	-37, 10
					CT	164	-114	-281, 53	4	-8, 16
					CC	202	4590	4478, 4702	172	163, 180
--	--	--	rs5742692	2.1	AG	17	27	-373, 426	18	-11, 46
					AA	382	4526	4444, 4609	171	166, 177
2	--	--	rs11111262	9.9	AG, AA	76	-181	-385, 23	-4	-19, 10
					GG	321	4567	4478, 4657	173	167, 180
2	3	4	rs1520220	19.1	GG	13	-539	-989, -88	-33	-65, 0
					CG	124	-164	-337, 9	7	-6, 19
					CC	258	4590	4491, 4688	171	164, 178

--	--	--	rs3730204	2.3	CT	18	-206	-595, 183	-24	-52, 3
					TT	378	4538	4455, 4621	173	168, 179
Outside	3	5	rs6214	39.8	TT	67	178	-59, 414	17	0, 34
					CT	183	-59	-236, 119	-2	-14, 11
					CC	148	4524	4392, 4656	170	161, 180
3	4	1	rs6219	10.0	CT, TT	78	-200	-402, 3	-6	-20, 9
					CC	321	4565	4476, 4655	174	167, 180
3	4	2	rs2946834	33.0	AA	39	-287	-571, -3	-7	-28, 13
					AG	181	-78	-248, 92	1	-11, 13
					GG	175	4585	4463, 4706	172	164, 181

NOTE: MAF, minor allele frequency.

* Blocks consist of SNPs with MAF ≥ 5% in strong linkage disequilibrium (LD) (95% of pairwise SNP comparisons with one-sided 95% confidence intervals for the D prime statistic within 0.7-0.98). Group refers to SNP combination for diplotype estimation. Position refers to SNP order in each group for diplotype estimation; the position for each group begins with "1". Values for diplotype group and position (--) were not listed for SNPs excluded from diplotype estimation based on either: 1) Tagger algorithm ($r^2 \geq 0.8$); 2) MAF < 5% (excluded from blocks (--)). rs6214 was combined with block 2 SNPs for diplotype estimation.

† Based on total N = 402.

‡ N = 401, excluded 1 woman missing plasma IGFBP-3 and IGF-I levels.

§ β's (95% CIs) in last row of each SNP represent the mean plasma IGFBP-3 or IGF-I levels for homozygous wildtype genotypes (reference) estimated from linear regression intercepts. β's for heterozygous and homozygous variant genotypes represent differences in the mean plasma IGFBP-3 or IGF-I levels for the reference genotypes estimated from linear regression.

‖ SNPs selected *a priori*. rs3729846, synonymous SNP; rs17884626, nonsynonymous SNP.

Table 4.5. Unadjusted linear regression of *IGF-I* single nucleotide polymorphisms (SNPs) on plasma IGFBP-3 and IGF-I levels among African Americans

Block*	Diplotype Group*	Diplotype Position*	SNP Reference ID	MAF (%)†	Genotype	N‡	Plasma IGFBP-3 (ng/mL)		Plasma IGF-I (ng/mL)	
							β§	95% CI§	β§	95% CI§
1	1	1	rs35767	40.8	AA	99	40	-166, 245	-1	-19, 17
					AG	264	-68	-224, 89	-2	-16, 11
					GG	205	4109	3992, 4227	166	156, 176
--	--	--	rs5742612	3.3	AG, GG	35	154	-139, 448	5	-20, 31
					AA	536	4079	4006, 4152	164	158, 171
1	1	2	rs5742614	11.5	CG, CC	124	-48	-219, 123	8	-6, 23
					GG	444	4097	4017, 4177	163	156, 170
--	--	--	rs3729846‖	0.8	CT	9	291	-266, 848	33	-16, 82
					CC	557	4070	3999, 4140	164	158, 170
Outside	1	3	rs12821878	5.8	AG, AA	64	-180	-404, 43	-15	-34, 5
					GG	505	4106	4031, 4181	167	160, 173
Outside	2	1	rs10860869	32.3	TT	59	265	23, 507	11	-10, 32
					AT	249	54	-95, 202	12	-1, 25
					AA	262	4034	3930, 4137	158	149, 167
--	--	--	rs1019731	3.4	AC	39	-140	-420, 140	-10	-34, 14
					CC	528	4096	4023, 4170	165	159, 172
2	2	2	rs7956547	26.2	CC	43	152	-118, 422	8	-16, 32
					CT	210	73	-75, 221	10	-3, 23
					TT	314	4042	3949, 4136	160	152, 168
2	2	3	rs5742626	7.2	CT, CC	78	27	-178, 233	-8	-26, 9
					TT	491	4084	4007, 4160	166	159, 173
--	--	--	rs17880975	2.3	AG	26	319	-19, 657	8	-21, 38
					GG	540	4075	4002, 4147	165	158, 171
--	--	--	rs2033178	4.7	AG, AA	50	-135	-385, 115	-24	-46, -3
					GG	513	4099	4025, 4174	167	161, 174
--	--	--	rs17884646	0	TT	570	4084	4014, 4155	165	159, 171

2	--	--	rs5742657	12.1	AG, GG	128	75	-94, 243	3	-11, 18
					AA	440	4063	3984, 4143	164	157, 170
2	--	--	rs5742663	11.6	GT, GG	123	102	-69, 274	8	-6, 23
					TT	444	4063	3983, 4143	163	156, 170
2	--	--	rs11829586	11.5	AG, AA	121	116	-56, 288	7	-8, 22
					GG	445	4054	3974, 4133	163	156, 170
2	2	4	rs4764884	23.7	TT	34	102	-202, 405	10	-16, 36
					CT	200	11	-140, 162	5	-8, 18
					CC	333	4073	3981, 4165	162	154, 170
2	--	--	rs5742683	10.6	AG, GG	113	112	-65, 289	5	-11, 20
					AA	454	4061	3982, 4140	164	157, 171
--	--	--	rs17884626‖	1.0	CT	11	104	-408, 617	13	-31, 57
					CC	559	4085	4014, 4156	164	158, 171
Outside	2	5	rs5009837	41.3	TT	105	135	-67, 337	8	-9, 26
					CT	260	96	-61, 254	10	-3, 24
					CC	203	4014	3896, 4132	158	148, 168
--	--	--	rs17727841	4.7	CG, CC	49	74	-178, 326	23	1, 44
					GG	520	4079	4005, 4153	163	156, 169
3	--	--	rs4764883	45.8	TT	132	-128	-321, 66	-12	-29, 5
					CT	248	-105	-270, 60	-11	-25, 3
					CC	180	4169	4043, 4295	173	162, 183
3	3	1	rs9308315	49.8	TT	150	-85	-277, 107	-12	-28, 5
					AT	262	-94	-264, 76	-11	-26, 4
					AA	152	4145	4009, 4280	173	161, 185
3	--	--	rs978458	37.5	TT	83	113	-104, 329	9	-10, 28
					CT	259	72	-82, 226	7	-6, 20
					CC	224	4039	3926, 4151	160	151, 170
3	3	2	rs5742692	11.7	GG	11	69	-443, 581	14	-31, 58
					AG	111	120	-58, 299	4	-12, 19
					AA	444	4054	3974, 4134	164	157, 170
--	--	--	rs11111262	2.7	AG, AA	27	81	-252, 413	40	11, 68
					GG	536	4083	4010, 4155	163	157, 169

3	3	3	rs1520220	34.7	GG	73	112	-113, 336	10	-9, 29
					CG	250	111	-40, 263	6	-7, 19
					CC	245	4024	3917, 4132	161	151, 170
--	--	--	rs3730204	0.3	CT	3	-736	-1711, 239	6	-79, 90
					TT	562	4097	4026, 4168	165	159, 171
Outside	3	4	rs6214	45.2	CC	124	-63	-259, 133	-3	-20, 14
					CT	263	61	-102, 223	-4	-18, 10
					TT	178	4068	3942, 4193	167	156, 178
4	4	1	rs6219	9.1	CT, TT	94	60	-130, 251	16	-1, 32
					CC	475	4074	3996, 4151	162	155, 169
4	4	2	rs2946834	49.3	AA	142	149	-49, 346	16	0, 33
					AG	279	52	-119, 223	2	-12, 17
					GG	147	4026	3888, 4165	159	147, 171

NOTE: MAF, minor allele frequency.

* Blocks consist of SNPs with MAF ≥ 5% in strong linkage disequilibrium (LD) (95% of pairwise SNP comparisons with one-sided 95% confidence intervals for the D prime statistic within 0.7-0.98). Group refers to SNP combination for diplotype estimation. Position refers to SNP order in each group for diplotype estimation; the position for each group begins with "1". Values for diplotype group and position (--) were not listed for SNPs excluded from diplotype estimation based on either: 1) Tagger algorithm ($r^2 \geq 0.8$); 2) MAF < 5% (excluded from blocks (--)). rs12821878 was combined with block 1 SNPs for diplotype estimation. rs10860869 and rs5009837 were combined with block 2 SNPs for diplotype estimation. rs9308315 was substituted for rs4764883 (Hardy-Weinberg p value <0.01). rs6214 was combined with block 3 SNPs for diplotype estimation.

† Based on total N = 582.

‡ Total N = 574, excluded 8 women missing plasma IGFBP-3 and IGF-I levels.

§ β's (95% CIs) in last row of each SNP represent the mean plasma IGFBP-3 or IGF-I levels for homozygous wildtype genotypes (reference) estimated from linear regression intercepts. β's for heterozygous and homozygous variant genotypes represent differences in the mean plasma IGFBP-3 or IGF-I levels for the reference genotypes estimated from linear regression.

‖ SNPs selected *a priori*. rs3729846, synonymous SNP; rs17884626, nonsynonymous SNP.

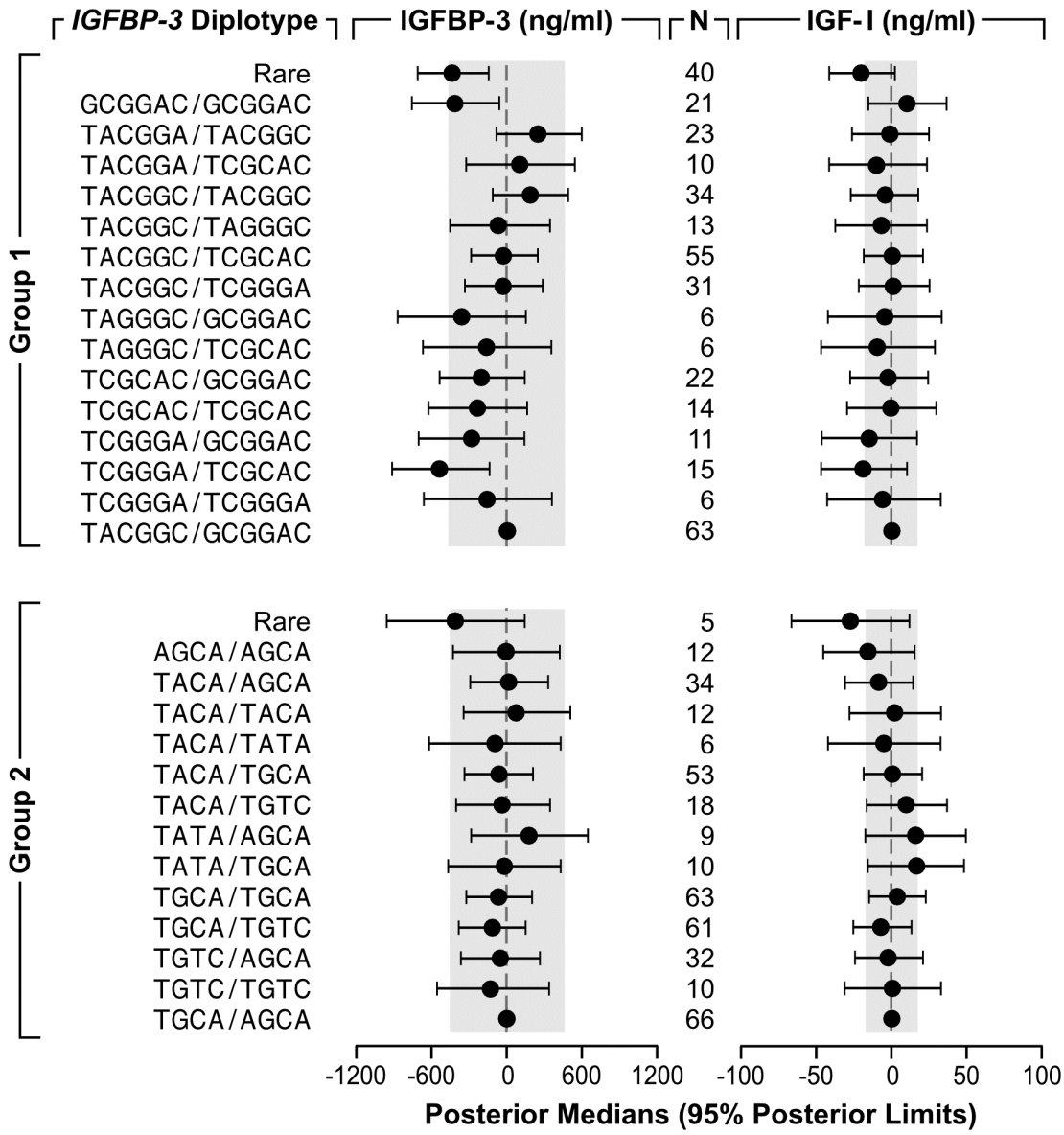

Figure 4.1. Unadjusted linear regression of *IGFBP-3* diplotypes on plasma IGFBP-3 and IGF-I levels among Caucasians. Diplotype group-specific models with posterior medians and 95% posterior limits based on prior mean of 0 and prior variance corresponding to one standard deviation of the mean plasma IGFBP-3 and IGF-I levels. Groups were created by combining individual SNPs that were not included in a block with an adjacent block and combining adjacent blocks with each other, if doing so resulted in diplotypes estimated with at least 90% posterior probability for at least 90% of Caucasians. After excluding 1 woman missing IGFBP-3 and IGF-I, Group 1 (N=370) represents SNPs from blocks 1 and 2; Group 2 (N=391) represents 2 SNPs outside block 3 combined with block 3 SNPs. Last diplotype listed in each group is most common (referent) diplotype. Shaded area indicates ±10% of the mean IGFBP-3 (Group 1: 4661 ng/ml; Group 2: 4587 ng/ml) or IGF-I (Group 1: 177 ng/ml; Group 2: 174 ng/ml) for each group-specific wildtype diplotype among Caucasians.

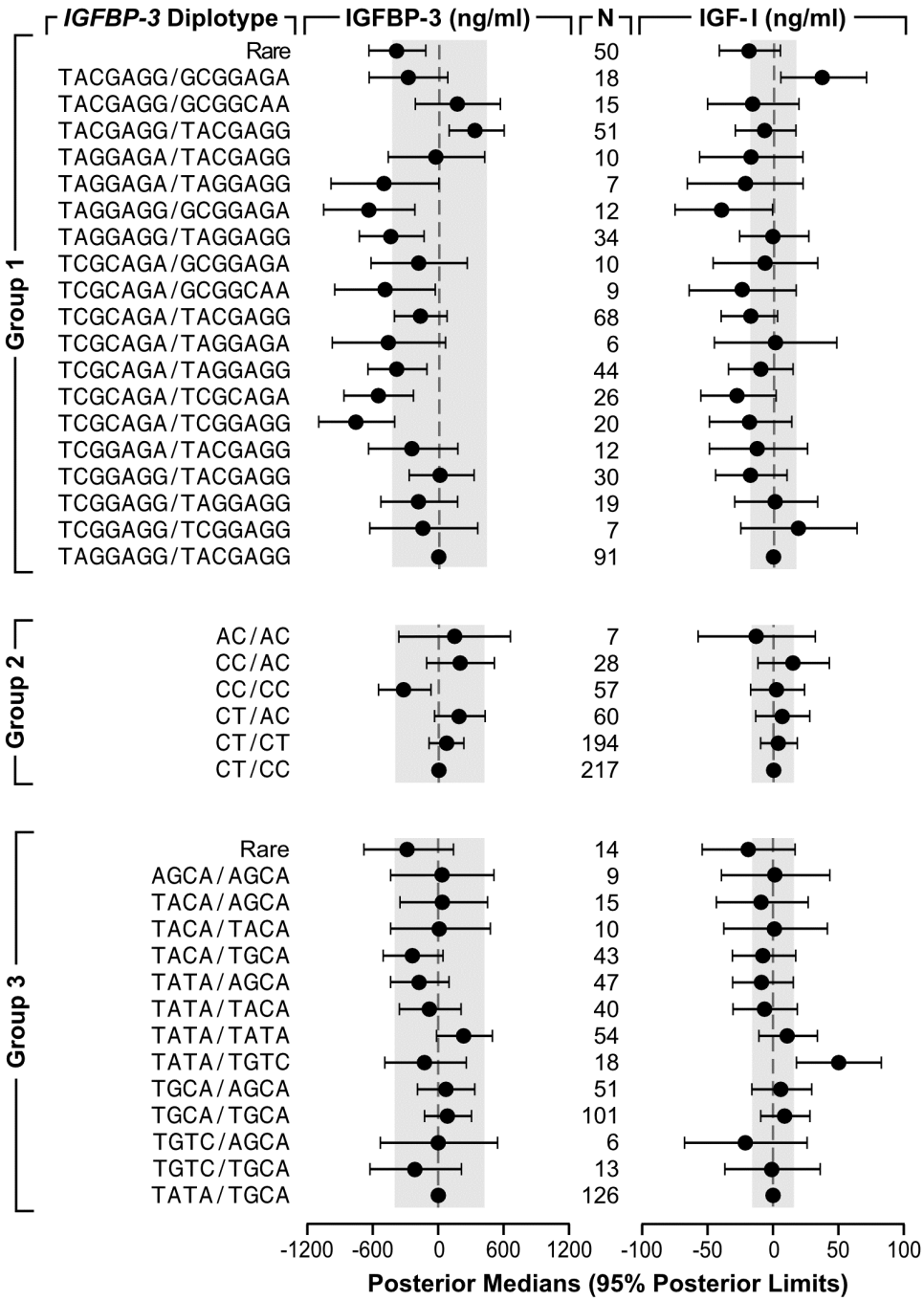

Figure 4.2. Unadjusted linear regression of *IGFBP-3* diplotypes on plasma IGFBP-3 and IGF-I levels among African Americans. Diplotype group-specific models with posterior medians and 95% posterior limits based on prior mean of 0 and prior variance corresponding to one standard deviation of the mean plasma IGFBP-3 and IGF-I levels. Groups were created by combining individual SNPs that were not included in a block with an adjacent block and combining adjacent blocks with each other, if doing so resulted in diplotypes estimated with at least 90% posterior probability for at least 90% of African Americans. After excluding 8 women missing IGFBP-3 and IGF-I, Group 1 (N=539) represents 3 SNPs outside block 1 combined with block 1 SNPs; Group 2 (N=563) represents block 2 SNPs; Group 3 (N=547) represents SNPs from blocks 3 and 4. Last diplotype listed in each group is most common (referent) diplotype. Shaded area indicates ±10% of the mean IGFBP-3 (Group 1: 4327 ng/ml; Group 2: 4060 ng/ml; Group 3: 4106 ng/ml) or IGF-I (Group 1: 175 ng/ml; Group 2: 162 ng/ml; Group 3: 162 ng/ml) for each group-specific wildtype diplotype among African Americans.

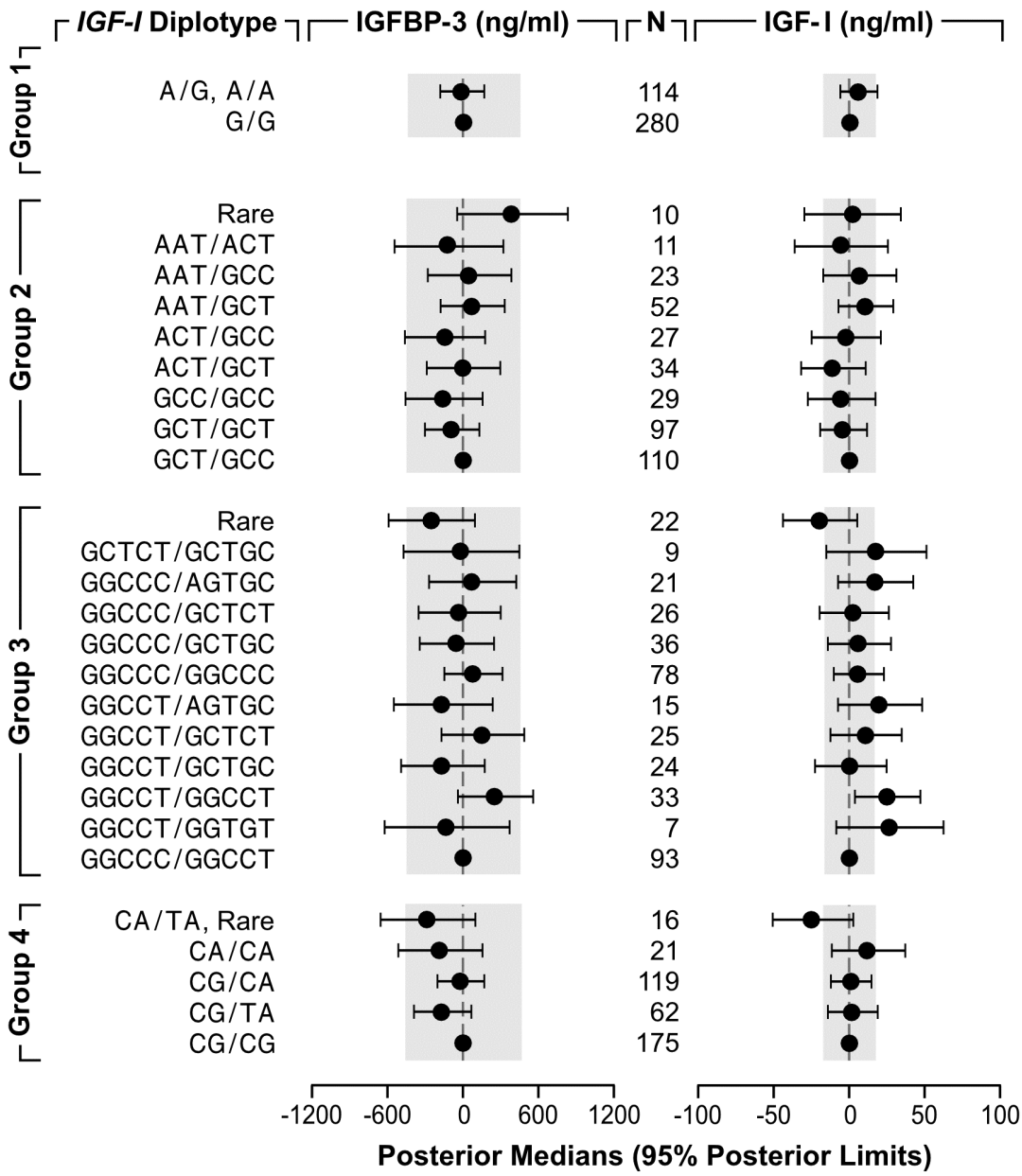

Figure 4.3. Unadjusted linear regression of *IGF-I* diplotypes on plasma IGFBP-3 and IGF-I levels among Caucasians. Diplotype group-specific models with posterior medians and 95% posterior limits based on prior mean of 0 and prior variance corresponding to one standard deviation of the mean plasma IGFBP-3 and IGF-I levels. Groups were created by combining individual SNPs that were not included in a block with an adjacent block and combining adjacent blocks with each other, if doing so resulted in diplotypes estimated with at least 90% posterior probability for at least 90% of Caucasians. After excluding 1 woman missing IGFBP-3 and IGF-I, Group 1 (N=394) represents 1 SNP (rs35767) outside block 1 since it could not be combined with any adjacent SNPS; Group 2 (N=393) represents block 1 SNPs; Group 3 (N=389) represents block 2 SNPs combined with 1 SNP outside block 2; Group 4 (N=393) represents block 3 SNPs. Last diplotype listed in each group is most common (referent) diplotype. Shaded area indicates ±10% of the mean IGFBP-3 (Group 1: 4525 ng/ml; Group 2: 4559 ng/ml; Group 3: 4518 ng/ml; Group 4: 4585 ng/ml) or IGF-I (Group 1: 171 ng/ml; Group 2: 173 ng/ml; Group 3: 165 ng/ml; Group 4: 172 ng/ml) for each group-specific wildtype diplotype among Caucasians.

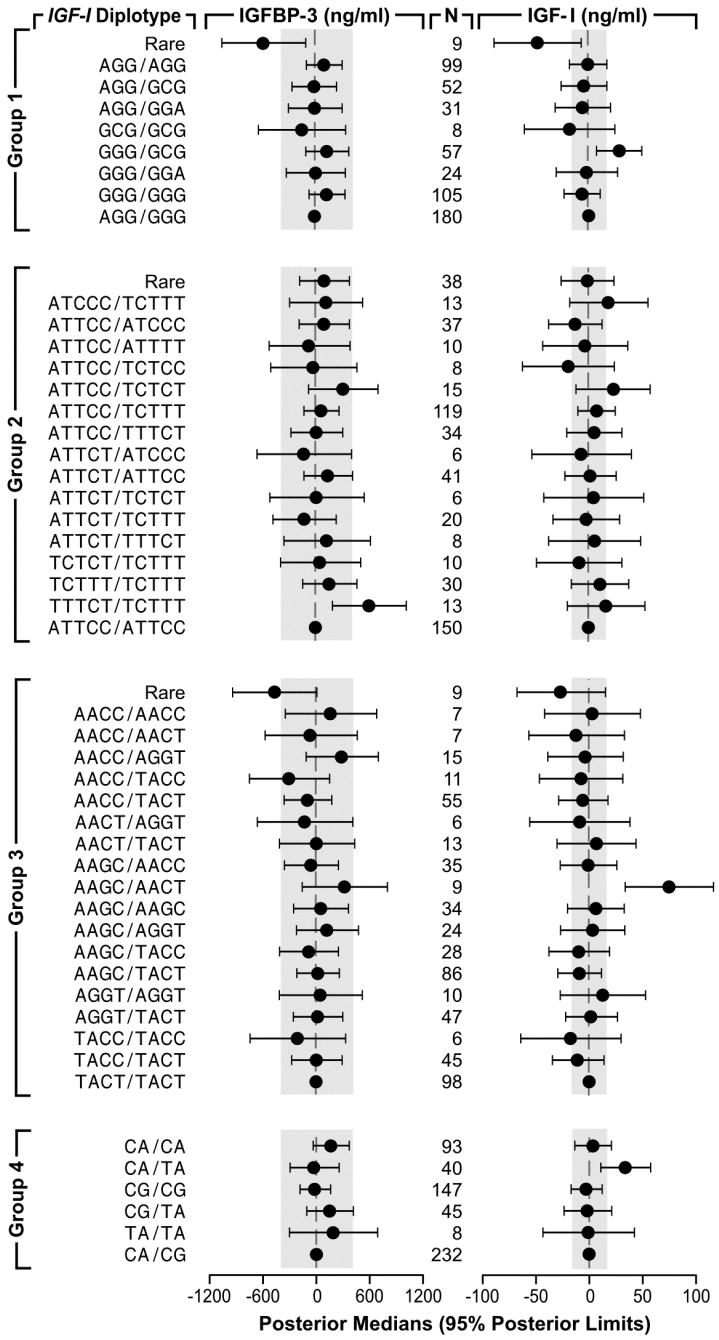

Figure 4.4. Unadjusted linear regression of *IGF-I* diplotypes on plasma IGFBP-3 and IGF-I levels among African Americans. Diplotype group-specific models with posterior medians and 95% posterior limits based on prior mean of 0 and prior variance corresponding to one standard deviation of the mean plasma IGFBP-3 and IGF-I levels. Groups were created by combining individual SNPs that were not included in a block with an adjacent block and combining adjacent blocks with each other, if doing so resulted in diplotypes estimated with at least 90% posterior probability for at least 90% of African Americans. After excluding 8 women missing IGFBP-3 and IGF-I, Group 1 (N=565) represents block 1 SNPs combined with 1 SNP outside block 1; Group 2 (N=558) represents 2 SNPs outside block 2 combined with block 2 SNPs; Group 3 (N=545) represents block 3 SNPs combined with 1 SNP outside block 3; Group 4 (N=565) represents block 4 SNPs. Last diplotype listed in each group is most common (referent) diplotype. Shaded area indicates ±10% of the mean IGFBP-3 (Group 1: 4043 ng/ml; Group 2: 3999 ng/ml; Group 3: 4083 ng/ml; Group 4: 4048 ng/ml) or IGF-I (Group 1: 166 ng/ml; Group 2: 162 ng/ml; Group 3: 166 ng/ml; Group 4: 162 ng/ml) for each group-specific wildtype diplotype among African Americans.

CHAPTER V

PAPER 2: ASSOCIATION OF *IGF-I* AND *IGFBP-3* POLYMORPHISMS WITH UTERINE LEIOMYOMATA AMONG AFRICAN AMERICAN AND CAUCASIAN WOMEN

Abstract

Uterine leiomyomata (fibroids) are responsible for substantial morbidity, especially among African Americans. Insulin-like growth factor-I (IGF-I), a potent mitogen, is hypothesized to be involved in fibroid pathogenesis based on gene expression studies, and IGFBP-3 may be involved in IGF-I mediated mechanisms based on influencing bioavailability of IGF-I. We evaluated 30 *IGF-I* and 15 *IGFBP-3* SNPs and estimated diplotypes in relation to fibroid prevalence (any and 2+ cm) among 984 premenopausal African American and Caucasian women. We noted associations between *IGF-I* and *IGFBP-3* variants and fibroids, but findings with diplotypes generally did not support SNP associations. Relatively precise *IGF-I* and *IGFBP-3* SNP variant associations with magnitudes of at least 10% were predominantly reported among African Americans, including functionally relevant *IGFBP-3* SNPs of rs9282734 (nonsynonymous) (overall: PD = -0.130, 95% CI: -0.294, 0.034) and rs2475551 (splice site) (overall: PD = 0.208, 95% CI: 0.095, 0.320) and *IGF-I* SNP rs35767 (promoter) (overall: PD = 0.208, 95% CI: 0.095, 0.320). Associations with larger fibroids were generally consistent or slightly weaker than associations with any fibroids, and did not provide stronger evidence for

IGF-I involvement in promoting growth rather than tumorigenesis. This is the first epidemiologic study to evaluate *IGF-I* and *IGFBP-3* variants in relation to fibroids. In conclusion, this study provides support for hypothesis of IGF-I involvement with fibroids. Future research should validate findings in other populations represented by African Americans and examine other genes involved in the IGF-I pathway.

Introduction

Uterine leiomyomata (fibroids) are hormonally-dependent benign tumors of smooth muscle origin that are the principal indication for hysterectomies in the United States (1, 2). Age and ethnicity have been the most consistently reported risk factors for uterine fibroids, with advancing age until menopause and African American versus Caucasian race associated with an increased incidence of fibroids and related morbidity (3-7). The National Institute of Environmental Health Sciences (NIEHS) Uterine Fibroid Study, which used ultrasound screening to detect fibroids in cohort participants, estimated that the risk of fibroids by age 50 is over 80% among African Americans and over 60% among Caucasians (3).

Estrogens and progesterones are hypothesized to promote fibroid development and growth by accelerating the activity of cytokines and growth factors (8). In particular, insulin-like growth factor-I (IGF-I), a polypeptide with structural similarities to insulin, has been implicated in fibroid etiology based on *in vitro* promotion of fibroid cell growth (9) and evidence of increased IGF-I mRNA or peptides (10-16) in fibroids relative to normal myometrium. A recent study reported elevated IGF-I mRNA and protein after estrogen treatment in fibroids relative to normal myometrium, in addition to relative

changes in levels of two IGF-I pathway factors (17). IGF Binding Protein-3 (IGFBP-3) binds approximately 90% of circulating IGF-I (18), and may contribute to IGF-I mediated pathogenic mechanisms. Although biologic evidence supporting IGFBP-3 involvement with uterine fibroids is limited, one gene expression study found elevated IGFBP-3 mRNA in fibroids compared to normal myometrium (16).

We examined whether IGF-I and IGFBP-3 were associated with uterine fibroid development and growth by evaluating relations between *IGF-I* and *IGFBP-3* polymorphisms and the prevalence of uterine fibroids. Specifically, we evaluated 45 single nucleotide polymorphisms (SNPs) in *IGF-I* and *IGFBP-3* and estimated diplotypes (paired haplotypes) in association with the prevalence of 1) all fibroids versus no fibroids and 2) larger (≥2cm) fibroids versus small fibroids or no fibroids among premenopausal African American and Caucasian participants in the National Institute for Environmental Health Sciences (NIEHS) Uterine Fibroid Study.

Methods

Study population

The study population consisted of 984 premenopausal women (582 African Americans and 402 Caucasians) with available DNA samples for genotyping from the NIEHS Uterine Fibroid Study (UFS). The parent study was designed to estimate the prevalence of uterine leiomyomata (fibroids) among African American and Caucasian women, and to evaluate potential etiologic factors for fibroids. The consent form specified use of biological samples for genetic polymorphism analyses. Details of the parent study, which was approved by the NIEHS and George Washington University

Human Subject's Review Boards, were previously described (3, 19, 20). Briefly, participants in the parent study were recruited from a random sample of 35 to 49-year old George Washington University health plan members. Approximately 20% of eligible women refused participation, resulting in a total of 1,430 participants enrolled between 1996 through 1999 (19).

We restricted the current study population to women who self-reported their race as African American or Caucasian (n = 1,323) to facilitate race-specific analyses, and excluded women who were postmenopausal at enrollment since they were not asked to participate in ultrasound screening and blood collection (n = 178). Race and menopausal status criteria were met by 1,145 women, and DNA was extracted for 984 of the 1,003 women with collected blood samples.

Uterine fibroid assessment

Many women have asymptomatic fibroids; therefore, the parent study conducted ultrasound screening of premenopausal participants to provide a better estimate of fibroid prevalence than previous studies based on clinical diagnosis or treatment by hysterectomy only. Both transabdominal and transvaginal ultrasound procedures were performed by trained sonographers certified by the American Registry of Diagnostic Medical Sonographers under the supervision of a radiologist who verified their assessments. Information collected by sonographers included the size and location of the two largest fibroids and the number of tumors within the uterus (3). Study sonograms were not performed for women who had a recent ultrasound at the clinic that conducted UFS study sonograms; instead, medical records were abstracted to obtain fibroid information for these women (N = 204, approximately 20% of the current study sample).

In addition, 19 women in our study sample were classified with regard to the size and presence of fibroids based on self-reported previous diagnoses only. Fibroid status could not be assessed in seven African Americans and eight Caucasians in our study sample; however, these women were genotyped based on analyses with other outcomes including circulating IGF-I and IGFBP-3 levels (not shown in current paper).

Selection of genetic polymorphisms

Race-specific haplotype-tagging SNPs in *IGF-I* and *IGFBP-3* were selected using Genome Variation Server (GVS) software (sponsored by the Seattle SNPs Program for Genomic Applications (PGA)) (21). We used the Seattle SNPs database as the reference population for *IGF-I*, and the HapMap database for *IGFBP-3*, which has not been evaluated by Seattle SNPs (21). In addition to SNPs within each gene, we included SNPs within 5 kilobases (kb) of the 5' and 3' ends of *IGF-I* and *IGFBP-3*, and used a value of 0.8 as the pairwise correlation coefficient (r^2) for identifying haplotype-tagging SNPs. We selected only haplotype-tagging SNPs with a minor allele frequency (MAF) greater than 5% among women in at least one racial group. There were 29 haplotype-tagging SNPs for *IGF-I* and 12 haplotype-tagging SNPs for *IGFBP-3* overall. In addition, we selected four SNPs *a priori* based on functional significance, including one nonsynonymous *IGF-I* SNP (rs17884626), one synonymous *IGF-I* SNP (rs3729846), and two nonsynonymous *IGFBP-3* SNPs (rs2854746, rs9282734) and an *IGFBP-3* promoter SNP (-202 A/C, rs2854744) *a priori* based on previous studies of associations with circulating IGF-I and IGFBP-3 levels and health outcomes (22-29).

Sample collection and genotyping

Blood specimens were collected through venipuncture after an 8-hour fast. Genomic DNA was initially extracted from whole blood using a phenol:chloroform procedure, but was later replaced by a safer and more efficient modified salt precipitation procedure (GenQuik Protocol). Genotyping was performed using the TaqMan genotyping approach (30-32) at an outside laboratory (Mammalian Genotyping Core, Lineberger Comprehensive Cancer Center; Chapel Hill, NC). Allele-specific oligonucleotide probes for 39 selected SNPs were purchased from Applied Biosystems (ABI; Foster City, CA) "TaqMan® Validated and Coding SNP or Pre-Designed SNP Genotyping Assays". In addition, ABI attempted to develop custom assays for the 6 remaining SNPs through their "Custom TaqMan® SNP Genotyping Assays" service. Two of the initially selected *IGF-I* haplotype tagging SNPs were removed from analyses, including one for which a custom assay could not be developed and one with a pre-designed assay that did not meet ABI technical specifications. We also genotyped an alternate *IGF-I* haplotype-tagging SNP after one seemed inconsistent with Hardy-Weinberg equilibrium among African Americans. Thus, 30 *IGF-I* and 15 *IGFBP-3* SNPs were genotyped, including 40 haplotype-tagging SNPs and 5 SNPs selected *a priori*.

PCR amplification was performed on an ABI GeneAmp® PCR System 9700 thermal cycler with dual 384-well-blocks, and endpoint plates were read using the ABI 7900HT system. VIC and 6-FAM reporter dyes were used as the fluorescent signals to distinguish wild type and variant alleles. Alleles were called automatically through Sequence Detection System (SDS) 2.3 software, and confirmed through a review of all output by experienced operators. The DNA concentration of each sample was validated using a

NanoDrop® ND-1000 Spectrophotometer prior to dilution to 5 ng/ul (using DNA grade sterile water). Samples were placed into eleven 96-well microtiter plates and aliquoted into three 384-well PCR plates for analysis. Each of the eleven microtiter plates contained four randomly assigned controls, including two blank samples and two with a known DNA standard (Control DNA CEPH Individual 1347-02, ABI). Quality control measures also included blinded genotyping of 28 duplicate samples representing 22 women, which produced concordant results for all samples. The overall call rate was 98.8%, and only 5 women had less than 50% of complete allele calls for the 45 SNPs assayed. We confirmed that SNP genotype frequencies were consistent with Hardy-Weinberg equilibrium (HWE) within each racial group using the exact test statistic with one degree of freedom ($\alpha = 0.01$) (33).

Diplotype estimation

In addition to estimating associations with individual SNPs, we estimated associations with race-specific diplotypes (paired haplotypes) imputed for groups of related SNPs within each gene, as described below. We excluded SNPs from race-specific diplotype analyses if their MAF was below 5% (for haplotype-tagging SNPs) or 3% (for *a priori* SNPs) within the racial category being evaluated. In addition, women missing genotype data for more than 50% of the SNPs relevant for diplotype estimation within a gene were excluded from diplotype analyses for that gene (one Caucasian and three African Americans for *IGFBP-3*, three Caucasians and two African Americans for *IGF-I*). We then evaluated race-specific linkage disequilibrium (LD) patterns using Haploview software (34) to identify SNPs in each gene that could be grouped together to estimate diplotypes. First, we identified blocks consisting of individual SNPs (with MAF

at least 5%) in LD based on 95% of pairwise SNP comparisons with one-sided 95% confidence intervals for the D prime statistic within 0.7 to 0.98 (35). Next, we used the Tagger approach (36) in Haploview to identify pairs of redundant SNPs in strong LD (pairwise r^2 values ≥ 0.8), and excluded one member of each redundant pair from diplotype estimation unless both SNPs had been selected *a priori*. Race-specific pairwise r^2 values are available for *IGFBP-3* and *IGF-I* SNPs in Tables A.1-A.4.

Race-specific diplotypes for defined groups of SNPs were then estimated using PHASE version 2.1 (37, 38), which uses a Markov chain-Monte Carlo (MCMC) algorithm to allocate the most likely diplotypes for each person, with the prior assumption that frequently observed haplotypes with less ambiguity are more probable. PHASE also provides a posterior probability estimate that expresses the uncertainty associated with each diplotype assignment. To simplify diplotype analyses, we created race-specific groups for diplotype estimation by combining individual SNPs that were not included in a block with adjacent blocks, and combined adjacent blocks with each other, if doing so resulted in diplotypes that were estimated with at least 90% certainty (posterior probability) for at least 90% of individuals within the racial category. Otherwise, diplotype groups represented SNPs within a single block, or individual SNPs that were not included in a block. We assigned women to their most probable diplotype for each group; however, if their most probable diplotype had a posterior probablility below 90%, we classified women as missing for that diplotype group.

Statistical analysis

All statistical analyses were stratified by race and conducted using SAS V9.1 (SAS Institute Inc., Cary, NC). We estimated race-specific prevalence differences (PDs) for

associations between *IGF-I* and *IGFBP-3* gene variants and prevalent fibroids, including associations with individual SNPs and associations with race-specific diplotypes. We performed separate analyses to 1) estimate associations between gene variants and any uterine fibroids versus no uterine fibroids, and 2) estimate associations with fibroids at least 2 cm in diameter versus no fibroids and fibroids < 2cm. The latter analysis focuses on women with moderate to larger fibroids to explore the hypothesis that IGF-I stimulates fibroid growth rather than incidence. We reported unadjusted associations since there are no known factors other than race that would predict both fibroids and gene variants. We estimated 95% confidence intervals for crude prevalence differences (PD) in SAS V9.1, which calculates standard errors from the square root of the sum of the fibroid prevalence variances based on formula for independent proportions. For diplotype analyses, we used Poisson regression with the robust variance option in PROC GENMOD to estimate prevalence differences with 95% confidence intervals (39).

For the analysis of individual SNPs, we first considered a codominant (general) model of inheritance in which heterozygous and homozygous variant genotypes were separately evaluated relative to the homozygous wildtype genotype. Codominant models make the fewest assumptions with regard to the inheritance pattern but could be used only for common variants. Therefore, for less common variants, we used a dominant model in which we combined homozygous variant and heterozygous genotypes and compared to those with the homozygous wildtype genotype. Associations with individual SNPs were not estimated if there were 5 or fewer women in a racial group with heterozygous and homozygous variant genoypes combined.

We estimated associations between *IGF-I* or *IGFBP-3* diplotypes and fibroids using separate race-specific models for each diplotype group, with the most common diplotype as the reference category. Diplotypes assigned to 5 or fewer women were combined into a "rare diplotype" category. We used a Bayesian method of information-weighted averaging (40) to improve the validity and precision of estimates. This method shrinks imprecise estimates based on smaller numbers of observations further toward the prior mean than more precise regression estimates. Since we did not have information to group diplotypes according to the direction or strength of their potential associations with fibroids, we assumed a prior mean of 0, and specified a prior variance, $(0.5/3.92)^2 \approx 0.016$, that assumed with 95% probability that regression estimates would vary within a range of ± 0.25. For each diplotype, we applied the shrinkage estimator and report posterior medians (50^{th} percentile of the posterior probability distribution) and 95% posterior limits (2.5^{th} and 97.5^{th} percentiles of the posterior probability distribution). Regression estimates and 95% confidence intervals estimated directly by Poisson regression with robust variances are available in Tables A.15-A.22.

Results

Participant characteristics

The study population included 582 African American women and 402 Caucasian women that were similar with regard to age (35 to 49 years) based on the study design. The overall prevalence of fibroids was 72% among African Americans and 50% among Caucasians, with 56% of African Americans having at least one fibroid that was 2 cm diameter or larger, compared with 33% of Caucasians (Table 5.1).

IGFBP-3 SNPs

We emphasize SNP associations where fibroid prevalence is increased or decreased by at least 10% with variant genotypes relative to homozygous wildtype genotypes, and we consider less support for imprecise associations based on small numbers with variant genotypes under comparison. Among Caucasians, the homozygous variant versus the homozygous wildtype genotypes for two SNPs (rs2453840, rs2453839) in strong LD (pairwise r^2 value = 0.87) were inversely associated with any fibroids (rs2453840: PD = -0.105; 95% CI: -0.390, 0.181; rs2453839: PD = -0.088; 95% CI: -0.354, 0.179), but were positively associated with larger fibroids (at least 2 cm in diameter) (rs2453840: PD = 0.077, 95% CI: -0.208, 0.362; rs2453839: PD = 0.091, 95% CI: -0.175, 0.357) (Table 5.2). In addition, the homozygous variant versus the homoyzygous wildtype genotypes for two SNPs (rs903889, rs2270628) were positively associated with any fibroids (rs903889: PD = 0.095, 95% CI: -0.120, 0.310; rs2270628: PD = 0.104, 95% CI: -0.152, 0.360), but only rs2270628 variants were associated with larger fibroids (PD = 0.149; 95% CI: -0.111, 0.408). These two SNPs (rs903889, rs2270628) were not in LD with each other nor with any SNPs associated with fibroids (pairwise r^2 values < 0.2). Associations with all four SNPs noted above were imprecise based on fewer than 25 observations with homozygous variant genotypes.

Among African Americans, there was an association in the opposite direction to Caucasians between rs903889 (combined GT and GG versus TT genotypes) and any fibroids (PD = -0.113; 95% CI: -0.216, -0.010), but no association with larger fibroids (Table 5.3). In addition, there were fibroid associations with three additional SNPs (rs2471551, rs9282734, rs12671457) among African Americans that were not reported

for Caucasians. The homozygous variant genotype (CC) for rs2471551 was positively associated with any fibroids (PD = 0.208; 95% CI: 0.095, 0.320) and larger fibroids (PD = 0.167; 95% CI: -0.011, 0.345) relative to the homozygous wildtype genotype (GG), with little difference in the prevalence of either fibroid outcome in association with the CG genotype. The combined AC and CC genotypes for the rare rs9283734 (nonsynonymous, MAF ≈ 3%) and the rs12671457 (MAF ≈ 5%) SNPs were inversely associated with prevalence of any fibroids (rs9282734: PD = -0.130; 95%CI: -0.294, 0.034; rs12671457: PD = -0.081, 95% CI: -0.213, 0.050) and larger fibroids (rs9282734: PD = -0.135, 95%CI: -0.303, 0.032; rs12671457: 0.121; 95% CI: -0.259, 0.018) relative to the AA genotype. None of the SNPs with reported fibroid associations among African Americans were in LD with each other (pairwise r^2 values < 0.4).

IGFBP-3 diplotypes

We generally emphasize associations between diplotypes and fibroid prevalence based on posterior medians with magnitudes of at least 10%, and we focus on more precise associations where diplotypes represent at least 10 women. Patterns of *IGFBP-3* diplotype associations among either Caucasians or African Americans could not be explained by associations with individual *IGFBP-3* SNPs. Among Caucasians, *IGFBP-3* group 1 diplotypes (TACGGA/TACGGC, TCGCAC/TCGCAC) relative to the referent diplotype (TACGGC/GCGGAC) were inversely associated with prevalence of any fibroids (TACGGA/TACGGC: -0.102, 95% posterior limits: -0.274, 0.070; TCGCAC/TCGCAC: -0.111, 95% posterior limits: -0.297, 0.076) and larger fibroids (at least 2 cm in diameter) (TACGGA/TACGGC: -0.105, 95% posterior limits: -0.268, 0.057; TCGCAC/TCGCAC: -0.138, 95% posterior limits: -0.308, 0.032) (Figure 5.1).

The *IGFBP-3* group 2 diplotype of TACA/TGCA, was inversely associated with prevalence of any fibroids (-0.078, 95% posterior limits: -0.225, 0.069) and larger fibroids (-0.069, 95% posterior limits: -0.203, 0.066). Despite representing a weaker association with fibroid prevalence (<10%), it was a relatively precise estimate based on greater than 10% (n = 51) of Caucasians with the TACA/TGCA diplotype.

Among African Americans, two *IGFBP-3* group 1 diplotypes (TCGCAGA/GCGGAGA, TCGCAGA/TCGCAGA) were positively associated with prevalence of any fibroids (TCGCAGA/GCGGAGA: 0.112, 95% posterior limits: -0.048, 0.272; TCGCAGA/TCGCAGA: 0.158; 95% posterior limits: 0.035, 0.281) and larger fibroids (TCGCAGA/GCGGAGA: 0.187, 95% posterior limits: 0.026, 0.349; TCGCAGA/TCGCAGA: 0.086, 95% posterior limits: -0.072, 0.243) relative to the referent diplotype (TAGGAGG/TACGAGG) (Figure 5.2). Three *IGFBP-3* group 3 diplotypes (TACA/AGCA, TGCA/TGCA, TGTC/TGCA) were inversely associated with prevalence of larger fibroids (TACA/AGCA: -0.158, 95% posterior limits: -0.336, 0.020; TGCA/TGCA: -0.137, 95% posterior limits: -0.250, -0.023; TGTC/TGCA: -0.120; 95% posterior limits: -0.306, 0.065) relative to the referent diplotype (TATA/TGCA), with slightly smaller associations with any fibroids. Of the three *IGFBP-3* group 3 diplotypes noted above, only the TGCA/TGCA diplotype represented a considerable proportion of African American women in our study population (18%, n = 104).

IGF-I SNPs

We focus on SNPs where fibroid prevalence is increased or decreased by at least 10% with variant genotypes relative to homozygous wildtype genotypes, and we place less emphasis on imprecise associations based on small numbers with variant genotypes

under comparison. Among Caucasians, prevalence of any fibroids was decreased in association with homozygous variant genotypes of rs12821878 (PD = -0.111, 95% CI: -0.335, 0.113) and rs2946834 (PD = -0.119, 95% CI: -0.289, 0.052) and increased in association with homozygous variant genotypes of rs10860869 (PD = 0.133, 95% CI: -0.053, 0.320) and rs7956547 (PD = 0.114, 95% CI: -0.079, 0.308) relative to homozygous wildtype genotypes, with weaker associations estimated in the same direction with larger fibroids (Table 5.4). The pairwise r^2 value for rs10860869 and rs7956547 was 0.91, indicating strong LD among Caucasians, but neither rs12821878 nor rs2946834 were in strong LD with each other or with either of the two former *IGF-I* SNPs. Imprecise associations based on fewer than 20 women with variants were estimated between the prevalence of any fibroids and the heterozygous genotypes (no homozygous variants) for the rare rs5742614 (MAF ≈ 2%) (PD = -0.159; 95% CI: -0.415, 0.097) and rs3730204 (MAF ≈ 2%) (PD = 0.116; 95% CI: -0.127, 0.358) SNPs relative to homozygous wildtype genotypes, but no associations were evident for larger fibroids.

Among African Americans, there were three common SNPs (rs35767, rs5742614, rs4764884) with fibroid associations. Specifically, rs35767 AA versus GG genotypes and rs4764884 TT versus CC genotypes were positively associated with prevalence of any fibroids (rs35767: PD = 0.116, 95% CI: 0.015, 0.216; rs4764884: PD = 0.094, 95% CI: -0.040, 0.227) and larger fibroids (rs35767: PD = 0.104, 95% CI: -0.012, 0.221; rs4764884: 0.097, 95% CI: -0.069, 0.263), which no fibroid associations with either SNP were noted among Caucasians (Table 5.5). The combined CG and CC genotypes versus the GG genotype for rs5742614 were inversely associated with prevalence of any fibroids (PD = -0.120; 95% CI: -0.215, -0.026 and larger fibroids (PD = -0.144; 95% CI: -0.243, -

0.044). Neither of these three SNPs (rs35767, rs5742614, rs4764884) were in LD with each other (pairwise r^2 values < 0.1) among African Americans. Positive associations between two rare (MAF < 5%) SNP variants and prevalence of overall (rs17880975: PD = 0.113, 95% CI: -0.036, 0.262; rs11111262: PD = 0.135, 95% CI: 0.004, 0.266) and larger fibroids (rs17880975: PD = 0.157, 95% CI: -0.024, 0.338; rs11111262: PD = 0.163, 95% CI: -0.005, 0.331) were noted relative to homozygous wildtype genotypes.

IGF-I diplotypes

We highlighted diplotype associations with fibroid prevalence based on posterior medians with magnitudes of at least 10% and more precise associations where diplotypes represent at least 10 women. Among Caucasians, four *IGF-I* group 3 diplotypes (GGCCC/GCTGC: -0.139, 95% posterior limits: -0.290, 0.012; GGCCC/GGCCC: -0.107, 95% posterior limits: -0.236, 0.021; GGCCT/AGTGC: -0.119, 95% posterior limits: -0.298, 0.060; GGCCT/GGCCT: -0.161, 95% posterior limits: -0.313, -0.009) versus referent diplotype (GGCCC/GGCCT) were inversely associated with prevalence of any fibroids (Figure 5.3). Weaker inverse associations with larger fibroids were noted for the GGCCC/GCTGC and GGCCT/GGCCT diplotypes, with a stronger inverse association for the GGCCT/AGTGC diplotype (-0.162; 95% posterior limits: -0.313, -0.011). However, there was no association with larger fibroids for the common GGCCC/GGCCC diplotype representing almost 20% of Caucasians (n = 76). Two *IGF-I* group 4 diplotypes (versus the referent diplotype CG/CG) were inversely associated with prevalence of any fibroids (CA/CA: -0.104, 95% posterior limits: -0.267, 0.059; CG/TA: -0.077, 95% posterior limits: -0.203, 0.050) and larger fibroids (CA/CA: -0.128, 95% posterior limits: -0.273, 0.016; CG/TA: -0.080, 95% posterior limits: -0.198, 0.039). The

latter diplotype (CG/TA) has a weaker association with fibroids (<10%), but estimates are relatively precise based on approximately 15% (n = 59) of Caucasians with CG/TA diplotype. There were no patterns in individual SNP associations to explain any of the estimated Caucasian *IGF-I* diplotype associations.

Among African Americans, the *IGF-I* group 1 GGG/GCG diplotype, which represented approximately 10% of women (n = 55), was inversely associated with prevalence of any fibroids (-0.132, 95% posterior limits: -0.257, -0.007) and larger fibroids (-0.134, 95% posterior limits: -0.261, -0.006) compared to the referent diplotype (AGG/GGG) (Figure 5.4). The *IGF-I* group 2 TTTCT/TCTTT diplotype was inversely associated with any fibroids (-0.131, 95% posterior limits: -0.314, 0.052), while there was a positive association between the TCTTT/TCTTT diplotype and fibroids of any size (0.122, 95% posterior limits: 0.001, 0.243) relative to the referent diplotype (ATTCC/ATTCC), with weaker associations for both diplotypes estimated with larger fibroids. The latter diplotype (TCTTT/TCTTT) was the only common diplotype that included the TT genotype for rs4764884, which was positively associated with fibroids. Three *IGF-I* group 3 diplotypes (versus the referent TACT/TACT diplotype) were inversely associated with any fibroids, including AACC/AGGT (-0.156, 95% posterior limits: -0.340, 0.028), AACT/TACT (-0.101, 95% posterior limits: -0.289, 0.086), and AAGC/AACC (-0.087, 95% posterior limits: -0.232, 0.057), but only the AACC/AGGT (-0.124, 95% posterior limits: -0.303, 0.055) diplotype was associated with larger fibroids. There were no consistent patterns in estimated associations with individual *IGF-I* SNPs and any of the African American diplotypes, with the exception of the *IGF-I* group 3 TCTTT/TCTTT diplotype.

Discussion

This is the first study to evaluate *IGF-I* and *IGFBP-3* polymorphisms in association with the prevalence of fibroids. In summary, we noted fibroid prevalence among African Americans was associated with five *IGF-I* (rs35767, rs5742614, rs17880975, rs4764884, rs11111262) and four *IGFBP-3* (rs903889, rs2471551, rs9282734, rs12671457) SNPs, which included two *IGF-I* (rs17880975, rs11111262) and two *IGFBP-3* SNPs (rs9282734, rs12671457) with MAF 5% or less. However, among Caucasians, fibroid prevalence was associated with six *IGF-I* SNPs (rs5742614, rs12821878, rs10860869, rs7956547, rs3730204, rs2946834) and four *IGFBP-3* SNPs (rs903889, rs2453840, rs2453839, rs2270628); however, the majority of associations, including three *IGF-I* SNPs (rs5742614, rs12821878, rs3730204) and four *IGFBP-3* SNPs, were relatively imprecise due to fewer than 25 observations with variant genotypes.

Only one of the fibroid associations involved an *a priori* selected SNP, the rare (MAF ≈ 3%) nonsynonymous *IGFBP-3* rs9282734 SNP (His158Pro), in which variants were inversely associated with fibroids of any size and at least 2 cm in diameter among African Americans. This SNP could not be evaluated among Caucasians, since less than five women had the variant genotypes. Based on protein sequencing analysis reported in the UniProtKB/Swiss-Prot database (41), the amino acid coded by rs9282734 is not within the protein region affiliated with IGF binding but instead is within a compositionally biased region of the IGFBP-3 protein, which is predominantly rich with serine and threonine. Associations with other health outcomes or circulating IGFBP-3 protein levels have not been reported with this SNP.

We reported strong positive associations between another functionally relevant *IGFBP-3* SNP (rs2471551 CC versus GG genotypes) and fibroid prevalence among African Americans. Specifically, rs2471551 is located at a splice site (intron 1 within less than 20 kb of exon 2). Inverse associations with CC versus GG variants were also estimated with circulating IGFBP levels among Caucasians and African Americans in our study (not published). In addition, Canzian et al.(22) reported that rs2471551 (homozygous variant versus homozygous wildtype genotypes) was inversely associated with circulating IGFBP-3 but not with circulating IGF-I or with breast cancer for a predominantly Caucasian case-control study nested within the European Prospective Investigation into Cancer and Nutrition. Inverse associations were also reported among African Americans between fibroids and two *IGFBP-3* SNPs, with possible functional effects based on their location outside the 5' (rs903889) and 3' (rs12671457) ends of the gene; however, other studies have not reported associations with health outcomes or circulating levels with these SNPs. Our study estimated associations between rs903889 and plasma IGFBP-3 and IGF-I among Caucasians, and associations with plasma IGFBP-3 only among African Americans (unpublished).

Relatively imprecise associations between *IGFBP-3* SNPs (rs903889, rs2453840, rs2453839, rs2270628) and fibroids among Caucasians are likely to be spurious findings, especially given that the association with rs2453830 and rs2453839 are in the opposite direction with larger fibroids. Canzian et al. reported that rs2453839 was not associated with circulating IGF-I or IGFBP-3 levels or with breast cancer (22). Also, Tamimi et al. estimated no association between *IGFBP-3* SNPs, including rs2453839 and rs2270628,

and mammographic density among Nurses' Heath Study controls nested within a breast cancer case-control study (42).

We estimated a positive association between a common *IGF-I* SNP (rs35767 AA versus GG genotypes), which is located within the promoter region, and fibroids among African Americans. Canzian et al. reported inverse associations for rs35767 homozygous variant versus homozygous wildtype genotypes with circulating IGFBP-3 and breast cancer among a predominantly Caucasian study population (22). In addition, a common *IGF-I* SNP (rs2946834 AA versus GG genotypes), which is located downstream of *IGF-I* at the 3' end, was inversely associated with fibroids among Caucasians. Tamimi et al. reported an inverse association for rs2946834 AA versus GG genotypes and percentage mammographic density within the Nurses' Health Study (42).

Current biological evidence supports a role of IGF-I in fibroid pathogenesis, but is lacking for IGFBP-3; however, we hypothesized that IGFBP-3 may indirectly affect fibroid pathogenesis by influencing levels of bioavailable IGF-I. Studies have found increased *IGF-I* mRNA and peptides in fibroids versus normal myometrium (10-16) and in fibroids versus normal myometrium in response to estrogen treatment, as well as decreases in IGF-I mRNA and protein in response to progesterone (43). Overall, these findings support the hypothesis that fibroid tumorigenesis or growth may be promoted by hormonally-mediated alteration of cytokines or growth factors (8). However, since IGF-I is a strong mitogen, we would expect that *IGF-I or IGFBP-3* variant associations with larger fibroids would be stronger than those with fibroids of any size.

From previous unpublished analyses in the NIEHS UFS, plasma IGF-I levels were inversely associated with fibroid prevalence among Caucasians but not African

Americans, while plasma IGFBP-3 levels were not associated with fibroids. We did not find support for fibroid associations with the common nonsynonymous *IGFBP-3* SNP of rs2854746 or with the rs2854744 (-202 A/C) SNP, both of which were associated with plasma IGFBP-3 levels among both Caucasians and African Americans in NIEHS Uterine Fibroid Study (cite 1st paper). However, lack of fibroid associations with circulating IGF-I or IGFBP-3 levels or with SNPs strongly predicting circulating levels does not negate involvement of IGF-I or IGFBP-3 in fibroid etiology since biological evidence for IGF-I in fibroid pathogenesis emphasizes IGF-I levels in the uterine tissue rather than the circulation. In addition, *IGF-I or IGFBP-3* variants associated with circulating IGF-I or IGFBP-3 may not predict levels within uterine tissue.

A major strength of our study was the high proportion of African Americans, who have consistently been reported to have a greater prevalence of fibroids and fibroid-related morbidity than Caucasians (3, 4, 6, 7). Etiologically relevant SNPs between races may differ based on differences in race-specific distributions of respective genotypes. Given racial differences in fibroid prevalence, it was not surprising that we noted more evidence for associations with *IGF-I* and *IGFBP-3* variants among African Americans based on relative precision of estimates within racial groups. We were able to examine associations with not only overall fibroid prevalence, but also with larger fibroids (at least 2 cm in diameter) to assess whether *IGF-I* and *IGFBP-3* variants might influence tumor growth versus incidence. Associations with two fibroid outcomes were generally in the same direction, but typically there were slightly stronger associations with any fibroids rather than larger fibroids, which was not expected since IGF-I is known for its growth-promoting actions. However, differences between analyses were not consistent and large

enough to suggest that genetic variants may be more relevant for incidence rather than growth of fibroids. Our study based fibroid status predominantly on ultrasound screening rather than clinical diagnosis or treatment by hysterectomy alone, which reduced the potential for misclassification bias if *IGF-I* or *IGFBP-3* variants were associated with the incidence of all fibroid tumors, regardless of size.

We restricted our study to Caucasians and African Americans and stratified all analyses by race; however, population stratification within each racial group may still be a potential bias with our study (44-47). In addition, we did not have statistical power to explore whether gene-environmental interactions could further explain findings. Race-specific diplotype associations with fibroids generally did not support patterns with SNP associations, suggesting that diplotype associations may be either due to spurious findings or if etiologically relevant may be due to interactions between the SNPs within the particular diplotype group. We attempted to reduce spurious findings for less common diplotypes by using a Bayesian method of information-weighted averaging that increased the precision of estimates, particularly for diplotypes with small numbers of women; however, these methods were not applied to individual SNP associations. Despite the increase in bias for posterior medians from this method, simulation studies indicate that there is a reduction in the overall mean square error based on the greater decrease in variance of estimates (48).

Since fibroids represent a common health outcome related to substantial morbidity, even *IGF-I* or *IGFBP-3* variant associations with magnitudes of less than 20% as reported in our study could be important. Some of the reported SNP or diplotype associations may reflect random error or bias, especially associations based on small

numbers with SNP variant or diplotype. Validation of findings with other study populations well-represented by African Americans would be important for assessing whether any of the *IGF-I* and *IGFBP-3* SNPs are relevant with fibroid etiology. Future research should also consider other genes besides *IGF-I* and *IGFBP-3* that are involved with the IGF-I pathway based on support for IGF-I involvement with fibroid pathogenesis.

REFERENCES

1. Farquhar, C. M., and Steiner, C. A. Hysterectomy rates in the United States 1990-1997. Obstet Gynecol, *99:* 229-34, 2002.

2. Wilcox, L. S., Koonin, L. M., Pokras, R., Strauss, L. T., Xia, Z., and Peterson, H. B. Hysterectomy in the United States, 1988-1990. Obstet Gynecol, *83:* 549-55, 1994.

3. Baird, D. D., Dunson, D. B., Hill, M. C., Cousins, D., and Schectman, J. M. High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence. Am J Obstet Gynecol, *188:* 100-7, 2003.

4. Marshall, L. M., Spiegelman, D., Barbieri, R. L., Goldman, M. B., Manson, J. E., Colditz, G. A., Willett, W. C., and Hunter, D. J. Variation in the incidence of uterine leiomyoma among premenopausal women by age and race. Obstet Gynecol, *90:* 967-73, 1997.

5. Wise, L. A., Palmer, J. R., Stewart, E. A., and Rosenberg, L. Age-specific incidence rates for self-reported uterine leiomyomata in the Black Women's Health Study. Obstet Gynecol, *105:* 563-8, 2005.

6. Samadi, A. R., Lee, N. C., Flanders, W. D., Boring, J. R., 3rd, and Parris, E. B. Risk factors for self-reported uterine fibroids: a case-control study. Am J Public Health, *86:* 858-62, 1996.

7. Faerstein, E., Szklo, M., and Rosenshein, N. Risk factors for uterine leiomyoma: a practice-based case-control study. I. African-American heritage, reproductive history, body size, and smoking. Am J Epidemiol, *153:* 1-10, 2001.

8. Sozen, I., and Arici, A. Cellular biology of myomas: interaction of sex steroids with cytokines and growth factors. Obstet Gynecol Clin North Am, *33:* 41-58, 2006.

9. Strawn, E. Y., Jr., Novy, M. J., Burry, K. A., and Bethea, C. L. Insulin-like growth factor I promotes leiomyoma cell growth in vitro. Am J Obstet Gynecol, *172:* 1837-43; discussion 1843-4, 1995.

10. Wolanska, M., and Bankowski, E. An accumulation of insulin-like growth factor I (IGF-I) in human myometrium and uterine leiomyomas in various stages of tumour growth. Eur Cytokine Netw, *15:* 359-63, 2004.

11. Van der Ven, L. T., Roholl, P. J., Gloudemans, T., Van Buul-Offers, S. C., Welters, M. J., Bladergroen, B. A., Faber, J. A., Sussenbach, J. S., and Den Otter, W. Expression of insulin-like growth factors (IGFs), their receptors and IGF

binding protein-3 in normal, benign and malignant smooth muscle tissues. Br J Cancer, *75:* 1631-40, 1997.

12. Van der Ven, L. T., Gloudemans, T., Roholl, P. J., Van Buul-Offers, S. C., Bladergroen, B. A., Welters, M. J., Sussenbach, J. S., and den Otter, W. Growth advantage of human leiomyoma cells compared to normal smooth-muscle cells due to enhanced sensitivity toward insulin-like growth factor I. Int J Cancer, *59:* 427-34, 1994.

13. Dixon, D., He, H., and Haseman, J. K. Immunohistochemical localization of growth factors and their receptors in uterine leiomyomas and matched myometrium. Environ Health Perspect, *108 Suppl 5:* 795-802, 2000.

14. Englund, K., Lindblom, B., Carlstrom, K., Gustavsson, I., Sjoblom, P., and Blanck, A. Gene expression and tissue concentrations of IGF-I in human myometrium and fibroids under different hormonal conditions. Mol Hum Reprod, *6:* 915-20, 2000.

15. Hoppener, J. W., Mosselman, S., Roholl, P. J., Lambrechts, C., Slebos, R. J., de Pagter-Holthuizen, P., Lips, C. J., Jansz, H. S., and Sussenbach, J. S. Expression of insulin-like growth factor-I and -II genes in human smooth muscle tumours. Embo J, *7:* 1379-85, 1988.

16. Giudice, L. C., Irwin, J. C., Dsupin, B. A., Pannier, E. M., Jin, I. H., Vu, T. H., and Hoffman, A. R. Insulin-like growth factor (IGF), IGF binding protein (IGFBP), and IGF receptor gene expression and IGFBP synthesis in human uterine leiomyomata. Hum Reprod, *8:* 1796-806, 1993.

17. Swartz, C. D., Afshari, C. A., Yu, L., Hall, K. E., and Dixon, D. Estrogen-induced changes in IGF-I, Myb family and MAP kinase pathway genes in human uterine leiomyoma and normal uterine smooth muscle cell lines. Mol Hum Reprod, *11:* 441-50, 2005.

18. Juul, A., Main, K., Blum, W. F., Lindholm, J., Ranke, M. B., and Skakkebaek, N. E. The ratio between serum levels of insulin-like growth factor (IGF)-I and the IGF binding proteins (IGFBP-1, 2 and 3) decreases with age in healthy adults and is increased in acromegalic patients. Clin Endocrinol (Oxf), *41:* 85-93, 1994.

19. Baird, D. D., Dunson, D. B., Hill, M. C., Cousins, D., and Schectman, J. M. Association of physical activity with development of uterine leiomyoma. Am J Epidemiol, *165:* 157-63, 2007.

20. Baird, D. D., Kesner, J. S., and Dunson, D. B. Luteinizing hormone in premenopausal women may stimulate uterine leiomyomata development. J Soc Gynecol Investig, *13:* 130-5, 2006.

21. Seattle SNPs Program for Genomic Applications (PGA). Genome Variation Server (GVS) Version 1.04. pp. http://gvs.gs.washington.edu/GVS/, 2006.

22. Canzian, F., McKay, J. D., Cleveland, R. J., Dossus, L., Biessy, C., Rinaldi, S., Landi, S., Boillot, C., Monnier, S., Chajes, V., Clavel-Chapelon, F., Tehard, B., Chang-Claude, J., Linseisen, J., Lahmann, P. H., Pischon, T., Trichopoulos, D., Trichopoulou, A., Zilis, D., Palli, D., Tumino, R., Vineis, P., Berrino, F., Bueno-de-Mesquita, H. B., van Gils, C. H., Peeters, P. H., Pera, G., Ardanaz, E., Chirlaque, M. D., Quiros, J. R., Larranaga, N., Martinez-Garcia, C., Allen, N. E., Key, T. J., Bingham, S. A., Khaw, K. T., Slimani, N., Norat, T., Riboli, E., and Kaaks, R. Polymorphisms of genes coding for insulin-like growth factor 1 and its major binding proteins, circulating levels of IGF-I and IGFBP-3 and breast cancer risk: results from the EPIC study. Br J Cancer, *94:* 299-307, 2006.

23. Deal, C., Ma, J., Wilkin, F., Paquette, J., Rozen, F., Ge, B., Hudson, T., Stampfer, M., and Pollak, M. Novel promoter polymorphism in insulin-like growth factor-binding protein-3: correlation with serum levels and interaction with known regulators. J Clin Endocrinol Metab, *86:* 1274-80, 2001.

24. Al-Zahrani, A., Sandhu, M. S., Luben, R. N., Thompson, D., Baynes, C., Pooley, K. A., Luccarini, C., Munday, H., Perkins, B., Smith, P., Pharoah, P. D., Wareham, N. J., Easton, D. F., Ponder, B. A., and Dunning, A. M. IGF1 and IGFBP3 tagging polymorphisms are associated with circulating levels of IGF1, IGFBP3 and risk of breast cancer. Hum Mol Genet, *15:* 1-10, 2006.

25. Jernstrom, H., Deal, C., Wilkin, F., Chu, W., Tao, Y., Majeed, N., Hudson, T., Narod, S. A., and Pollak, M. Genetic and nongenetic factors associated with variation of plasma levels of insulin-like growth factor-I and insulin-like growth factor-binding protein-3 in healthy premenopausal women. Cancer Epidemiol Biomarkers Prev, *10:* 377-84, 2001.

26. Lai, J. H., Vesprini, D., Zhang, W., Yaffe, M. J., Pollak, M., and Narod, S. A. A polymorphic locus in the promoter region of the IGFBP3 gene is related to mammographic breast density. Cancer Epidemiol Biomarkers Prev, *13:* 573-82, 2004.

27. Schernhammer, E. S., Hankinson, S. E., Hunter, D. J., Blouin, M. J., and Pollak, M. N. Polymorphic variation at the -202 locus in IGFBP3: Influence on serum levels of insulin-like growth factors, interaction with plasma retinol and vitamin D and breast cancer risk. Int J Cancer, *107:* 60-4, 2003.

28. Cheng, I., Henderson, K. D., Haiman, C. A., Kolonel, L. N., Henderson, B. E., Freedman, M. L., and Le Marchand, L. Genetic Determinants of Circulating IGF-I, IGFBP-1 and IGFBP-3 Levels in a Multiethnic Population. J Clin Endocrinol Metab, 2007.

29. Ren, Z., Cai, Q., Shu, X. O., Cai, H., Li, C., Yu, H., Gao, Y. T., and Zheng, W. Genetic polymorphisms in the IGFBP3 gene: association with breast cancer risk and blood IGFBP-3 protein levels among Chinese women. Cancer Epidemiol Biomarkers Prev, *13:* 1290-5, 2004.

30. Behrens, M., and Lange, R. A highly reproducible and economically competitive SNP analysis of several well characterized human mutations. Clin Lab, *50:* 305-16, 2004.

31. Livak, K. J. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. Genet Anal, *14:* 143-9, 1999.

32. McGuigan, F. E., and Ralston, S. H. Single nucleotide polymorphism detection: allelic discrimination using TaqMan. Psychiatr Genet, *12:* 133-6, 2002.

33. Wigginton, J. E., Cutler, D. J., and Abecasis, G. R. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet, *76:* 887-93, 2005.

34. Barrett, J. C., Fry, B., Maller, J., and Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics, *21:* 263-5, 2005.

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

36. de Bakker, P. I., Yelensky, R., Pe'er, I., Gabriel, S. B., Daly, M. J., and Altshuler, D. Efficiency and power in genetic association studies. Nat Genet, *37:* 1217-23, 2005.

37. Stephens, M., and Donnelly, P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet, *73:* 1162-9, 2003.

38. Stephens, M., Smith, N. J., and Donnelly, P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet, *68:* 978-89, 2001.

39. Spiegelman, D., and Hertzmark, E. Easy SAS calculations for risk or prevalence ratios and differences. Am J Epidemiol, *162:* 199-200, 2005.

40. Greenland, S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. Int J Epidemiol, *35:* 765-75, 2006.

41. Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N., and Yeh, L. S. The Universal Protein Resource (UniProt). Nucleic Acids Res, *33:* D154-9, 2005.

42. Tamimi, R. M., Cox, D. G., Kraft, P., Pollak, M. N., Haiman, C. A., Cheng, I., Freedman, M. L., Hankinson, S. E., Hunter, D. J., and Colditz, G. A. Common genetic variation in IGF1, IGFBP-1, and IGFBP-3 in relation to mammographic density: a cross-sectional study. Breast Cancer Res, *9:* R18, 2007.

43. Yamada, T., Nakago, S., Kurachi, O., Wang, J., Takekida, S., Matsuo, H., and Maruo, T. Progesterone down-regulates insulin-like growth factor-I expression in cultured human uterine leiomyoma cells. Hum Reprod, *19:* 815-21, 2004.

44. Wang, Y., Localio, R., and Rebbeck, T. R. Evaluating bias due to population stratification in case-control association studies of admixed populations. Genet Epidemiol, *27:* 14-20, 2004.

45. Wang, Y., Localio, R., and Rebbeck, T. R. Evaluating bias due to population stratification in epidemiologic studies of gene-gene or gene-environment interactions. Cancer Epidemiol Biomarkers Prev, *15:* 124-32, 2006.

46. Ardlie, K. G., Lunetta, K. L., and Seielstad, M. Testing for population subdivision and association in four case-control studies. Am J Hum Genet, *71:* 304-11, 2002.

47. Wacholder, S., Rothman, N., and Caporaso, N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst, *92:* 1151-8, 2000.

48. Greenland, S., and Poole, C. Empirical-Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. Arch Environ Health, *49:* 9-16, 1994.

Table 5.1. Uterine fibroid status of premenopausal women with genotype information by race from NIEHS Uterine Fibroid Study

Uterine fibroids	African American (N=582) N (%)	Caucasian (N=402) N (%)
None	154 (26.5)	194 (48.3)
Any	421 (72.3)	200 (49.8)
<2 cm	94 (16.2)	69 (17.2)
2 - <4 cm	187 (32.1)	87 (21.6)
≥4 cm	140 (24.1)	44 (10.9)
Missing	7 (1.2)	8 (2.0)

Table 5.2. Unadjusted prevalence differences (PD) and 95% confidence intervals (CI) for uterine fibroids in association with *IGFBP-3* single nucleotide polymorphisms (SNPs) among Caucasians

Block*	Diplotype Group*	Diplotype Position*	SNP Reference ID	MAF (%)†	Genotype	N(any fibroids): total N‡	Uterine Fibroids (any)		N(2+ cm fibroids): total N‡	Uterine Fibroids (2+ cm)	
							PD§	95% CI		PD§	95% CI
1	1	1	rs903889	22.0	GG	13:22	0.095	-0.120, 0.310	7:22	-0.008	-0.212, 0.195
					GT	67:128	0.028	-0.080, 0.135	45:128	0.025	-0.077, 0.127
					TT	120:242	0		79:242	0	
1	--	--	rs924140	46.8	TT	38:80	-0.039	-0.184, 0.106	23:80	-0.012	-0.143, 0.120
					CT	106:205	0.003	-0.114, 0.120	76:205	0.072	-0.037, 0.181
					CC	55:107	0		32:107	0	
1	1	2	rs2854744‖	46.3	AA	36:77	-0.047	-0.193, 0.100	22:77	-0.013	-0.146, 0.120
					AC	104:199	0.009	-0.109, 0.126	75:199	0.078	-0.032, 0.188
					CC	55:107	0		32:107	0	
1	1	3	rs2854746‖	41.7	CC	32:63	-0.019	-0.170, 0.131	18:63	-0.040	-0.178, 0.098
					CG	99:199	-0.030	-0.140, 0.081	71:199	0.031	-0.074, 0.136
					GG	68:129	0		42:129	0	
1	1	4	rs2471551	19.4	CC	7:16	-0.063	-0.313, 0.188	5:16	-0.026	-0.261, 0.208
					CG	64:119	0.038	-0.071, 0.147	39:119	-0.011	-0.113, 0.092
					GG	127:254	0		86:254	0	
--	--	--	rs9282734‖	0.4	AC	1:3	--		1:3	--	
					AA	199:386	0		130:386	0	
--	--	--	rs2453837	0	GG	199:390	--		130:390	--	
--	--	--	rs6953668	0.6	AG	1:5	--		1:5	--	
					GG	199:387	0		130:387	0	
1	1	5	rs3110697	41.4	AA	32:62	0.036	-0.116, 0.187	18:62	-0.043	-0.182, 0.096
					AG	106:198	0.055	-0.056, 0.166	70:198	0.020	-0.085, 0.125
					GG	62:129	0		43:129	0	

2	1	6	rs2453840	18.3	AA	5:12	-0.105	-0.390, 0.181	5:12	0.077	-0.208, 0.362
					AC	59:118	-0.021	-0.130, 0.088	38:118	-0.018	-0.120, 0.084
					CC	135:259	0		88:259	0	
2	--	--	rs2453839	20.1	CC	6:14	-0.088	-0.354, 0.179	6:14	0.091	-0.175, 0.357
					CT	64:128	-0.016	-0.123, 0.091	42:128	-0.009	-0.110, 0.091
					TT	127:246	0		83:246	0	
Outside	2	1	rs6670	21.2	AT, AA	85:155	0.068	-0.034, 0.169	57:155	0.053	-0.044, 0.149
					TT	113:235	0		74:235	0	
Outside	2	2	rs13223993	21.4	AA	9:19	-0.041	-0.274, 0.192	5:19	-0.069	-0.276, 0.138
					AG	65:127	-0.003	-0.110, 0.105	45:127	0.022	-0.080, 0.125
					GG	124:241	0		80:241	0	
3	2	3	rs2270628	21.0	TT	9:15	0.104	-0.152, 0.360	7:15	0.149	-0.111, 0.408
					CT	71:135	0.030	-0.075, 0.135	47:135	0.030	-0.070, 0.130
					CC	120:242	0		77:242	0	
3	2	4	rs12671457	16.5	AC, CC	61:117	0.020	-0.090, 0.128	40:117	0.009	-0.095, 0.112
					AA	131:261	0		87:261	0	

NOTE: MAF, minor allele frequency.

* Blocks consist of SNPs with MAF ≥ 5% in strong linkage disequilibrium (LD) (95% of pairwise SNP comparisons with one-sided 95% confidence intervals for the D prime statistic within 0.7-0.98). Group refers to combination of SNPs for diplotype estimation. Position refers to SNP order in each group for diplotype estimation.; the position for each group begins with "1". Values for diplotype group and position (--) were not listed for SNPs excluded from diplotype estimation based on either: 1) Tagger algorithm ($r^2 \geq 0.8$); 2) MAF < 5% (excluded from blocks (--)). rs2453840 (block 2) was combined with block 1 SNPs for diplotype estimation. rs6670 and rs13223993 were combined with block 3 SNPs for diplotype estimation.

† Based on total N = 402.

‡ Total N = 394, excluded 8 women missing uterine fibroid status.

§ PDs were not estimated for SNPs with 5 or fewer women with combined heterozygous and homozygous variant genotypes.

‖ SNPs selected *a priori*. rs2854744 is also known as -202 A/C; rs2854746 and rs9282734, nonsynonymous SNPs.

Table 5.3. Unadjusted prevalence differences (PD) and 95% confidence intervals (CI) for uterine fibroids in association with *IGFBP-3* single nucleotide polymorphisms (SNPs) among African Americans

Block*	Diplotype Group*	Diplotype Position*	SNP Reference ID	MAF (%)†	Genotypes	N(any fibroids): total N‡	Uterine Fibroids (any)		N(2+ cm fibroids): total N‡	Uterine Fibroids (2+ cm)	
							PD	95% CI		PD	95% CI
Outside	1	1	rs903889	9.4	GT, GG	62:97	-0.113	-0.216, -0.010	53:97	-0.030	-0.139, 0.079
					TT	355:472	0		272:472	0	
1	--	--	rs924140	38.9	CC	71:93	0.021	-0.083, 0.125	59:93	0.060	-0.058, 0.177
					CT	181:255	-0.032	-0.113, 0.048	137:255	-0.037	-0.127, 0.052
					TT	164:221	0		127:221	0	
1	1	2	rs2854744§	42.2	CC	77:105	-0.004	-0.109, 0.101	64:105	0.006	-0.110, 0.122
					AC	189:262	-0.016	-0.098, 0.067	138:262	-0.076	-0.168, 0.015
					AA	143:194	0		117:194	0	
1	1	3	rs2854746§	32.1	CC	39:52	0.004	-0.125, 0.133	31:52	0.002	-0.144, 0.149
					CG	185:260	-0.035	-0.111, 0.042	141:260	-0.051	-0.137, 0.034
					GG	191:256	0		152:256	0	
1	1	4	rs2471551	20.3	CC	24:26	0.208	0.095, 0.320	19:26	0.167	-0.011, 0.345
					CG	130:177	0.019	-0.061, 0.099	98:177	-0.010	-0.099, 0.079
					GG	259:362	0		204:362	0	
--	1	5	rs9282734§	3.3	AC, CC	22:36	-0.130	-0.294, 0.034	16:36	-0.135	-0.303, 0.032
					AA	395:533	0		309:533	0	
--	--	--	rs2453837	0	GG	417:569	--		323:569	--	
1	1	6	rs6953668	5.0	AG, AA	36:54	-0.069	-0.201, 0.062	29:54	-0.036	-0.176, 0.104
					GG	379:515	0		295:515	0	
Outside	1	7	rs3110697	36.1	AA	59:78	0.021	-0.089, 0.132	47:78	0.044	-0.082, 0.169
					AG	181:252	-0.017	-0.096, 0.062	142:252	0.005	-0.083, 0.093
					GG	175:238	0		133:238	0	
2	2	1	rs2453840	9.2	AC, AA	68:98	-0.047	-0.147, 0.052	54:98	-0.022	-0.130, 0.086
					CC	349:471	0		270:471	0	

2	2	2	rs2453839	40.9	CC	69:94	-0.004	-0.113, 0.104	55:94	0.036	-0.085, 0.158
					CT	200:278	-0.019	-0.100, 0.062	158:278	0.020	-0.071, 0.111
					TT	144:195	0		107:195	0	
3	3	1	rs6670	12.6	AT, AA	102:135	0.033	-0.051, 0.117	79:135	0.019	-0.076, 0.115
					TT	313:433	0		245:433	0	
3	3	2	rs13223993	43.6	AA	87:114	0.044	-0.057, 0.145	68:114	0.067	-0.047, 0.182
					AG	192:262	0.013	-0.070, 0.097	155:262	0.063	-0.030, 0.155
					GG	136:189	0		100:189	0	
4	3	3	rs2270628	36.2	TT	62:81	0.056	-0.053, 0.165	45:81	0.012	-0.113, 0.137
					CT	183:245	0.037	-0.042, 0.116	148:245	0.061	-0.027, 0.148
					CC	171:241	0		131:241	0	
4	3	4	rs12671457	5.1	AC, CC	36:55	-0.081	-0.213, 0.050	25:55	-0.121	-0.259, 0.018
					AA	376:511	0		294:511	0	

NOTE: MAF, minor allele frequency.

* Blocks consist of SNPs with MAF ≥ 5% in strong linkage disequilibrium (LD) (95% of pairwise SNP comparisons with one-sided 95% confidence intervals for the D prime statistic within 0.7-0.98). Group refers to combination of SNPs for diplotype estimation. Position refers to SNP order in each group for diplotype estimation; the position for each group begins with "1". Values for diplotype group and position (--) were not listed for SNPs excluded from diplotype estimation based on either: 1) Tagger algorithm ($r^2 \geq 0.8$); 2) MAF < 5% (excluded from blocks (--)) or MAF < 3% for SNPs selected *a priori*. rs903889 and rs3110697 were combined with block 1 SNPs for diplotype estimation. SNPs in blocks 3 and 4 were combined together for diplotype estimation.

† Based on total N = 582.

‡ Total N = 575, excluded 7 women missing uterine fibroid status.

§ SNPs selected *a priori*. rs2854744 is also known as -202 A/C; rs2854746 and rs9282734, nonsynonymous SNPs.

Table 5.4. Unadjusted prevalence differences (PD) and 95% confidence intervals (CI) for uterine fibroids in association with *IGF-I* single nucleotide polymorphisms (SNPs) among Caucasians

Block*	Diplotype Group*	Diplotype Position*	SNP Reference ID	MAF (%)†	Genotypes	N(any fibroids): total N‡	Uterine Fibroids (any)		N(2+ cm fibroids): total N‡	Uterine Fibroids (2+ cm)	
							PD§	95% CI		PD§	95% CI
Outside	1	1	rs35767	16.0	AG, AA	59:116	-0.001	-0.109, 0.108	40:116	0.012	-0.092, 0.115
					GG	139:273	0		91:273	0	
--	--	--	rs5742612	3.9	AG, GG	14:30	-0.046	-0.232, 0.140	8:30	-0.074	-0.240, 0.092
					AA	185:361	0		123:361	0	
--	--	--	rs5742614	1.8	CG	5:14	-0.159	-0.415, 0.097	4:14	-0.050	-0.292, 0.191
					GG	195:378	0		127:378	0	
--	--	--	rs3729846‖	0	CC	200:392	--		131:392	--	
1	2	1	rs12821878	22.7	AA	8:20	-0.111	-0.335, 0.113	5:20	-0.082	-0.281, 0.117
					AG	70:135	0.008	-0.098, 0.114	47:135	0.016	-0.084, 0.117
					GG	120:235	0		78:235	0	
1	--	--	rs10860869	29.1	TT	19:30	0.133	-0.053, 0.320	13:30	0.098	-0.091, 0.288
					AT	83:166	0.000	-0.104, 0.104	52:166	-0.022	-0.119, 0.075
					AA	97:194	0		65:194	0	
1	2	2	rs1019731	12.3	AC, AA	41:89	-0.064	-0.182, 0.054	28:89	-0.028	-0.138, 0.083
					CC	158:301	0		103:301	0	
1	2	3	rs7956547	27.3	CC	17:28	0.114	-0.079, 0.308	11:28	0.069	-0.123, 0.261
					CT	81:156	0.027	-0.077, 0.130	53:156	0.016	-0.082, 0.114
					TT	102:207	0		67:207	0	
--	--	--	rs5742626	0.5	CT	3:4	--		2:4	--	
					TT	196:387	0		129:387	0	
--	--	--	rs17880975	0.1	AG	1:1	--		1:1	--	
					GG	199:388	0		130:388	0	
2	3	1	rs2033178	6.2	AG, AA	22:48	-0.056	-0.207, 0.094	16:48	0.000	-0.142, 0.142
					GG	176:342	0		114:342	0	
--	--	--	rs17884646	0	TT	199:390	--		131:390	--	

--	--	--	rs5742657	2.2	AG, GG	8:17	-0.040	-0.283, 0.203	5:17	-0.041	-0.263, 0.181
					AA	192:376	0		126:376	0	
--	--	--	rs5742663	0.1	GT	0:1	--		0:1	--	
					TT	200:389	0		131:389	0	
--	--	--	rs11829586	2.3	AG	9:18	-0.009	-0.246, 0.227	6:18	-0.002	-0.225, 0.221
					GG	190:373	0		125:373	0	
2	--	--	rs4764884	25.7	TT	11:24	-0.053	-0.264, 0.157	8:24	-0.002	-0.200, 0.197
					CT	77:150	0.002	-0.103, 0.106	50:150	-0.002	-0.100, 0.097
					CC	110:215	0		72:215	0	
--	--	--	rs5742683	0.1	AG	0:1	--		0:1	--	
					AA	198:389	0		129:389	0	
--	--	--	rs17884626‖	0	CC	200:392	--		131:392	--	
2	--	--	rs5009837	30.3	TT	20:35	0.048	-0.131, 0.226	14:35	0.051	-0.125, 0.227
					CT	80:164	-0.036	-0.141, 0.069	51:164	-0.038	-0.136, 0.060
					CC	99:189	0		66:189	0	
2	3	2	rs17727841	19.1	CC	6:14	-0.081	-0.347, 0.185	4:14	-0.050	-0.294, 0.193
					CG	61:119	0.003	-0.106, 0.112	39:119	-0.008	-0.110, 0.094
					GG	132:259	0		87:259	0	
2	--	--	rs4764883	30.9	CC	20:36	0.020	-0.158, 0.198	14:36	0.034	-0.140, 0.207
					CT	78:164	-0.060	-0.165, 0.045	50:164	-0.050	-0.149, 0.049
					TT	98:183	0		65:183	0	
2	--	--	rs9308315	28.2	AA	13:27	-0.044	-0.245, 0.157	10:27	0.022	-0.172, 0.216
					AT	80:163	-0.035	-0.138, 0.069	52:163	-0.030	-0.127, 0.068
					TT	104:198	0		69:198	0	
2	3	3	rs978458	28.0	TT	13:27	-0.041	-0.242, 0.160	10:27	0.024	-0.170, 0.217
					CT	80:161	-0.026	-0.130, 0.078	52:161	-0.024	-0.122, 0.074
					CC	104:199	0		69:199	0	
--	--	--	rs5742692	2.1	AG	8:17	-0.041	-0.284, 0.201	5:17	-0.042	-0.264, 0.180
					AA	192:375	0		126:375	0	
2	--	--	rs11111262	9.9	AG, AA	33:73	-0.068	-0.195, 0.058	24:73	-0.006	-0.125, 0.114
					GG	165:317	0		106:317	0	

2	3	4	rs1520220	19.1	GG	7:13	0.005	-0.273, 0.283	7:13	0.195	-0.083, 0.472
					CG	55:122	-0.083	-0.190, 0.025	36:122	-0.049	-0.149, 0.051
					CC	135:253	0		87:253	0	
--	--	--	rs3730204	2.3	CT	10:16	0.116	-0.127, 0.358	5:16	-0.025	-0.257, 0.207
					TT	190:373	0		126:373	0	
Outside	3	5	rs6214	39.8	TT	30:66	-0.014	-0.159, 0.131	18:66	-0.086	-0.219, 0.047
					CT	102:180	0.098	-0.011, 0.207	61:180	-0.020	-0.124, 0.085
					CC	68:145	0		52:145	0	
3	4	1	rs6219	10.0	CT, TT	35:75	-0.054	-0.179, 0.072	24:75	-0.018	-0.135, 0.100
					CC	165:317	0		107:317	0	
3	4	2	rs2946834	33.0	AA	17:40	-0.119	-0.289, 0.052	12:40	-0.074	-0.234, 0.085
					AG	88:177	-0.047	-0.152, 0.058	55:177	-0.064	-0.163, 0.036
					GG	93:171	0		64:171	0	

NOTE: MAF, minor allele frequency.

* Blocks consist of SNPs with MAF ≥ 5% in strong linkage disequilibrium (LD) (95% of pairwise SNP comparisons with one-sided 95% confidence intervals for the D prime statistic within 0.7-0.98). Group refers to SNP combination for diplotype estimation. Position refers to SNP order in each group for diplotype estimation; the position for each group begins with "1". Values for diplotype group and position (--) were not listed for SNPs excluded from diplotype estimation based on either: 1) Tagger algorithm ($r^2 \geq 0.8$); 2) MAF < 5% (excluded from blocks (--)). rs6214 was combined with block 2 SNPs for diplotype estimation.

† Based on total N = 402.

‡ Total N = 394, excluded 8 women missing uterine fibroid status.

§ PDs were not estimated for SNPs with 5 or fewer women with combined heterozygous and homozygous variant genotypes.

‖ SNPs selected *a priori*. rs3729846, synonymous SNP; rs17884626, nonsynonymous SNP.

Table 5.5. Unadjusted prevalence differences (PD) and 95% confidence intervals (CI) for uterine fibroids in association with *IGF-I* single nucleotide polymorphisms (SNPs) among African Americans

Block*	Diplotype Group*	Diplotype Position*	SNP Reference ID	MAF (%)†	Genotypes	N(any fibroids): total N‡	Uterine Fibroids (any)		N(2+ cm fibroids): total N‡	Uterine Fibroids (2+ cm)	
							PD§	95% CI		PD§	95% CI
1	1	1	rs35767	40.8	AA	81:101	0.116	0.015, 0.216	64:101	0.104	-0.012, 0.221
					AG	195:264	0.052	-0.031, 0.135	151:264	0.043	-0.048, 0.133
					GG	140:204	0		108:204	0	
--	--	--	rs5742612	3.3	AG, GG	27:36	0.021	-0.126, 0.167	22:36	0.048	-0.117, 0.212
					AA	391:536	0		302:536	0	
1	1	2	rs5742614	11.5	CG, CC	77:121	-0.120	-0.215, -0.026	55:121	-0.144	-0.243, -0.044
					GG	339:448	0		268:448	0	
--	--	--	rs3729846‖	0.8	CT	6:9	-0.066	-0.377, 0.244	4:9	-0.125	-0.453, 0.202
					CC	409:558	0		318:558	0	
Outside	1	3	rs12821878	5.8	AG, AA	47:63	0.014	-0.100, 0.129	37:63	0.021	-0.108, 0.150
					GG	371:507	0		287:507	0	
Outside	2	1	rs10860869	32.3	TT	42:61	-0.038	-0.167, 0.090	34:61	0.015	-0.124, 0.154
					AT	186:250	0.017	-0.060, 0.094	149:250	0.054	-0.032, 0.140
					AA	189:260	0		141:260	0	
--	--	--	rs1019731	3.4	AC	26:38	-0.050	-0.202, 0.103	20:38	-0.045	-0.210, 0.119
					CC	389:530	0		303:530	0	
2	2	2	rs7956547	26.2	CC	35:44	0.071	-0.058, 0.200	27:44	0.066	-0.089, 0.220
					CT	154:212	0.002	-0.076, 0.080	123:212	0.032	-0.054, 0.119
					TT	226:312	0		171:312	0	
2	2	3	rs5742626	7.2	CT, CC	60:78	0.046	-0.056, 0.147	43:78	-0.016	-0.135, 0.103
					TT	356:492	0		279:492	0	
--	--	--	rs17880975	2.3	AG	21:25	0.113	-0.036, 0.262	18:25	0.157	-0.024, 0.338
					GG	394:542	0		305:542	0	
--	--	--	rs2033178	4.7	AG, AA	33:50	-0.079	-0.216, 0.057	27:50	-0.030	-0.175, 0.115
					GG	380:514	0		293:514	0	

--	--	--	rs17884646	0	TT	417:571	--		324:571	--	
2	--	--	rs5742657	12.1	AG, GG	91:128	-0.024	-0.112, 0.065	73:128	0.006	-0.092, 0.103
					AA	324:441	0		249:441	0	
2	--	--	rs5742663	11.6	GT, GG	88:123	-0.019	-0.109, 0.070	70:123	0.003	-0.096, 0.102
					TT	327:445	0		252:445	0	
2	--	--	rs11829586	11.5	AG, AA	87:122	-0.022	-0.112, 0.068	71:122	0.018	-0.081, 0.117
					GG	327:445	0		251:445	0	
2	2	4	rs4764884	23.7	TT	29:35	0.094	-0.040, 0.227	23:35	0.097	-0.069, 0.263
					CT	142:201	-0.029	-0.107, 0.050	113:201	0.002	-0.085, 0.089
					CC	244:332	0		186:332	0	
2	--	--	rs5742683	10.6	AG, GG	80:113	-0.028	-0.121, 0.065	63:113	-0.012	-0.114, 0.091
					AA	335:455	0		259:455	0	
--	--	--	rs17884626‖	1.0	CT	9:11	0.088	-0.143, 0.319	8:11	0.163	-0.103, 0.429
					CC	409:560	0		316:560	0	
Outside	2	5	rs5009837	41.3	TT	77:107	0.007	-0.099, 0.112	62:107	0.050	-0.066, 0.166
					CT	195:260	0.037	-0.045, 0.119	153:260	0.059	-0.032, 0.150
					CC	144:202	0		107:202	0	
--	--	--	rs17727841	4.7	CG, CC	37:51	-0.005	-0.133, 0.124	29:51	0.004	-0.138, 0.147
					GG	379:519	0		293:519	0	
3	--	--	rs4764883	45.8	TT	97:133	0.026	-0.075, 0.127	70:133	-0.012	-0.124, 0.099
					CT	186:245	0.056	-0.029, 0.141	151:245	0.078	-0.017, 0.173
					CC	128:182	0		98:182	0	
3	3	1	rs9308315	49.8	TT	112:152	0.047	-0.055, 0.148	80:152	0.004	-0.108, 0.116
					AT	195:258	0.066	-0.024, 0.155	161:258	0.102	0.003, 0.200
					AA	107:155	0		81:155	0	
3	--	--	rs978458	37.5	TT	64:85	0.016	-0.092, 0.125	52:85	0.058	-0.064, 0.181
					CT	187:258	-0.012	-0.091, 0.068	147:258	0.016	-0.073, 0.105
					CC	165:224	0		124:224	0	
3	3	2	rs5742692	11.7	AG, GG	89:122	-0.005	-0.094, 0.084	72:122	0.026	-0.073, 0.125
					AA	327:445	0		251:445	0	
--	--	--	rs11111262	2.7	AG, AA	25:29	0.135	0.004, 0.266	21:29	0.163	-0.005, 0.331
					GG	389:535	0		300:535	0	

3	3	3	rs1520220	34.7	GG	57:74	0.044	-0.067, 0.155	46:74	0.083	-0.044, 0.210
					CG	180:250	-0.007	-0.085, 0.072	143:250	0.033	-0.054, 0.121
					CC	178:245	0		132:245	0	
--	--	--	rs3730204	0.3	CT	1:3	--		1:3	--	
					TT	414:563	0		322:563	0	
Outside	3	4	rs6214	45.2	CC	93:125	-0.002	-0.101, 0.010	75:125	0.048	-0.065, 0.160
					CT	185:260	-0.034	-0.118, 0.050	147:260	0.013	-0.081, 0.107
					TT	135:181	0		100:181	0	
4	4	1	rs6219	9.1	CT, TT	75:96	0.062	-0.030, 0.154	58:96	0.045	-0.063, 0.153
					CC	341:474	0		265:474	0	
4	4	2	rs2946834	49.3	AA	111:143	0.042	-0.057, 0.140	88:143	0.058	-0.056, 0.171
					AG	197:279	-0.029	-0.118, 0.061	152:279	-0.013	-0.112, 0.086
					GG	108:147	0		82:147	0	

NOTE: MAF, minor allele frequency.

* Blocks consist of SNPs with MAF ≥ 5% in strong linkage disequilibrium (LD) (95% of pairwise SNP comparisons with one-sided 95% confidence intervals for the D prime statistic within 0.7-0.98). Group refers to SNP combination for diplotype estimation. Position refers to SNP order in each group for diplotype estimation; the position for each group begins with "1". Values for diplotype group and position (--) were not listed for SNPs excluded from diplotype estimation based on either: 1) Tagger algorithm ($r^2 \geq 0.8$); 2) MAF < 5% (excluded from blocks (--)). rs12821878 was combined with block 1 SNPs for diplotype estimation. rs10860869 and rs5009837 were combined with block 2 SNPs for diplotype estimation. rs9308315 was substituted for rs4764883 (Hardy-Weinberg p value <0.01). rs6214 was combined with block 3 SNPs for diplotype estimation.

† Based on total N = 582.

‡ Total N = 575, excluded 7 women missing uterine fibroid status.

§ PDs were not estimated for SNPs with 5 or fewer women with combined heterozygous and homozygous variant genotypes.

‖ SNPs selected *a priori*. rs3729846, synonymous SNP; rs17884626, nonsynonymous SNP.

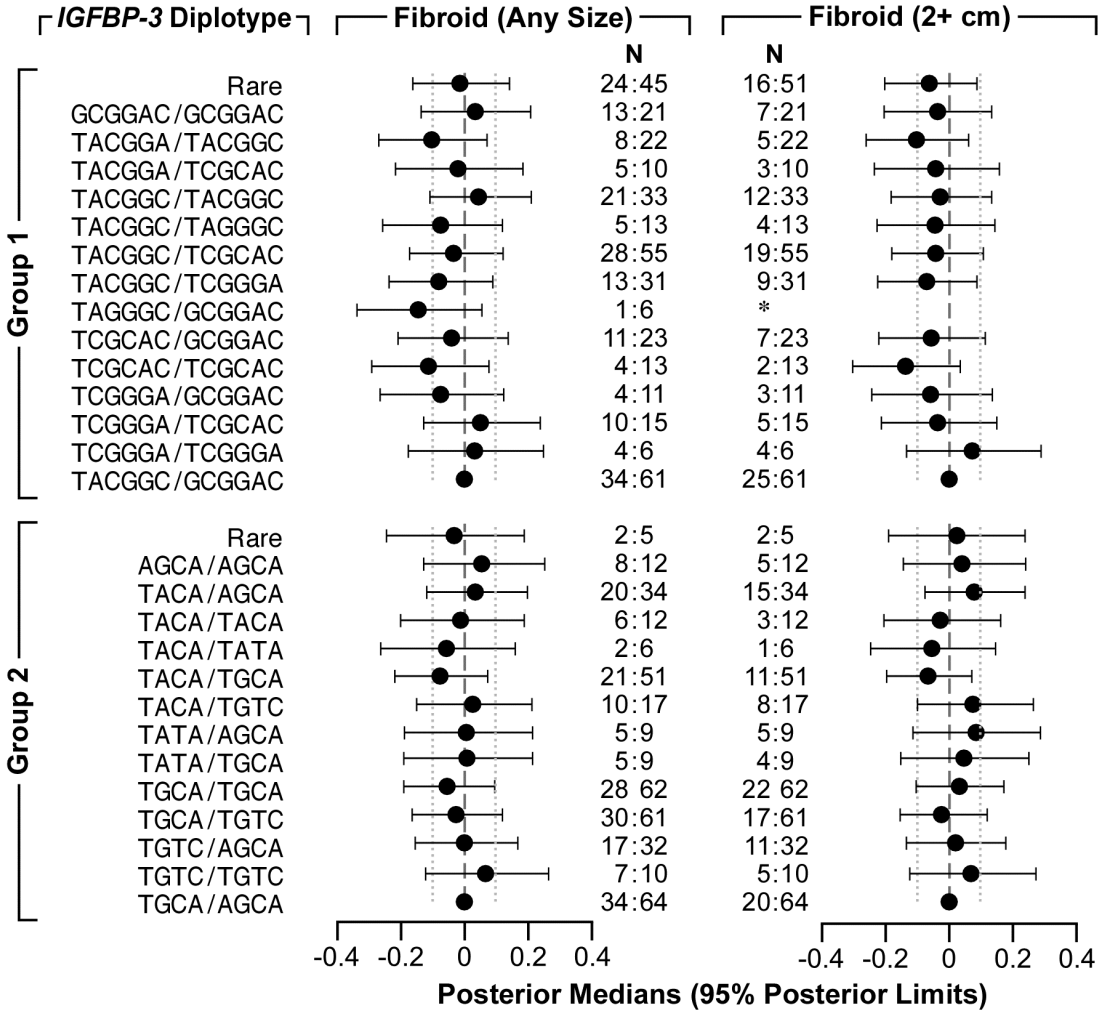

Figure 5.1. Unadjusted posterior medians of uterine fibroids with *IGFBP-3* diplotypes among Caucasians. Diplotype group-specific models for prevalence of fibroids (any size and those 2+ cm in diameter) with posterior medians and 95% posterior limits based on prior mean of 0 and prior variance of $(0.5/3.92)^2 \approx 0.016$. Groups were created by combining individual SNPs that were not included in a block with an adjacent block and combining adjacent blocks with each other, if doing so resulted in diplotypes estimated with at least 90% posterior probability for at least 90% of Caucasians. After excluding women missing fibroid status (Group 1: missing = 6; Group 2: missing = 8), Group 1 (N=365) represents SNPs from blocks 1 and 2; Group 2 (N=384) represents 2 SNPs outside block 3 combined with block 3 SNPs. Diplotype category with "*" for fibroid (2+ cm) was included with rare diplotypes. Last diplotype listed in each group is the most common (referent) diplotype. Gray dotted lines indicate ± 10% for posterior medians.

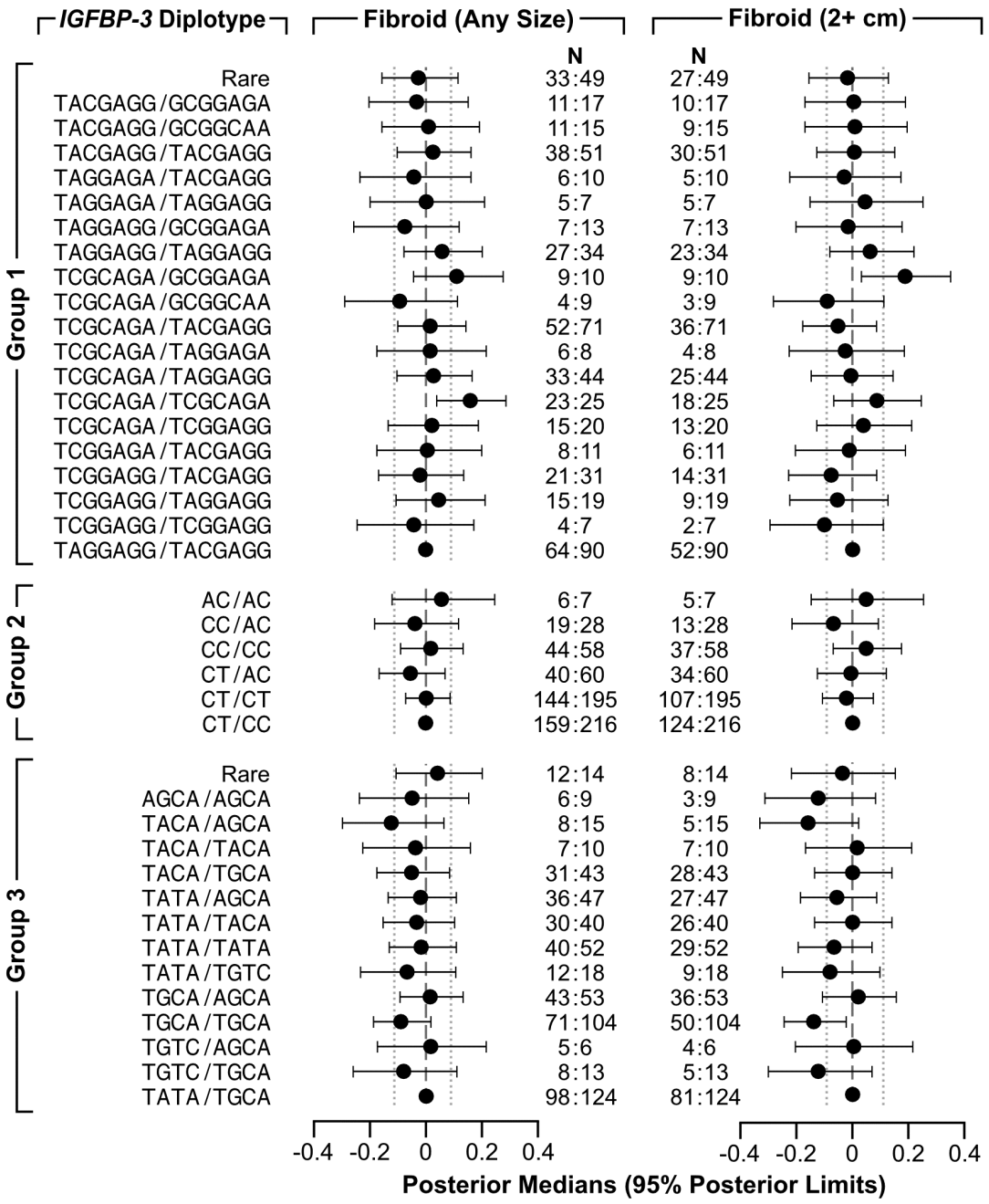

Figure 5.2. Unadjusted posterior medians of uterine fibroids with *IGFBP-3* diplotypes among African Americans. Diplotype group-specific models for prevalence of fibroids (any size and those 2+ cm in diameter) with posterior medians and 95% posterior limits based on prior mean of 0 and prior variance of $(0.5/3.92)^2 \approx 0.016$. Groups were created by combining individual SNPs that were not included in a block with an adjacent block and combining adjacent blocks with each other, if doing so resulted in diplotypes estimated with at least 90% posterior probability for at least 90% of African Americans. After excluding women missing fibroid status (Group 1: missing = 6; Group 2: missing = 7; Group 3: missing = 7), Group 1 (N=541) represents 3 SNPs outside block 1 combined with block 1 SNPs; Group 2 (N=564) represents block 2 SNPs; Group 3 (N=548) represents SNPs from blocks 3 and 4. Last diplotype listed in each group is the most common (referent) diplotype. Gray dotted lines indicate ± 10% for posterior medians.

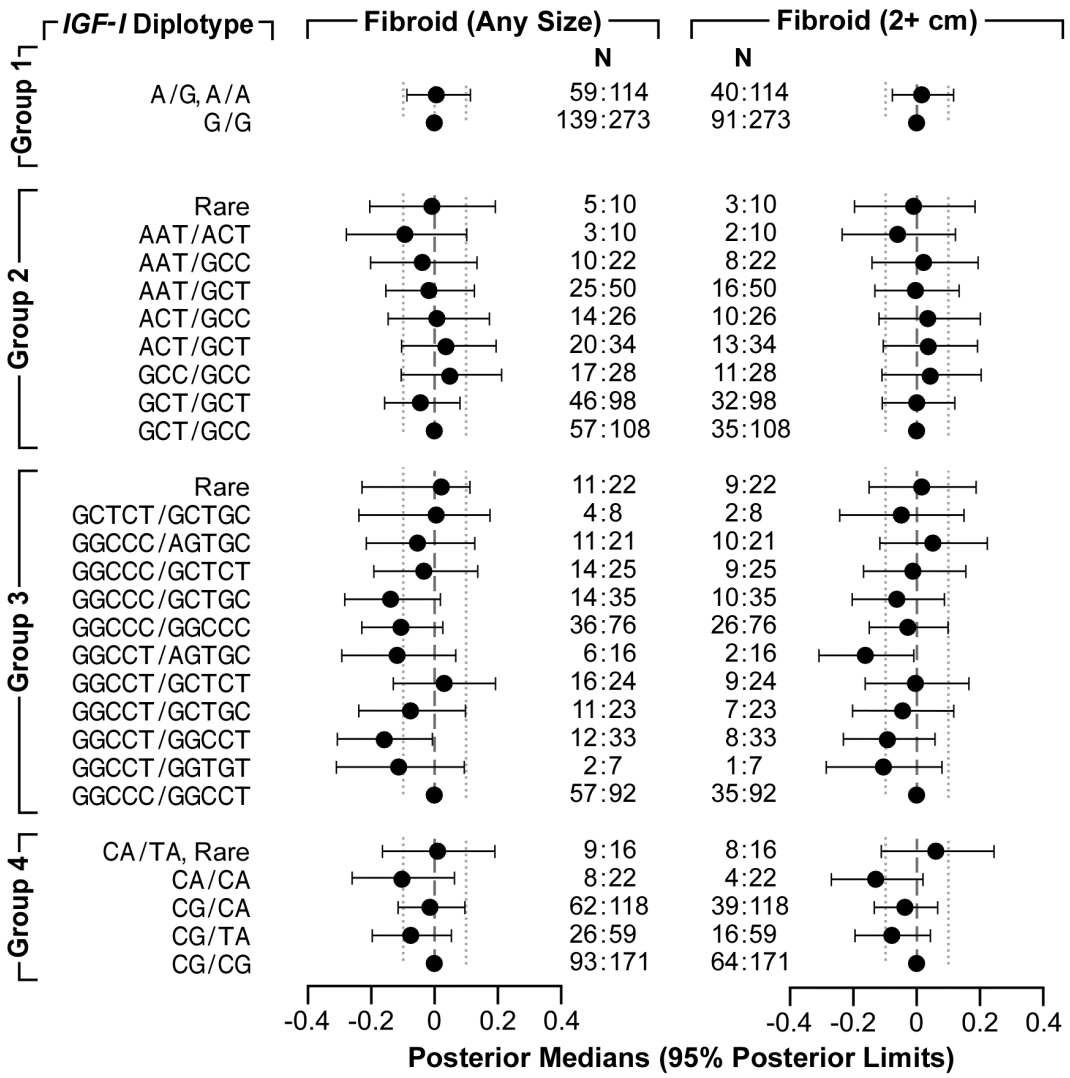

Figure 5.3.Unadjusted posterior medians of uterine fibroids with *IGF-I* diplotypes among Caucasians. Diplotype group-specific models for prevalence of fibroids (any size and those 2+ cm in diameter) with posterior medians and 95% posterior limits based on prior mean of 0 and prior variance of $(0.5/3.92)^2 \approx 0.016$. Groups were created by combining individual SNPs that were not included in a block with an adjacent block and combining adjacent blocks with each other, if doing so resulted in diplotypes estimated with at least 90% posterior probability for at least 90% of Caucasians. After excluding 8 women missing fibroid status, Group 1 (N=387) represents 1 SNP (rs35767) outside block 1 since it could not be combined with any adjacent SNPS; Group 2 (N=386) represents block 1 SNPs; Group 3 (N=382) represents block 2 SNPs combined with 1 SNP outside block 2; Group 4 (N=386) represents block 3 SNPs. Last diplotype listed in each group is the most common (referent) diplotype. Gray dotted lines indicate ± 10% for posterior medians.

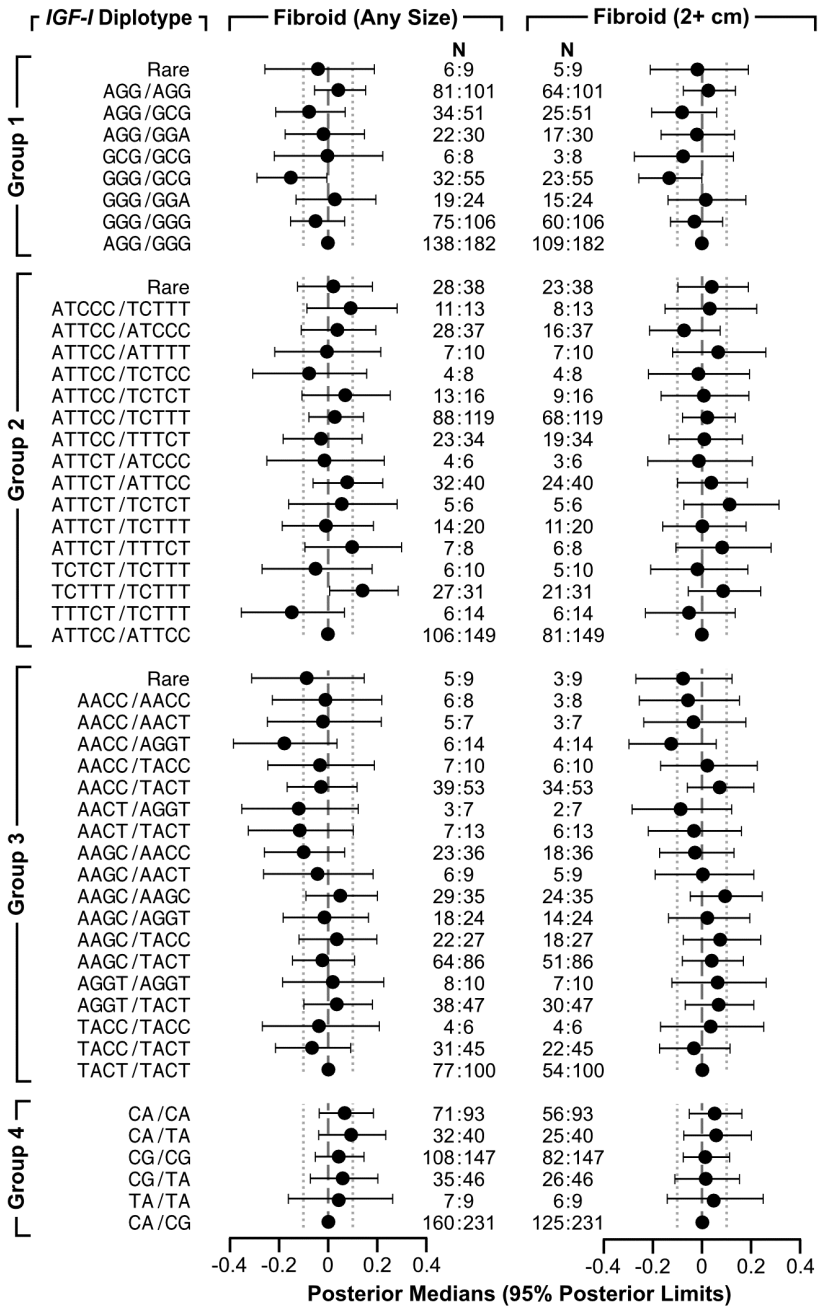

Figure 5.4. Unadjusted posterior medians of uterine fibroids with *IGF-I* diplotypes among African Americans. Diplotype group-specific models for prevalence of fibroids (any size and those 2+ cm in diameter) with posterior medians and 95% posterior limits based on prior mean of 0 and prior variance of $(0.5/3.92)2 \approx 0.016$. Groups were created by combining individual SNPs that were not included in a block with an adjacent block and combining adjacent blocks with each other, if doing so resulted in diplotypes estimated with at least 90% posterior probability for at least 90% of African Americans. After excluding 7 women missing fibroid status, Group 1 (N=566) represents block 1 SNPs combined with 1 SNP outside block 1; Group 2 (N=559) represents 2 SNPs outside block 2 combined with block 2 SNPs; Group 3 (N=546) represents block 3 SNPs combined with 1 SNP outside block 3; Group 4 (N=566) represents block 4 SNPs. Last diplotype listed in each group is the most common (referent) diplotype. Gray dotted lines indicate ± 10% for posterior medians.

CHAPTER VI

CONCLUSIONS

Introduction

Insulin-like growth factor-one (IGF-I), a peptide with structural similarities to insulin, is important for many cellular functions including cell cycle regulation, differentiation, and inhibition of apoptosis (1). IGF binding protein-3 (IGFBP-3) influences IGF-I bioavailability and degradation but may act independently from IGF-I such as inhibiting growth and promoting apoptosis. Associations with circulating IGF-I and IGFBP-3 have been evaluated with cardiovascular disease, diabetes, and cancer. Genetic factors are important determinants of circulating IGF-I (2, 3) and IGFBP-3 levels (2, 4).

The common theme with both studies is that we evaluated variation in *IGF-I* and *IGFBP-3* through single nucleotide polymorphisms and estimated diplotypes among a premenopausal population well-represented by both Caucasians and African Americans. Dense evaluation *IGF-I* and *IGFBP-3* SNPs in relation to their circulating levels among African Americans has been limited to the Multiethnic Cohort study, which included a small African American sample (N = 150) (5). Based on biological interrelations between IGF-I and IGFBP-3, we evaluated whether *IGF-I* and *IGFBP-3* variants were associated with circulating levels of both proteins in our first study, which had previously been done in only two studies (6, 7).

In our second study, we analyzed associations between *IGF-I* and *IGFBP-3* variants and prevalence of uterine fibroids. Uterine fibroids are hormonally dependent benign tumors, which are the primary indicator for hysterectomies in the United States, and occur disproportionately among African Americans. However, much of their etiology has still not been characterized. One hypothesis is that estrogens and progesterones increase cytokine and growth factor activity, which promotes fibroid development and growth. Tissue-specific evidence supports involvement of IGF-I in fibroid pathogenesis including *in vitro* promotion of fibroid growth (8) and increased IGF-I mRNA or peptides (9-15) (16) in fibroids relative to normal myometrium. In particular, one gene expression study measured elevated IGF-I mRNA and protein in addition to changes in two IGF-I pathway factors after estrogen treatment (16). Although tissue-specific evidence supporting IGFBP-3 involvement with fibroid etiology is limited to one study with elevated IGFBP-3 mRNA in fibroids (15), IGFBP-3 may contribute to IGF-I mediated mechanisms based on influence of IGFBP-3 on IGF-I bioavailability and degradation. However, no epidemiological study has previously investigated associations between variants of either gene and fibroids.

For both studies, the study population was 984 premenopausal African American and Caucasian participants with available DNA from the NIEHS Uterine Fibroid Study. The parent study conducted ultrasound screening of premenopausal women for fibroids to estimate their prevalence and investigate etiologic factors (17).

Summary of results

Among African Americans and Causasians, one of the most consistent findings was the association between *IGFBP-3* rs2854746 (Ala32Gly) and plasma IGFBP-3. In particular, this SNP results in an amino acid change from alanine to glycine, and protein sequencing tools suggest that it is located in the region responsible for IGF binding. We reported a dose response relation in which the CC genotype is positively associated with plasma IGFBP-3 relative to the GG genotype, with intermediate plasma IGFBP-3 levels associated with the CG genotype, which two other studies including the Multiethnic Cohort Study reported similar findings (5, 7). In addition, diplotypes for both racial groups support individual rs2854746 associations. We also reported high plasma IGFBP-3 levels in association with AA versus CC genotypes of an *IGFBP-3* promoter SNP (rs2854744, -202 A/C), but consistencies with diplotype findings were not as strong as with rs2854746, especially for African Americans. Seven studies also report similar associations between rs2854744 and plasma IGFBP-3 (6, 18-23). However, previous studies of this promoter SNP have been predominantly among Caucasians and have not evaluated rs2854746. We reported strong LD between rs2854744 and rs2854746 among Caucasians only ($r^2 = 0.82$), which may explain Caucasian associations between rs2854744 and plasma IGFBP-3. Strong LD between these two SNPs was not present among African Americans ($r^2 = 0.34$). We also estimated associations between *IGF-I* SNPs and plasma IGF-I although we did not find patterns between SNPs and diplotype associations with plasma IGF-I, and other studies evaluating *IGF-I* SNPs were inconsistent with our findings.

We found more precise associations for *IGF-I* and *IGFBP-3* SNPs with prevalence of fibroids among African Americans than Causasians, which had primarily imprecise associations due to 25 or fewer observations with homozygous variants. In particular, two *IGFBP-3* SNPs associated with fibroids among African Americans have potential functional relevance including rs9282734 (His158Pro), which results in an amino acid change from histidine to proline, and rs2475551, which is located at a splice site within less than 20 kilobases (kb) from exon 2. However, rs9282734 and rs2475551 have not been reported to be associated with diseases. Among a Caucasian study population, Canzian et al. reported no association for rs2475551 with breast cancer but an inverse association with plasma IGFBP-3 (6), which the latter is consistent with our study findings. Other fibroid associations among African Americans included SNPs outside *IGFBP-3* at the 5' (rs903889) and 3' ends (rs12671457) and a SNP outside *IGF-I* at the 5' end (rs35767), and there was an association with a SNP outside *IGF-I* at the 3' end (rs2946834) among Caucasians. For the two *IGF-I* SNPs noted above, associations with mammographic density (rs2946834) (24) and breast cancer (rs35767) (6) have been reported among predominantly Caucasian study populations. Generally, we estimated slightly stronger associations in a similar direction with prevalence of any fibroids rather than larger fibroids for *IGF-I* and *IGFBP-3* variants; however, differences in fibroid analyses were not consistent and substantial enough to conclude that variants preferentially affect fibroid incidence versus growth.

Strengths and limitations

The study population was well-represented by African Americans (about 60%), which was a major strength for both studies. In addition, we attempted to capture variation across both genes by selecting haplotype-tagging SNPs from race-specific reference populations using a pairwise r^2 value of 0.8. In particular, the first paper contributes information on distribution of SNPs as well as associations with circulating IGF-I and IGFBP-3 among African Americans, who have been evaluated with dense coverage of *IGF-I* or *IGFBP-3* polymorphisms by only one previous study (5). For the second study, the high proportion of African Americans is particularly relevant given that greater prevalence of fibroids and related morbidity has been consistently reported among this racial group (17, 25-28).

We are the first epidemiologic study to evaluate *IGF-I* and *IGFBP-3* variants in association with uterine fibroids. A major strength of our study population is that women were screened for uterine fibroids using ultrasounds, which reduces misclassification of fibroids since many women with fibroids in the parent study were asymptomatic. Our investigation of the hypothesis that IGF-I would preferentially affect fibroid growth versus incidence was limited because we did not have statistical power to evaluate each of the fibroid size categories in relation to women with no fibroids. Instead, we were only able to evaluate women with larger fibroids compared to women with small fibroids and no fibroids combined.

Our genotyping was very successful, with only five women having less than 50% of complete allele calls for the 45 SNPs in study. In addition, exclusions from uncertain estimated diplotypes (posterior probability < 90%) were negligible since 93 to 99% of

women were included for race-specific analyses of each diplotype group. Given the heterogeneity of diplotypes represented within race-specific diplotype groups, validity and precision of regression estimates was a potential limitation with analyses. However, we utilized an empirical-Bayes method (information-weighted averaging) (29), which shrunk imprecise regression estimates based on smaller numbers of observations further toward the prior mean of 0 based than more precise estimates. Overall mean square error is reduced with this method based on a greater reduction in precision than the gain in bias with estimates (30).

Possible selection bias is a limitation for both studies if excluded women differed from the remaining study population with respect to their genotypes or study outcomes (i.e. circulating protein levels or fibroid status). To evaluate selection bias, we would need to consider not only women who were excluded from both studies because there was no DNA available, but also women who were dropped earlier from the study population for other reasons including postmenopausal at enrollment and refusal to participate. We restricted study population to only African Americans and Caucasians and stratified all analyses by race for both studies, but population stratification could still be a bias within racial groups of our study population (31-34). We had limited power to evaluate rare SNPs with both studies. Rare estimated diplotypes (5 or fewer observations) were combined into one category for all analyses, but heterogeneity of diplotypes within these categories made interpretation of their estimated associations impossible. Finally, we did not have statistical power to explore gene-environmental or gene-gene interactions, which may be relevant to our estimated associations, especially with differences in associations across racial groups.

Future research

Findings from the first study reported associations between several *IGF-I* and *IGFBP-3* SNPs and their circulating protein levels. In particular, we had a highly consistent finding with evidence suggesting a causal association between *IGFBP-3* rs2854746 (Ala32Gly) and plasma IGFBP-3 among both Caucasians and African Americans. Validation of this association with rs2854746 and with other SNPs in our study would be needed in other study populations. Further study is needed to assess functional significance of the SNPs with reported associations in either of our studies. For SNPs that have no known functional relevance, some of their associations with circulating protein levels or fibroids may be based on spurious findings, but instead they may tag a functional variant or affect expression. Investigation of SNPs having associations with circulating protein levels with risk of cancer and cardiovascular disease would further our understanding in whether IGF-I mediated mechanisms and IGFBP-3 independent actions are involved in their etiology, since issue of confounding would be lessened for associations with gene variants, unlike with circulating levels of their proteins.

Given that this is the first epidemiologic study to evaluate *IGF-I* and *IGFBP-3* in relation to fibroids, estimated associations cannot be considered conclusive without replication in other study populations. Findings from the second study estimated precise associations with fibroids predominantly among African Americans. Therefore, it would be important to replicate our findings within a study population with a substantial proportion of African Americans. Other issues that should be considered in future studies are gene-gene interactions and gene-environment since IGF-I and IGFBP-3 have strong

biological interrelations, and their tissue-specific and circulating levels may be affected by environmental factors. Finally, other genes involved in the IGF-I pathway should be considered to further investigate whether IGF-I is involved with fibroid pathogenesis.

REFERENCES

1. Jones, J. I., and Clemmons, D. R. Insulin-like growth factors and their binding proteins: biological actions. Endocr Rev, *16:* 3-34, 1995.

2. Harrela, M., Koistinen, H., Kaprio, J., Lehtovirta, M., Tuomilehto, J., Eriksson, J., Toivanen, L., Koskenvuo, M., Leinonen, P., Koistinen, R., and Seppala, M. Genetic and environmental components of interindividual variation in circulating levels of IGF-I, IGF-II, IGFBP-1, and IGFBP-3. J Clin Invest, *98:* 2612-5, 1996.

3. Pantsulaia, I., Trofimov, S., Kobyliansky, E., and Livshits, G. Genetic regulation of the variation of circulating insulin-like growth factors and leptin in human pedigrees. Metabolism, *54:* 975-81, 2005.

4. Pantsulaia, I., Trofimov, S., Kobyliansky, E., and Livshits, G. Pedigree-based quantitative genetic analysis of interindividual variation in circulating levels of IGFBP-3. J Bone Miner Metab, *20:* 156-63, 2002.

5. Cheng, I., Henderson, K. D., Haiman, C. A., Kolonel, L. N., Henderson, B. E., Freedman, M. L., and Le Marchand, L. Genetic Determinants of Circulating IGF-I, IGFBP-1 and IGFBP-3 Levels in a Multiethnic Population. J Clin Endocrinol Metab, 2007.

6. Canzian, F., McKay, J. D., Cleveland, R. J., Dossus, L., Biessy, C., Rinaldi, S., Landi, S., Boillot, C., Monnier, S., Chajes, V., Clavel-Chapelon, F., Tehard, B., Chang-Claude, J., Linseisen, J., Lahmann, P. H., Pischon, T., Trichopoulos, D., Trichopoulou, A., Zilis, D., Palli, D., Tumino, R., Vineis, P., Berrino, F., Bueno-de-Mesquita, H. B., van Gils, C. H., Peeters, P. H., Pera, G., Ardanaz, E., Chirlaque, M. D., Quiros, J. R., Larranaga, N., Martinez-Garcia, C., Allen, N. E., Key, T. J., Bingham, S. A., Khaw, K. T., Slimani, N., Norat, T., Riboli, E., and Kaaks, R. Polymorphisms of genes coding for insulin-like growth factor 1 and its major binding proteins, circulating levels of IGF-I and IGFBP-3 and breast cancer risk: results from the EPIC study. Br J Cancer, *94:* 299-307, 2006.

7. Morimoto, L. M., Newcomb, P. A., White, E., Bigler, J., and Potter, J. D. Variation in plasma insulin-like growth factor-1 and insulin-like growth factor binding protein-3: genetic factors. Cancer Epidemiol Biomarkers Prev, *14:* 1394-401, 2005.

8. Strawn, E. Y., Jr., Novy, M. J., Burry, K. A., and Bethea, C. L. Insulin-like growth factor I promotes leiomyoma cell growth in vitro. Am J Obstet Gynecol, *172:* 1837-43; discussion 1843-4, 1995.

9. Wolanska, M., and Bankowski, E. An accumulation of insulin-like growth factor I (IGF-I) in human myometrium and uterine leiomyomas in various stages of tumour growth. Eur Cytokine Netw, *15:* 359-63, 2004.

10. Van der Ven, L. T., Roholl, P. J., Gloudemans, T., Van Buul-Offers, S. C., Welters, M. J., Bladergroen, B. A., Faber, J. A., Sussenbach, J. S., and Den Otter, W. Expression of insulin-like growth factors (IGFs), their receptors and IGF binding protein-3 in normal, benign and malignant smooth muscle tissues. Br J Cancer, *75:* 1631-40, 1997.

11. Van der Ven, L. T., Gloudemans, T., Roholl, P. J., Van Buul-Offers, S. C., Bladergroen, B. A., Welters, M. J., Sussenbach, J. S., and den Otter, W. Growth advantage of human leiomyoma cells compared to normal smooth-muscle cells due to enhanced sensitivity toward insulin-like growth factor I. Int J Cancer, *59:* 427-34, 1994.

12. Dixon, D., He, H., and Haseman, J. K. Immunohistochemical localization of growth factors and their receptors in uterine leiomyomas and matched myometrium. Environ Health Perspect, *108 Suppl 5:* 795-802, 2000.

13. Englund, K., Lindblom, B., Carlstrom, K., Gustavsson, I., Sjoblom, P., and Blanck, A. Gene expression and tissue concentrations of IGF-I in human myometrium and fibroids under different hormonal conditions. Mol Hum Reprod, *6:* 915-20, 2000.

14. Hoppener, J. W., Mosselman, S., Roholl, P. J., Lambrechts, C., Slebos, R. J., de Pagter-Holthuizen, P., Lips, C. J., Jansz, H. S., and Sussenbach, J. S. Expression of insulin-like growth factor-I and -II genes in human smooth muscle tumours. Embo J, *7:* 1379-85, 1988.

15. Giudice, L. C., Irwin, J. C., Dsupin, B. A., Pannier, E. M., Jin, I. H., Vu, T. H., and Hoffman, A. R. Insulin-like growth factor (IGF), IGF binding protein (IGFBP), and IGF receptor gene expression and IGFBP synthesis in human uterine leiomyomata. Hum Reprod, *8:* 1796-806, 1993.

16. Swartz, C. D., Afshari, C. A., Yu, L., Hall, K. E., and Dixon, D. Estrogen-induced changes in IGF-I, Myb family and MAP kinase pathway genes in human uterine leiomyoma and normal uterine smooth muscle cell lines. Mol Hum Reprod, *11:* 441-50, 2005.

17. Baird, D. D., Dunson, D. B., Hill, M. C., Cousins, D., and Schectman, J. M. High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence. Am J Obstet Gynecol, *188:* 100-7, 2003.

18. Al-Zahrani, A., Sandhu, M. S., Luben, R. N., Thompson, D., Baynes, C., Pooley, K. A., Luccarini, C., Munday, H., Perkins, B., Smith, P., Pharoah, P. D., Wareham, N. J., Easton, D. F., Ponder, B. A., and Dunning, A. M. IGF1 and IGFBP3 tagging polymorphisms are associated with circulating levels of IGF1, IGFBP3 and risk of breast cancer. Hum Mol Genet, *15:* 1-10, 2006.

19. Jernstrom, H., Deal, C., Wilkin, F., Chu, W., Tao, Y., Majeed, N., Hudson, T., Narod, S. A., and Pollak, M. Genetic and nongenetic factors associated with variation of plasma levels of insulin-like growth factor-I and insulin-like growth factor-binding protein-3 in healthy premenopausal women. Cancer Epidemiol Biomarkers Prev, *10:* 377-84, 2001.

20. Lai, J. H., Vesprini, D., Zhang, W., Yaffe, M. J., Pollak, M., and Narod, S. A. A polymorphic locus in the promoter region of the IGFBP3 gene is related to mammographic breast density. Cancer Epidemiol Biomarkers Prev, *13:* 573-82, 2004.

21. Schernhammer, E. S., Hankinson, S. E., Hunter, D. J., Blouin, M. J., and Pollak, M. N. Polymorphic variation at the -202 locus in IGFBP3: Influence on serum levels of insulin-like growth factors, interaction with plasma retinol and vitamin D and breast cancer risk. Int J Cancer, *107:* 60-4, 2003.

22. Ren, Z., Cai, Q., Shu, X. O., Cai, H., Li, C., Yu, H., Gao, Y. T., and Zheng, W. Genetic polymorphisms in the IGFBP3 gene: association with breast cancer risk and blood IGFBP-3 protein levels among Chinese women. Cancer Epidemiol Biomarkers Prev, *13:* 1290-5, 2004.

23. Deal, C., Ma, J., Wilkin, F., Paquette, J., Rozen, F., Ge, B., Hudson, T., Stampfer, M., and Pollak, M. Novel promoter polymorphism in insulin-like growth factor-binding protein-3: correlation with serum levels and interaction with known regulators. J Clin Endocrinol Metab, *86:* 1274-80, 2001.

24. Tamimi, R. M., Cox, D. G., Kraft, P., Pollak, M. N., Haiman, C. A., Cheng, I., Freedman, M. L., Hankinson, S. E., Hunter, D. J., and Colditz, G. A. Common genetic variation in IGF1, IGFBP-1, and IGFBP-3 in relation to mammographic density: a cross-sectional study. Breast Cancer Res, *9:* R18, 2007.

25. Marshall, L. M., Spiegelman, D., Barbieri, R. L., Goldman, M. B., Manson, J. E., Colditz, G. A., Willett, W. C., and Hunter, D. J. Variation in the incidence of uterine leiomyoma among premenopausal women by age and race. Obstet Gynecol, *90:* 967-73, 1997.

26. Samadi, A. R., Lee, N. C., Flanders, W. D., Boring, J. R., 3rd, and Parris, E. B. Risk factors for self-reported uterine fibroids: a case-control study. Am J Public Health, *86:* 858-62, 1996.

27. Faerstein, E., Szklo, M., and Rosenshein, N. Risk factors for uterine leiomyoma: a practice-based case-control study. I. African-American heritage, reproductive history, body size, and smoking. Am J Epidemiol, *153:* 1-10, 2001.

28. Wise, L. A., Palmer, J. R., Stewart, E. A., and Rosenberg, L. Age-specific incidence rates for self-reported uterine leiomyomata in the Black Women's Health Study. Obstet Gynecol, *105:* 563-8, 2005.

29. Greenland, S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. Int J Epidemiol, *35:* 765-75, 2006.

30. Greenland, S., and Poole, C. Empirical-Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. Arch Environ Health, *49:* 9-16, 1994.

31. Wang, Y., Localio, R., and Rebbeck, T. R. Evaluating bias due to population stratification in case-control association studies of admixed populations. Genet Epidemiol, *27:* 14-20, 2004.

32. Wang, Y., Localio, R., and Rebbeck, T. R. Evaluating bias due to population stratification in epidemiologic studies of gene-gene or gene-environment interactions. Cancer Epidemiol Biomarkers Prev, *15:* 124-32, 2006.

33. Ardlie, K. G., Lunetta, K. L., and Seielstad, M. Testing for population subdivision and association in four case-control studies. Am J Hum Genet, *71:* 304-11, 2002.

34. Wacholder, S., Rothman, N., and Caporaso, N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst, *92:* 1151-8, 2000.

APPENDIX

Table A.1. Pairwise r^2 between *IGFBP-3* single nucleotide polymorphisms (SNPs) among Caucasians*

SNP rs #	903889	924140	2854744	2854746	2471551	3110697	2453840	2453839	6670	13223993	2270628	12671457
903889												
924140	0.24											
2854744	0.23	**0.99**										
2854746	0.19	**0.81**	**0.82**									
2471551	0.06	0.16	0.16	0.12								
3110697	0.39	0.55	0.55	0.44	0.32							
2453840	0.04	0.00	0.00	0.00	0.05	0.12						
2453839	0.04	0.00	0.00	0.00	0.04	0.11	**0..87**					
6670	0.04	0.01	0.01	0.01	0.08	0.00	0.01	0.00				
13223993	0.00	0.05	0.05	0.02	0.03	0.00	0.04	0.04	0.07			
2270628	0.11	0.03	0.03	0.08	0.00	0.06	0.03	0.04	0.07	0.00		
12671457	0.12	0.06	0.06	0.06	0.00	0.09	0.02	0.02	0.05	0.05	0.77	

* Based on N = 401, excludes 1 woman missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation. Excludes SNPs with minor allele frequencies < 1%. Pairwise r^2 values ≥ 0.80, which are in bold, indicate SNP pairs in strong linkage disequilibrium.

Table A.2. Pairwise r^2 between *IGFBP-3* single nucleotide polymorphisms (SNPs) among African Americans*

SNP rs #	903889	924140	2854744	2854746	2471551	9282734	6953668	3110697	2453840	2453839
903889										
924140	0.16									
2854744	0.14	**0.82**								
2854746	0.04	0.30	0.34							
2471551	0.02	0.38	0.33	0.10						
9282734	0.33	0.05	0.04	0.01	0.00					
6953668	0.28	0.08	0.07	0.02	0.01	0.64				
3110697	0.17	0.65	0.57	0.25	0.45	0.05	0.08			
2453840	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01		
2453839	0.04	0.02	0.03	0.09	0.00	0.02	0.02	0.04	0.14	
6670	0.00	0.01	0.04	0.01	0.05	0.00	0.00	0.02	0.00	0.03
13223993	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00
2270628	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00
12671457	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

* Based on N = 579, excludes 3 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation. Excludes SNPs with minor allele frequencies < 1%. Pairwise r^2 values ≥ 0.80, which are in bold, indicate SNP pairs in strong linkage disequilibrium.

SNP rs #	6670	13223993	2270628	12671457
903889				
924140				
2854744				
2854746				
2471551				
9282734				
6953668				
3110697				
2453840				
2453839				
6670				
13223993	0.09			
2270628	0.07	0.45		
12671457	0.00	0.00	0.09	

* Based on N = 579, excludes 3 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation. Excludes SNPs with minor allele frequencies < 1%. Pairwise r^2 values ≥ 0.80, which are in bold, indicate SNP pairs in strong linkage disequilibrium.

Table A.3. Pairwise r^2 between *IGF-I* single nucleotide polymorphisms (SNPs) among Caucasians*

SNP rs #	35767	5742612	5742614	12821878	10860869	1019731	7956547	2033178	5742657	11829586	4764884
35767											
5742612	0.21										
5742614	0.00	0.00									
12821878	0.04	0.01	0.00								
10860869	0.00	0.09	0.03	0.12							
1019731	0.02	0.00	0.00	0.47	0.05						
7956547	0.01	0.00	0.03	0.11	**0.91**	0.05					
2033178	0.35	0.00	0.00	0.01	0.02	0.00	0.02				
5742657	0.07	0.33	0.00	0.00	0.02	0.00	0.00	0.00			
11829586	0.07	0.33	0.00	0.00	0.02	0.00	0.00	0.00	**0.88**		
4764884	0.03	0.01	0.00	0.10	0.32	0.04	0.38	0.19	0.00	0.00	
5009837	0.10	0.04	0.00	0.12	0.45	0.06	0.39	0.15	0.05	0.04	0.79
17727841	0.04	0.00	0.00	0.06	0.55	0.03	0.60	0.01	0.00	0.00	0.67
4764883	0.09	0.04	0.00	0.13	0.45	0.06	0.39	0.15	0.05	0.04	0.78
9308315	0.06	0.00	0.00	0.11	0.38	0.05	0.32	0.17	0.04	0.04	**0.89**
978458	0.06	0.00	0.00	0.10	0.38	0.05	0.32	0.17	0.04	0.04	**0.88**
5742692	0.08	0.35	0.00	0.00	0.02	0.00	0.00	0.00	**0.94**	**0.94**	0.00
11111262	0.02	0.00	0.00	0.03	0.25	0.01	0.28	0.00	0.00	0.00	0.32
1520220	0.13	0.01	0.00	0.06	0.11	0.03	0.07	0.28	0.05	0.05	0.49
3730204	0.00	0.00	0.00	0.00	0.05	0.00	0.06	0.00	0.00	0.00	0.06
6214	0.01	0.03	0.01	0.10	0.00	0.03	0.00	0.04	0.01	0.01	0.00
6219	0.02	0.00	0.00	0.03	0.26	0.01	0.28	0.00	0.00	0.00	0.32
2946834	0.06	0.02	0.00	0.11	0.04	0.06	0.03	0.13	0.03	0.02	0.20

* Based on N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation. Excludes SNPs with minor allele frequencies < 1%. Pairwise r^2 values ≥ 0.80, which are in bold, indicate SNP pairs in strong linkage disequilibrium.

SNP rs #	5009837	17727841	4764883	9308315	978458	5742692	11111262	1520220	3730204	6214	6219	2946834
35767												
5742612												
5742614												
12821878												
10860869												
1019731												
7956547												
2033178												
5742657												
11829586												
4764884												
5009837												
17727841	0.53											
4764883	**0.99**	0.53										
9308315	**0.88**	0.60	**0.88**									
978458	**0.88**	0.59	**0.87**	**0.98**								
5742692	0.05	0.00	0.05	0.05	0.05							
11111262	0.25	0.47	0.25	0.28	0.28	0.00						
1520220	0.54	0.17	0.53	0.59	0.60	0.06	0.48					
3730204	0.05	0.10	0.05	0.06	0.06	0.00	0.00	0.00				
6214	0.00	0.00	0.00	0.00	0.00	0.01	0.07	0.08	0.03			
6219	0.25	0.46	0.25	0.28	0.28	0.00	**0.97**	0.48	0.00	0.07		
2946834	0.31	0.05	0.31	0.25	0.26	0.02	0.22	0.47	0.01	0.02	0.23	

* Based on N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation. Excludes SNPs with minor allele frequencies < 1%. Pairwise r^2 values ≥ 0.80, which are in bold, indicate SNP pairs in strong linkage disequilibrium.

Table A.4. Pairwise r^2 between *IGF-I* single nucleotide polymorphisms (SNPs) among African Americans*

SNP rs #	35767	5742612	5742614	12821878	10860869	1019731	7956547	5742626	17880975	2033178	5742657
35767											
5742612	0.05										
5742614	0.09	0.00									
12821878	0.04	0.00	0.00								
10860869	0.07	0.07	0.02	0.02							
1019731	0.02	0.00	0.00	0.56	0.01						
7956547	0.04	0.06	0.03	0.02	0.75	0.01					
5742626	0.08	0.00	0.01	0.00	0.03	0.00	0.02				
17880975	0.02	0.00	0.00	0.00	0.01	0.00	0.00	0.00			
2033178	0.00	0.00	0.00	0.00	0.02	0.00	0.03	0.00	0.00		
5742657	0.02	0.00	0.00	0.00	0.00	0.00	0.04	0.01	0.00	0.00	
5742663	0.01	0.00	0.00	0.00	0.00	0.00	0.03	0.01	0.00	0.00	**0.88**
11829586	0.01	0.00	0.00	0.00	0.01	0.00	0.03	0.01	0.00	0.00	**0.88**
4764884	0.05	0.00	0.03	0.01	0.48	0.01	0.65	0.01	0.00	0.15	0.04
5742683	0.01	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	**0.86**
17884626	0.01	0.29	0.00	0.00	0.02	0.00	0.02	0.00	0.00	0.00	0.00
5009837	0.00	0.03	0.04	0.03	0.57	0.01	0.42	0.04	0.01	0.06	0.17
17727841	0.03	0.00	0.00	0.00	0.10	0.00	0.13	0.00	0.00	0.00	0.00
4764883	0.01	0.02	0.06	0.06	0.34	0.03	0.25	0.04	0.02	0.04	0.11
9308315	0.02	0.01	0.08	0.05	0.25	0.03	0.17	0.03	0.01	0.04	0.12
978458	0.00	0.00	0.03	0.02	0.45	0.01	0.32	0.03	0.01	0.07	0.21
5742692	0.02	0.00	0.00	0.00	0.00	0.00	0.03	0.01	0.00	0.00	**0.87**
11111262	0.01	0.00	0.00	0.00	0.05	0.00	0.07	0.00	0.00	0.00	0.00
1520220	0.00	0.00	0.03	0.03	0.36	0.01	0.26	0.03	0.01	0.08	0.21
6214	0.13	0.01	0.04	0.02	0.08	0.01	0.14	0.02	0.01	0.06	0.09
6219	0.04	0.00	0.01	0.00	0.18	0.00	0.24	0.00	0.00	0.01	0.00
2946834	0.00	0.01	0.02	0.05	0.24	0.03	0.16	0.02	0.02	0.03	0.11

* Based on N = 580, excludes 2 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation. Excludes SNPs with minor allele frequencies < 1%. Pairwise r^2 values ≥ 0.80, which are in bold, indicate SNP pairs in strong linkage disequilibrium.

SNP rs #	5742663	11829586	4764884	5742683	17884626	5009837	17727841	4764883	9308315	978458
35767										
5742612										
5742614										
12821878										
10860869										
1019731										
7956547										
5742626										
17880975										
2033178										
5742657										
5742663										
11829586	**0.89**									
4764884	0.04	0.04								
5742683	**0.88**	**0.81**	0.03							
17884626	0.00	0.00	0.00	0.00						
5009837	0.17	0.17	0.43	0.16	0.01					
17727841	0.00	0.00	0.15	0.00	0.00	0.06				
4764883	0.11	0.11	0.26	0.10	0.00	0.60	0.04			
9308315	0.13	0.13	0.31	0.11	0.01	0.46	0.04	**0.84**		
978458	0.21	0.21	0.52	0.20	0.00	**0.83**	0.08	0.50	0.59	
5742692	**0.89**	**0.89**	0.04	**0.81**	0.00	0.16	0.00	0.10	0.12	0.21
11111262	0.00	0.00	0.08	0.00	0.00	0.03	0.56	0.02	0.02	0.04
1520220	0.21	0.21	0.44	0.19	0.00	0.71	0.00	0.44	0.52	**0.87**
6214	0.09	0.09	0.27	0.10	0.00	0.05	0.00	0.18	0.25	0.08
6219	0.00	0.00	0.30	0.01	0.00	0.14	0.11	0.08	0.09	0.16
2946834	0.11	0.11	0.18	0.10	0.01	0.46	0.00	0.23	0.17	0.39

* Based on N = 580, excludes 2 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation. Excludes SNPs with minor allele frequencies < 1%. Pairwise r^2 values ≥ 0.80, which are in bold, indicate SNP pairs in strong linkage disequilibrium.

SNP rs #	5742692	11111262	1520220	6214	6219	2946834
35767						
5742612						
5742614						
12821878						
10860869						
1019731						
7956547						
5742626						
17880975						
2033178						
5742657						
5742663						
11829586						
4764884						
5742683						
17884626						
5009837						
17727841						
4764883						
9308315						
978458						
5742692						
11111262	0.00					
1520220	0.24	0.05				
6214	0.09	0.03	0.11			
6219	0.00	0.24	0.18	0.10		
2946834	0.12	0.02	0.48	0.00	0.10	

* Based on N = 580, excludes 2 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation. Excludes SNPs with minor allele frequencies < 1%. Pairwise r^2 values ≥ 0.80, which are in bold, indicate SNP pairs in strong linkage disequilibrium.

Table A.5. Differences of mean differences comparing African American and Caucasian mean plasma IGFBP-3 differences associated with index genotypes for *IGFBP-3* single nucleotide polymorphisms (SNPs)*

SNP Reference ID	Genotypes	Difference of Mean Differences	95% CI
rs903889†	GT, GG	-17	-267, 233
	TT	0	
rs924140‡	TT	-21	-328, 286
	CT	-91	-362, 181
	CC	0	
rs2854744‡	AA	-58	-360, 244
	AC	-78	-340, 184
	CC	0	
rs2854746	CC	266	-70, 602
	CG	79	-144, 303
	GG	0	
rs2471551	CC	-120	-643, 404
	CG	-139	-375, 97
	GG	0	
rs3110697	AA	-126	-452, 201
	AG	-177	-410, 56
	GG	0	
rs2453840†	AC, AA	298	45, 551
	CC	0	
rs2453839	CC	157	-340, 654
	CT	86	-149, 320
	TT	0	
rs6670†	AT, AA	-123	-358, 112
	TT	0	
rs13223993	AA	-56	-494, 383
	AG	-190	-428, 49
	GG	0	
rs2270628	TT	400	-88, 888
	CT	-49	-281, 183
	CC	0	
rs12671457	AC, CC	-99	-397, 198
	AA	0	

* Only compares SNPs with at least 2% for minor allele frequencies among both racial groups. Combined data from both racial groups and fit linear regression models that included multiplicative gene by race interaction terms along with separate parameters for race and genotypes. Referent genotypes were homozygous wildtype genotypes among Caucasians.

† These SNPs had different models of inheritance (i.e. codominant, dominant) for each of the race-specific linear regression models. Dominant model was used for these SNPs in linear regression models with combined racial groups to estimate differences of mean differences.

‡ These SNPs had opposite homozygous wildtype genotypes for race-specific linear regression models among African Americans.

Table A.6. Differences of mean differences comparing African American and Caucasian mean plasma IGF-I differences associated with index genotypes for *IGF-I* single nucleotide polymorphisms (SNPs)*

SNP Reference ID	Genotypes	Difference of Mean Differences	95% CI
rs35767†	AG, AA	-7	-26, 11
	GG	0	
rs5742612	AG, GG	-15	-50, 19
	AA	0	
rs5742614	CG, CC	-9	-48, 30
	GG	0	
rs12821878†	AG, AA	-19	-41, 4
	GG	0	
rs1086869	TT	16	-16, 48
	AT	6	-12, 24
	AA	0	
rs1019731	AC, AA	-20	-47, 8
	CC	0	
rs7956547	CC	13	-21, 47
	CT	8	-10, 27
	TT	0	
rs2033178	AG, GG	-29	-58, 0
	AA	0	
rs5742657	AG, GG	-14	-50, 21
	AA	0	
rs11829586	AG, AA	-9	-44, 26
	GG	0	
rs4764884	TT	26	-11, 63
	CT	5	-13, 24
	CC	0	
rs5009837	TT	14	-15, 43
	CT	4	-14, 23
	CC	0	
rs17727841†	CG, CC	27	2, 51
	GG	0	
rs4764883‡	CC	20	-8, 49
	CT	-3	-23, 17
	TT	0	
rs9308315‡	AA	28	-3, 59
	AT	-2	-22, 17
	TT	0	
rs978458	TT	22	-10, 54
	CT	3	-16, 21
	CC	0	
rs5742692†	AG, GG	-13	-49, 23
	AA	0	
rs11111262	AG, AA	44	13, 75
	GG	0	

rs1520220	GG	43	1, 85
	CG	-1	-20, 18
	CC	0	
rs6214‡	TT	-14	-39, 11
	CT	1	-19, 22
	CC	0	
rs6219	CT, TT	21	-1, 44
	CC	0	
rs2946834	AA	24	-4, 52
	AG	1	-19, 21
	GG	0	

* Only compares SNPs with at least 2% for minor allele frequencies among both racial groups. Combined data from both racial groups and fit linear regression models that included multiplicative gene by race interaction terms along with separate parameters for race and genotypes. Referent genotypes were homozygous wildtype genotypes among Caucasians.

† These SNPs had different models of inheritance (i.e. codominant, dominant) for each of the race-specific linear regression models. Dominant model was used for these SNPs in linear regression models with combined racial groups to estimate differences of mean differences.

‡ These SNPs had opposite homozygous wildtype genotypes for race-specific linear regression models among African Americans.

Table A.7. Unadjusted linear regression of *IGFBP-3* group-specific diplotypes on plasma IGFBP-3 levels among Caucasians*

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		4661	4471, 4852		
1	Rare	40	-499	-804, -194	-438	-724, -152
1	GCGGAC/GCGGAC	21	-508	-889, -128	-418	-763, -73
1	TACGGA/TACGGC	23	299	-69, 666	249	-87, 584
1	TACGGA/TCGCAC	10	141	-373, 655	101	-334, 536
1	TACGGC/TACGGC	34	209	-112, 530	181	-118, 480
1	TACGGC/TAGGGC	13	-78	-538, 381	-60	-460, 341
1	TACGGC/TCGCAC	55	-29	-307, 250	-26	-290, 238
1	TACGGC/TCGGGA	31	-37	-368, 294	-32	-339, 275
1	TAGGGC/GCGGAC	6	-581	-1225, 64	-358	-864, 148
1	TAGGGC/TCGCAC	6	-260	-905, 384	-161	-667, 346
1	TCGCAC/GCGGAC	22	-245	-618, 129	-202	-542, 138
1	TCGCAC/TCGCAC	14	-304	-750, 142	-235	-626, 157
1	TCGGGA/GCGGAC	11	-386	-879, 107	-283	-705, 139
1	TCGGGA/TCGCAC	15	-681	-1114, -247	-531	-914, -148
1	TCGGGA/TCGGGA	6	-249	-893, 396	-153	-660, 353
1	TACGGC/GCGGAC	63	REF		REF	

* N = 401, excludes 1 woman missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.
† N = 370, excludes 30 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 1 woman missing plasma IGFBP-3 level.
‡ Units: ng/ml.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		4587	4392, 4782		
2	Rare	5	-751	-1487, -16	-415	-962, 132
2	AGCA/AGCA	12	-18	-516, 480	-13	-438, 412
2	TACA/AGCA	34	12	-323, 347	10	-300, 320
2	TACA/TACA	12	101	-397, 598	73	-352, 499
2	TACA/TATA	6	-174	-851, 502	-104	-625, 418
2	TACA/TGCA	53	-82	-375, 210	-73	-348, 203
2	TACA/TGTC	18	-51	-473, 371	-40	-415, 334
2	TATA/AGCA	9	254	-310, 817	172	-292, 636
2	TATA/TGCA	10	-41	-579, 497	-29	-478, 421
2	TGCA/TGCA	63	-80	-359, 200	-71	-336, 193
2	TGCA/TGTC	61	-142	-423, 140	-127	-393, 140
2	TGTC/AGCA	32	-68	-409, 274	-58	-373, 257
2	TGTC/TGTC	10	-163	-701, 376	-113	-563, 336
2	TGCA/AGCA	66	REF		REF	

* N = 401, excludes 1 woman missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 391, excludes 9 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 1 woman missing plasma IGFBP-3 level.

‡ Units: ng/ml.

Table A.8. Unadjusted linear regression of *IGFBP-3* group-specific diplotypes on plasma IGF-I levels among Caucasians*

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		177	163, 192		
1	Rare	40	-24	-47, 0	-20	-42, 1
1	GCGGAC/GCGGAC	21	12	-17, 41	10	-16, 36
1	TACGGA/TACGGC	23	0	-28, 28	0	-25, 25
1	TACGGA/TCGCAC	10	-13	-52, 26	-9	-42, 23
1	TACGGC/TACGGC	34	-5	-29, 19	-4	-27, 18
1	TACGGC/TAGGGC	13	-9	-44, 26	-7	-37, 23
1	TACGGC/TCGCAC	55	1	-20, 22	1	-19, 21
1	TACGGC/TCGGGA	31	2	-23, 27	2	-21, 25
1	TAGGGC/GCGGAC	6	-7	-56, 42	-4	-42, 33
1	TAGGGC/TCGCAC	6	-15	-64, 34	-9	-46, 29
1	TCGCAC/GCGGAC	22	-2	-30, 27	-1	-27, 24
1	TCGCAC/TCGCAC	14	0	-33, 34	0	-29, 29
1	TCGGGA/GCGGAC	11	-21	-58, 16	-15	-46, 17
1	TCGGGA/TCGCAC	15	-24	-57, 8	-19	-47, 10
1	TCGGGA/TCGGGA	6	-8	-57, 40	-5	-42, 33
1	TACGGC/GCGGAC	63	REF		REF	

* N = 401, excludes 1 woman missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.
† N = 370, excludes 30 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 1 woman missing plasma IGF-I level.
‡ Units: ng/ml.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		174	160, 188		
2	Rare	5	-50	-102, 2	-28	-67, 11
2	AGCA/AGCA	12	-22	-57, 14	-16	-46, 14
2	TACA/AGCA	34	-10	-34, 14	-9	-31, 13
2	TACA/TACA	12	2	-33, 38	2	-29, 32
2	TACA/TATA	6	-9	-58, 39	-6	-43, 32
2	TACA/TGCA	53	0	-21, 21	0	-19, 20
2	TACA/TGTC	18	12	-18, 42	9	-17, 36
2	TATA/AGCA	9	22	-18, 62	15	-18, 48
2	TATA/TGCA	10	22	-16, 61	16	-16, 48
2	TGCA/TGCA	63	4	-16, 24	3	-16, 22
2	TGCA/TGTC	61	-8	-28, 12	-7	-26, 12
2	TGTC/AGCA	32	-3	-27, 21	-2	-25, 20
2	TGTC/TGTC	10	0	-38, 39	0	-32, 32
2	TGCA/AGCA	66	REF		REF	

* N = 401, excludes 1 woman missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 391, excludes 9 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 1 woman missing plasma IGF-I level.

‡ Units: ng/ml.

Table A.9. Unadjusted linear regression of *IGFBP-3* group-specific diplotypes on plasma IGFBP-3 levels among African Americans*

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		4327	4166, 4488		
1	Rare	50	-420	-691, -149	-381	-639, -123
1	TACGAGG/GCGGAGA	18	-332	-729, 65	-272	-632, 87
1	TACGAGG/GCGGCAA	15	220	-209, 649	175	-208, 558
1	TACGAGG/TACGAGG	51	372	103, 641	338	81, 594
1	TAGGAGA/TACGAGG	10	-28	-541, 485	-21	-459, 418
1	TAGGAGA/TAGGAGG	7	-740	-1344, -136	-491	-982, 1
1	TAGGAGG/GCGGAGA	12	-835	-1307, -362	-636	-1049, -224
1	TAGGAGG/TAGGAGG	34	-487	-797, -178	-430	-720, -139
1	TCGCAGA/GCGGAGA	10	-242	-755, 271	-177	-616, 262
1	TCGCAGA/GCGGCAA	9	-681	-1219, -144	-486	-940, -32
1	TCGCAGA/TACGAGG	68	-182	-429, 65	-168	-405, 69
1	TCGCAGA/TAGGAGA	6	-719	-1368, -71	-453	-969, 62
1	TCGCAGA/TAGGAGG	44	-410	-693, -128	-369	-637, -101
1	TCGCAGA/TCGCAGA	26	-640	-982, -298	-550	-867, -233
1	TCGCAGA/TCGGAGG	20	-901	-1282, -521	-750	-1097, -403
1	TCGGAGA/TACGAGG	12	-309	-782, 164	-236	-649, 177
1	TCGGAGG/TACGAGG	30	25	-299, 349	22	-281, 325
1	TCGGAGG/TAGGAGG	19	-220	-608, 169	-182	-535, 171
1	TCGGAGG/TCGGAGG	7	-209	-812, 395	-138	-630, 353
1	TAGGAGG/TACGAGG	91	REF		REF	

* N = 579, excludes 3 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 539, excludes 32 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing plasma IGFBP-3 levels.

‡ Units: ng/ml.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		4060	3947, 4172		
2	AC/AC	7	225	-410, 860	144	-364, 652
2	CC/AC	28	229	-103, 560	198	-111, 507
2	CC/CC	57	-344	-590, -98	-317	-553, -81
2	CT/AC	60	205	-37, 446	189	-43, 421
2	CT/CT	194	70	-93, 233	67	-93, 228
2	CT/CC	217	REF		REF	

* N = 579, excludes 3 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 563, excludes 8 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing plasma IGFBP-3 levels.

‡ Units: ng/ml.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		4106	3959, 4254		
3	Rare	14	-360	-826, 106	-276	-685, 132
3	AGCA/AGCA	9	49	-521, 620	34	-439, 507
3	TACA/AGCA	15	55	-397, 507	43	-356, 441
3	TACA/TACA	10	21	-522, 565	15	-442, 473
3	TACA/TGCA	43	-261	-553, 31	-233	-509, 43
3	TATA/AGCA	47	-193	-476, 90	-174	-442, 95
3	TATA/TACA	40	-89	-389, 211	-79	-362, 204
3	TATA/TATA	54	257	-12, 526	234	-23, 490
3	TATA/TGTC	18	-149	-566, 268	-120	-494, 254
3	TGCA/AGCA	51	74	-200, 349	67	-194, 328
3	TGCA/TGCA	101	89	-132, 310	84	-130, 297
3	TGTC/AGCA	6	-1	-692, 690	-1	-536, 535
3	TGTC/TGCA	13	-279	-761, 203	-211	-629, 208
3	TATA/TGCA	126	REF		REF	

* N = 579, excludes 3 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 547, excludes 24 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing plasma IGFBP-3 levels.

‡ Units: ng/ml.

Table A.10. Unadjusted linear regression of *IGFBP-3* group-specific diplotypes on plasma IGF-I levels among African Americans*

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		175	161, 190		
1	Rare	50	-19	-44, 5	-17	-41, 6
1	TACGAGG/GCGGAGA	18	48	12, 84	39	6, 71
1	TACGAGG/GCGGCAA	15	-19	-58, 20	-15	-49, 20
1	TACGAGG/TACGAGG	51	-6	-31, 18	-6	-29, 17
1	TAGGAGA/TACGAGG	10	-23	-70, 23	-17	-56, 23
1	TAGGAGA/TAGGAGG	7	-33	-88, 22	-21	-65, 23
1	TAGGAGG/GCGGAGA	12	-50	-93, -7	-38	-75, 0
1	TAGGAGG/TAGGAGG	34	1	-27, 29	1	-25, 27
1	TCGCAGA/GCGGAGA	10	-8	-54, 39	-5	-45, 34
1	TCGCAGA/GCGGCAA	9	-33	-82, 16	-23	-64, 18
1	TCGCAGA/TACGAGG	68	-20	-42, 3	-18	-39, 3
1	TCGCAGA/TAGGAGA	6	4	-55, 63	2	-44, 48
1	TCGCAGA/TAGGAGG	44	-10	-35, 16	-9	-33, 16
1	TCGCAGA/TCGCAGA	26	-31	-62, 0	-26	-55, 2
1	TCGCAGA/TCGGAGG	20	-21	-55, 14	-17	-48, 14
1	TCGGAGA/TACGAGG	12	-15	-58, 28	-11	-48, 26
1	TCGGAGG/TACGAGG	30	-19	-49, 10	-17	-44, 11
1	TCGGAGG/TAGGAGG	19	3	-32, 38	2	-30, 34
1	TCGGAGG/TCGGAGG	7	30	-24, 85	20	-24, 64
1	TAGGAGG/TACGAGG	91	REF		REF	

* N = 579, excludes 3 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 539, excludes 32 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing plasma IGF-I levels.

‡ Units: ng/ml.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		162	152, 172		
2	AC/AC	7	-20	-76, 36	-13	-57, 32
2	CC/AC	28	18	-11, 47	15	-12, 43
2	CC/CC	57	3	-19, 25	3	-18, 24
2	CT/AC	60	7	-14, 29	7	-14, 27
2	CT/CT	194	4	-10, 19	4	-10, 18
2	CT/CC	217	REF		REF	

* N = 579, excludes 3 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 563, excludes 8 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing plasma IGF-I levels.

‡ Units: ng/ml.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		162	150, 175		
3	Rare	14	-25	-65, 16	-19	-54, 16
3	AGCA/AGCA	9	2	-48, 51	1	-40, 42
3	TACA/AGCA	15	-11	-50, 28	-9	-43, 26
3	TACA/TACA	10	2	-45, 49	1	-38, 41
3	TACA/TGCA	43	-8	-34, 17	-8	-31, 16
3	TATA/AGCA	47	-9	-34, 15	-8	-31, 15
3	TATA/TACA	40	-7	-33, 19	-7	-31, 18
3	TATA/TATA	54	12	-11, 35	11	-11, 33
3	TATA/TGTC	18	62	26, 98	50	17, 82
3	TGCA/AGCA	51	7	-17, 31	6	-16, 29
3	TGCA/TGCA	101	10	-10, 29	9	-10, 27
3	TGTC/AGCA	6	-35	-95, 25	-21	-68, 25
3	TGTC/TGCA	13	-1	-43, 40	-1	-37, 35
3	TATA/TGCA	126	REF		REF	

* N = 579, excludes 3 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 547, excludes 24 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing plasma IGF-I levels.

‡ Units: ng/ml.

Table A.11. Unadjusted linear regression of *IGF-I* group-specific diplotypes on plasma IGFBP-3 levels among Caucasians*

Group†	Diplotypes	N‡	Maximum Likelihood Estimates§	95% Confidence Interval§	Posterior Medians§	95% Posterior Limits§
	Intercept		4525	4429, 4620		
1	A/G, A/A	114	-15	-193, 164	-14	-188, 160
1	G/G	280	REF		REF	

* N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.
† rs35767 represents entire group because it could not be combined with block 1 SNPs for diplotype estimation.
‡ N = 394, excludes 4 women missing SNP genotype and 1 woman missing plasma IGFBP-3 level.
§ Units: ng/ml.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		4559	4407, 4711		
2	Rare	10	545	19, 1071	386	-57, 828
2	AAT/ACT	11	-170	-673, 334	-123	-552, 306
2	AAT/GCC	23	52	-313, 417	43	-290, 377
2	AAT/GCT	52	75	-193, 343	68	-187, 323
2	ACT/GCC	27	-173	-515, 169	-147	-463, 168
2	ACT/GCT	34	-1	-314, 311	-1	-293, 291
2	GCC/GCC	29	-185	-518, 147	-159	-467, 149
2	GCT/GCT	97	-100	-322, 122	-93	-307, 121
2	GCT/GCC	110	REF		REF	

* N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.
† N = 393, excludes 5 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 1 woman missing plasma IGFBP-3 level.
‡ Units: ng/ml.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		4518	4353, 4683		
3	Rare	22	-306	-683, 71	-252	-595, 90
3	GCTCT/GCTGC	9	-33	-588, 522	-23	-482, 437
3	GGCCC/AGTGC	21	84	-300, 468	69	-279, 417
3	GGCCC/GCTCT	26	-42	-395, 311	-35	-359, 289
3	GGCCC/GCTGC	36	-61	-373, 252	-53	-345, 239
3	GGCCC/GGCCC	78	82	-162, 326	75	-159, 309
3	GGCCT/AGTGC	15	-213	-656, 230	-165	-554, 225
3	GGCCT/GCTCT	25	179	-179, 537	150	-178, 478
3	GGCCT/GCTGC	24	-202	-566, 163	-168	-501, 165
3	GGCCT/GGCCT	33	291	-32, 613	251	-48, 551
3	GGCCT/GGTGT	7	-212	-835, 412	-134	-630, 362
3	GGCCC/GGCCT	93	REF		REF	

* N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.

† N = 389, excludes 9 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 1 woman missing plasma IGFBP-3 level.

‡ Units: ng/ml.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		4585	4463, 4706		
4	CA/TA, Rare	16	-360	-779, 59	-285	-658, 88
4	CA/CA	21	-221	-592, 150	-184	-521, 154
4	CG/CA	119	-23	-214, 167	-22	-208, 164
4	CG/TA	62	-183	-420, 54	-169	-397, 59
4	CG/CG	175	REF		REF	

* N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.

† N = 393, excludes 5 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 1 woman missing plasma IGFBP-3 level.

‡ Units: ng/ml.

Table A.12. Unadjusted linear regression of *IGF-I* group-specific diplotypes on plasma IGF-I levels among Caucasians*

Group†	Diplotypes	N‡	Maximum Likelihood Estimates§	95% Confidence Interval§	Posterior Medians§	95% Posterior Limits§
	Intercept		171	164, 177		
1	A/G, A/A	114	6	-7, 19	6	-7, 18
1	G/G	280	REF		REF	

* N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.
† rs35767 represents entire group because it could not be combined with block 1 SNPs for diplotype estimation.
‡ N = 394, excludes 4 women missing SNP genotype and 1 woman missing plasma IGF-I level.
§ Units: ng/ml.

.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		173	162, 184		
2	Rare	10	3	-35, 41	2	-30, 34
2	AAT/ACT	11	-7	-44, 29	-5	-36, 26
2	AAT/GCC	23	8	-18, 34	7	-17, 31
2	AAT/GCT	52	12	-8, 31	11	-8, 29
2	ACT/GCC	27	-3	-27, 22	-2	-25, 20
2	ACT/GCT	34	-13	-35, 10	-11	-32, 10
2	GCC/GCC	29	-6	-30, 18	-5	-28, 17
2	GCT/GCT	97	-4	-20, 12	-4	-20, 11
2	GCT/GCC	110	REF		REF	

* N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.
† N = 393, excludes 5 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 1 woman missing plasma IGF-I level.
‡ Units: ng/ml.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		165	153, 176		
3	Rare	22	-24	-51, 3	-20	-44, 5
3	GCTCT/GCTGC	9	25	-14, 64	17	-15, 50
3	GGCCC/AGTGC	21	21	-6, 48	17	-8, 42
3	GGCCC/GCTCT	26	3	-22, 28	3	-20, 26
3	GGCCC/GCTGC	36	7	-15, 29	6	-15, 27
3	GGCCC/GGCCC	78	6	-11, 24	6	-11, 23
3	GGCCT/AGTGC	15	25	-6, 57	20	-8, 48
3	GGCCT/GCTCT	25	13	-13, 38	11	-13, 34
3	GGCCT/GCTGC	24	1	-25, 27	1	-23, 24
3	GGCCT/GGCCT	33	29	6, 52	25	4, 46
3	GGCCT/GGTGT	7	41	-3, 86	26	-9, 62
3	GGCCC/GGCCT	93	REF		REF	

* N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.
† N = 389, excludes 9 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 1 woman missing plasma IGF-I level.
‡ Units: ng/ml.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		172	164, 181		
4	CA/TA, Rare	16	-31	-61, -1	-25	-52, 2
4	CA/CA	21	14	-12, 41	12	-12, 36
4	CG/CA	119	1	-13, 14	1	-13, 14
4	CG/TA	62	2	-15, 19	2	-14, 18
4	CG/CG	175	REF		REF	

* N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.
† N = 393, excludes 5 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 1 woman missing plasma IGF-I level.
‡ Units: ng/ml.

Table A.13. Unadjusted linear regression of *IGF-I* group-specific diplotypes on plasma IGFBP-3 levels among African Americans*

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		4043	3919, 4167		
1	Rare	9	-841	-1409, -272	-580	-1052, -108
1	AGG/AGG	99	106	-102, 314	100	-102, 302
1	AGG/GCG	52	-12	-274, 250	-11	-261, 239
1	AGG/GGA	31	-3	-327, 320	-3	-305, 299
1	GCG/GCG	8	-221	-822, 380	-147	-637, 343
1	GGG/GCG	57	148	-105, 401	136	-106, 378
1	GGG/GGA	24	7	-355, 368	6	-327, 338
1	GGG/GGG	105	137	-67, 342	130	-69, 328
1	AGG/GGG	180	REF		REF	

* N = 580, excludes 2 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.

† N = 565, excludes 7 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing plasma IGFBP-3 levels.

‡ Units: ng/ml.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		3999	3866, 4133		
2	Rare	38	109	-187, 405	97	-182, 377
2	ATCCC/TCTTT	13	153	-319, 624	117	-295, 529
2	ATTCC/ATCCC	37	107	-192, 406	95	-187, 377
2	ATTCC/ATTTT	10	-101	-633, 432	-72	-523, 378
2	ATTCC/TCTCC	8	-35	-626, 557	-23	-508, 462
2	ATTCC/TCTCT	15	392	-50, 833	308	-83, 700
2	ATTCC/TCTTT	119	67	-133, 267	64	-131, 258
2	ATTCC/TTTCT	34	10	-300, 319	9	-282, 299
2	ATTCT/ATCCC	6	-212	-891, 466	-129	-659, 400
2	ATTCT/ATTCC	41	154	-133, 442	138	-134, 411
2	ATTCT/TCTCT	6	19	-660, 698	12	-518, 541
2	ATTCT/TCTTT	20	-154	-542, 234	-127	-480, 226
2	ATTCT/TTTCT	8	186	-405, 778	125	-360, 610
2	TCTCT/TCTTT	10	73	-460, 605	52	-399, 503
2	TCTTT/TCTTT	30	179	-147, 505	156	-149, 460
2	TTTCT/TCTTT	13	783	312, 1255	598	186, 1010
2	ATTCC/ATTCC	150	REF		REF	

* N = 580, excludes 2 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.

† N = 558, excludes 14 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing plasma IGFBP-3 levels.

‡ Units: ng/ml.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		4083	3915, 4251		
3	Rare	9	-679	-1257, -101	-463	-941, 14
3	AACC/AACC	7	260	-389, 910	164	-351, 679
3	AACC/AACT	7	-101	-750, 549	-63	-579, 452
3	AACC/AGGT	15	371	-89, 832	287	-118, 691
3	AACC/TACC	11	-423	-950, 105	-304	-752, 143
3	AACC/TACT	55	-107	-386, 173	-96	-362, 169
3	AACT/AGGT	6	-217	-915, 481	-129	-668, 410
3	AACT/TACT	13	9	-481, 498	6	-418, 430
3	AAGC/AACC	35	-68	-394, 259	-59	-364, 246
3	AAGC/AACT	9	466	-112, 1044	318	-160, 795
3	AAGC/AAGC	34	58	-272, 388	50	-257, 358
3	AAGC/AGGT	24	149	-229, 527	124	-221, 470
3	AAGC/TACC	28	-99	-455, 257	-84	-412, 244
3	AAGC/TACT	86	19	-226, 264	18	-218, 253
3	AGGT/AGGT	10	66	-485, 617	47	-415, 508
3	AGGT/TACT	47	21	-274, 315	18	-260, 297
3	TACC/TACC	6	-351	-1049, 347	-209	-748, 329
3	TACC/TACT	45	3	-296, 302	3	-279, 285
3	TACT/TACT	98	REF		REF	

* N = 580, excludes 2 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.

† N = 545, excludes 27 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing plasma IGFBP-3 levels.

‡ Units: ng/ml.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		4048	3938, 4158		
4	CA/CA	93	171	-34, 377	162	-38, 362
4	CA/TA	40	-29	-316, 257	-26	-298, 245
4	CG/CG	147	-22	-199, 155	-21	-194, 152
4	CG/TA	45	162	-111, 435	147	-113, 407
4	TA/TA	8	283	-320, 885	188	-304, 679
4	CA/CG	232	REF		REF	

* N = 580, excludes 2 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.

† N = 565, excludes 7 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing plasma IGFBP-3 levels.

‡ Units: ng/ml.

Table A.14. Unadjusted linear regression of *IGF-I* group-specific diplotypes on plasma IGF-I levels among African Americans*

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		166	155, 176		
1	Rare	9	-68	-117, -20	-48	-88, -7
1	AGG/AGG	99	0	-18, 17	0	-18, 17
1	AGG/GCG	52	-5	-28, 17	-5	-26, 17
1	AGG/GGA	31	-7	-34, 21	-6	-32, 20
1	GCG/GCG	8	-27	-78, 25	-18	-60, 24
1	GGG/GCG	57	31	9, 52	28	8, 49
1	GGG/GGA	24	-2	-33, 29	-2	-30, 27
1	GGG/GGG	105	-7	-24, 11	-6	-23, 11
1	AGG/GGG	180	REF		REF	

* N = 580, excludes 2 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.

† N = 565, excludes 7 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing plasma IGF-I levels.

‡ Units: ng/ml.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		162	150, 174		
2	Rare	38	-1	-28, 25	-1	-26, 23
2	ATCCC/TCTTT	13	25	-17, 66	19	-17, 55
2	ATTCC/ATCCC	37	-14	-41, 12	-13	-37, 12
2	ATTCC/ATTTT	10	-5	-52, 42	-3	-43, 36
2	ATTCC/TCTCC	8	-29	-81, 24	-19	-62, 23
2	ATTCC/TCTCT	15	29	-10, 68	23	-12, 57
2	ATTCC/TCTTT	119	8	-10, 26	8	-10, 25
2	ATTCC/TTTCT	34	6	-21, 33	5	-20, 31
2	ATTCT/ATCCC	6	-12	-71, 48	-7	-53, 40
2	ATTCT/ATTCC	41	2	-24, 27	2	-22, 25
2	ATTCT/TCTCT	6	8	-52, 67	5	-42, 51
2	ATTCT/TCTTT	20	-3	-37, 31	-2	-33, 29
2	ATTCT/TTTCT	8	8	-44, 60	5	-37, 48
2	TCTCT/TCTTT	10	-13	-60, 34	-9	-49, 31
2	TCTTT/TCTTT	30	12	-17, 41	10	-16, 37
2	TTTCT/TCTTT	13	21	-20, 63	16	-20, 52
2	ATTCC/ATTCC	150	REF		REF	

* N = 580, excludes 2 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.

† N = 558, excludes 14 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing plasma IGF-I levels.

‡ Units: ng/ml.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		166	152, 181		
3	Rare	9	-38	-87, 12	-26	-67, 15
3	AACC/AACC	7	5	-51, 61	3	-41, 48
3	AACC/AACT	7	-18	-74, 38	-12	-56, 33
3	AACC/AGGT	15	-4	-43, 36	-3	-38, 32
3	AACC/TACC	11	-10	-55, 35	-7	-46, 31
3	AACC/TACT	55	-6	-30, 18	-5	-28, 18
3	AACT/AGGT	6	-14	-74, 46	-8	-55, 38
3	AACT/TACT	13	10	-33, 52	7	-29, 44
3	AAGC/AACC	35	0	-28, 28	0	-26, 26
3	AAGC/AACT	9	109	59, 159	75	34, 116
3	AAGC/AAGC	34	8	-21, 36	7	-20, 33
3	AAGC/AGGT	24	4	-28, 37	4	-26, 33
3	AAGC/TACC	28	-11	-41, 20	-9	-37, 19
3	AAGC/TACT	86	-9	-30, 12	-9	-29, 12
3	AGGT/AGGT	10	18	-29, 65	13	-27, 53
3	AGGT/TACT	47	2	-23, 28	2	-22, 26
3	TACC/TACC	6	-28	-88, 32	-17	-63, 30
3	TACC/TACT	45	-12	-37, 14	-10	-35, 14
3	TACT/TACT	98	REF		REF	

* N = 580, excludes 2 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.

† N = 545, excludes 27 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing plasma IGF-I levels.

‡ Units: ng/ml.

Group	Diplotypes	N†	Maximum Likelihood Estimates‡	95% Confidence Interval‡	Posterior Medians‡	95% Posterior Limits‡
	Intercept		162	152, 171		
4	CA/CA	93	4	-14, 22	4	-13, 21
4	CA/TA	40	38	13, 63	34	11, 57
4	CG/CG	147	-2	-18, 13	-2	-17, 12
4	CG/TA	45	-1	-25, 22	-1	-23, 21
4	TA/TA	8	-1	-52, 51	0	-43, 42
4	CA/CG	232	REF		REF	

* N = 580, excludes 2 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.
† N = 565, excludes 7 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing plasma IGF-I levels.
‡ Units: ng/ml.

Table A.15. Unadjusted prevalence differences of uterine fibroids (any size) with *IGFBP-3* group-specific diplotypes among Caucasians*

Group	Diplotypes	N (any fibroids)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
1	Rare	24	45	-0.024	-0.216, 0.168	-0.015	-0.167, 0.137
1	GCGGAC/GCGGAC	13	21	0.062	-0.181, 0.304	0.032	-0.142, 0.206
1	TACGGA/TACGGC	8	22	-0.194	-0.430, 0.043	-0.102	-0.274, 0.070
1	TACGGA/TCGCAC	5	10	-0.057	-0.391, 0.277	-0.021	-0.221, 0.180
1	TACGGC/TACGGC	21	33	0.079	-0.127, 0.285	0.047	-0.112, 0.206
1	TACGGC/TAGGGC	5	13	-0.173	-0.465, 0.120	-0.073	-0.263, 0.117
1	TACGGC/TCGCAC	28	55	-0.048	-0.230, 0.133	-0.032	-0.179, 0.115
1	TACGGC/TCGGGA	13	31	-0.138	-0.352, 0.076	-0.080	-0.242, 0.083
1	TAGGGC/GCGGAC	1	6	-0.391	-0.714, -0.068	-0.146	-0.344, 0.051
1	TCGCAC/GCGGAC	11	23	-0.079	-0.318, 0.160	-0.041	-0.214, 0.132
1	TCGCAC/TCGCAC	4	13	-0.250	-0.530, 0.030	-0.111	-0.297, 0.076
1	TCGGGA/GCGGAC	4	11	-0.194	-0.504, 0.117	-0.076	-0.271, 0.118
1	TCGGGA/TCGCAC	10	15	0.109	-0.160, 0.378	0.051	-0.133, 0.234
1	TCGGGA/TCGGGA	4	6	0.109	-0.288, 0.507	0.031	-0.181, 0.243
1	TACGGC/GCGGAC	34	61	REF		REF	

* N = 401, excludes 1 woman missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 365, excludes 30 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 6 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Group	Diplotypes	N (any fibroids)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
2	Rare	2	5	-0.131	-0.578, 0.315	-0.031	-0.249, 0.187
2	AGCA/AGCA	8	12	0.135	-0.158, 0.429	0.057	-0.133, 0.247
2	TACA/AGCA	20	34	0.057	-0.149, 0.263	0.034	-0.125, 0.193
2	TACA/TACA	6	12	-0.031	-0.339, 0.277	-0.012	-0.207, 0.182
2	TACA/TATA	2	6	-0.198	-0.594, 0.199	-0.056	-0.268, 0.155
2	TACA/TGCA	21	51	-0.119	-0.302, 0.063	-0.078	-0.225, 0.069
2	TACA/TGTC	10	17	0.057	-0.207, 0.321	0.027	-0.155, 0.208
2	TATA/AGCA	5	9	0.024	-0.323, 0.371	0.008	-0.195, 0.211
2	TATA/TGCA	5	9	0.024	-0.323, 0.371	0.008	-0.195, 0.211
2	TGCA/TGCA	28	62	-0.080	-0.254, 0.094	-0.054	-0.196, 0.089
2	TGCA/TGTC	30	61	-0.039	-0.215, 0.136	-0.026	-0.170, 0.117
2	TGTC/AGCA	17	32	0.000	-0.212, 0.212	0.000	-0.162, 0.162
2	TGTC/TGTC	7	10	0.169	-0.140, 0.478	0.067	-0.128, 0.261
2	TGCA/AGCA	34	64	REF		REF	

* N = 401, excludes 1 woman missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 384, excludes 9 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Table A.16. Unadjusted prevalence differences of uterine fibroids (2+ cm) with *IGFBP-3* group-specific diplotypes among Caucasians*

Group	Diplotypes	N (fibroids 2+ cm)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
1	Rare‖	16	51	-0.096	-0.273, 0.081	-0.064	-0.209, 0.081
1	GCGGAC/GCGGAC	7	21	-0.077	-0.313, 0.160	-0.040	-0.212, 0.131
1	TACGGA/TACGGC	5	22	-0.183	-0.397, 0.032	-0.105	-0.268, 0.057
1	TACGGA/TCGCAC	3	10	-0.110	-0.420, 0.200	-0.043	-0.238, 0.151
1	TACGGC/TACGGC	12	33	-0.046	-0.252, 0.159	-0.028	-0.186, 0.131
1	TACGGC/TAGGGC	4	13	-0.102	-0.382, 0.177	-0.045	-0.232, 0.141
1	TACGGC/TCGCAC	19	55	-0.064	-0.241, 0.112	-0.043	-0.187, 0.101
1	TACGGC/TCGGGA	9	31	-0.120	-0.321, 0.082	-0.072	-0.229, 0.085
1	TCGCAC/GCGGAC	7	23	-0.105	-0.330, 0.119	-0.058	-0.226, 0.109
1	TCGCAC/TCGCAC	2	13	-0.256	-0.488, -0.024	-0.138	-0.308, 0.032
1	TCGGGA/GCGGAC	3	11	-0.137	-0.428, 0.154	-0.058	-0.248, 0.131
1	TCGGGA/TCGCAC	5	15	-0.077	-0.345, 0.192	-0.036	-0.219, 0.147
1	TCGGGA/TCGGGA	4	6	0.257	-0.140, 0.654	0.073	-0.139, 0.284
1	TACGGC/GCGGAC	25	61	REF		REF	

* N = 401, excludes 1 woman missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 365, excludes 30 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 6 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

‖ Includes TAGGGC/GCGGAC diplotype, which was analyzed separately in relation to any fibroids. However, no women with this diplotype had fibroids 2+ cm, and estimating a prevalence difference for this diplotype resulted in convergences problems for model.

Group	Diplotypes	N (fibroids 2+ cm)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
2	Rare	2	5	0.088	-0.357, 0.532	0.021	-0.197, 0.239
2	AGCA/AGCA	5	12	0.104	-0.197, 0.405	0.042	-0.150, 0.235
2	TACA/AGCA	15	34	0.129	-0.073, 0.331	0.078	-0.079, 0.235
2	TACA/TACA	3	12	-0.062	-0.333, 0.208	-0.029	-0.212, 0.155
2	TACA/TATA	1	6	-0.146	-0.465, 0.173	-0.055	-0.252, 0.141
2	TACA/TGCA	11	51	-0.097	-0.257, 0.063	-0.069	-0.203, 0.066
2	TACA/TGTC	8	17	0.158	-0.105, 0.421	0.075	-0.106, 0.256
2	TATA/AGCA	5	9	0.243	-0.101, 0.587	0.084	-0.118, 0.286
2	TATA/TGCA	4	9	0.132	-0.212, 0.476	0.046	-0.157, 0.248
2	TGCA/TGCA	22	62	0.042	-0.122, 0.207	0.030	-0.108, 0.167
2	TGCA/TGTC	17	61	-0.034	-0.194, 0.126	-0.024	-0.159, 0.111
2	TGTC/AGCA	11	32	0.031	-0.169, 0.231	0.019	-0.137, 0.175
2	TGTC/TGTC	5	10	0.188	-0.143, 0.518	0.068	-0.131, 0.268
2	TGCA/AGCA	20	64	REF		REF	

* N = 401, excludes 1 woman missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 384, excludes 9 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Table A.17. Unadjusted prevalence differences of uterine fibroids (any size) with *IGFBP-3* group-specific diplotypes among African Americans*

Group	Diplotypes	N (any fibroids)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
1	Rare	33	49	-0.038	-0.199, 0.124	-0.027	-0.162, 0.109
1	TACGAGG/GCGGAGA	11	17	-0.064	-0.310, 0.182	-0.033	-0.208, 0.143
1	TACGAGG/GCGGCAA	11	15	0.022	-0.220, 0.265	0.011	-0.163, 0.186
1	TACGAGG/TACGAGG	38	51	0.034	-0.118, 0.186	0.025	-0.105, 0.155
1	TAGGAGA/TACGAGG	6	10	-0.111	-0.429, 0.207	-0.042	-0.239, 0.154
1	TAGGAGA/TAGGAGG	5	7	0.003	-0.344, 0.351	0.001	-0.202, 0.204
1	TAGGAGG/GCGGAGA	7	13	-0.173	-0.459, 0.114	-0.075	-0.263, 0.114
1	TAGGAGG/TAGGAGG	27	34	0.083	-0.082, 0.248	0.058	-0.080, 0.196
1	TCGCAGA/GCGGAGA	9	10	0.189	-0.019, 0.397	0.112	-0.048, 0.272
1	TCGCAGA/GCGGCAA	4	9	-0.267	-0.605, 0.071	-0.094	-0.295, 0.107
1	TCGCAGA/TACGAGG	52	71	0.021	-0.118, 0.160	0.016	-0.105, 0.138
1	TCGCAGA/TAGGAGA	6	8	0.039	-0.275, 0.353	0.015	-0.181, 0.211
1	TCGCAGA/TAGGAGG	33	44	0.039	-0.120, 0.197	0.028	-0.106, 0.162
1	TCGCAGA/TCGCAGA	23	25	0.209	0.067, 0.351	0.158	0.035, 0.281
1	TCGCAGA/TCGGAGG	15	20	0.039	-0.173, 0.251	0.023	-0.139, 0.184
1	TCGGAGA/TACGAGG	8	11	0.016	-0.263, 0.296	0.007	-0.179, 0.193
1	TCGGAGG/TACGAGG	21	31	-0.034	-0.223, 0.156	-0.021	-0.172, 0.130
1	TCGGAGG/TAGGAGG	15	19	0.078	-0.127, 0.284	0.047	-0.112, 0.206
1	TCGGAGG/TCGGAGG	4	7	-0.140	-0.518, 0.239	-0.042	-0.251, 0.166
1	TAGGAGG/TACGAGG	64	90	REF		REF	

* N = 579, excludes 3 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 541, excludes 32 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 6 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Group	Diplotypes	N (any fibroids)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
2	AC/AC	6	7	0.121	-0.145, 0.387	0.057	-0.125, 0.239
2	CC/AC	19	28	-0.058	-0.240, 0.125	-0.038	-0.185, 0.110
2	CC/CC	44	58	0.023	-0.102, 0.147	0.018	-0.094, 0.130
2	CT/AC	40	60	-0.069	-0.202, 0.064	-0.054	-0.172, 0.063
2	CT/CT	144	195	0.002	-0.083, 0.088	0.002	-0.079, 0.083
2	CT/CC	159	216	REF		REF	

* N = 579, excludes 3 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 564, excludes 8 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 7 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Group	Diplotypes	N (any fibroids)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
3	Rare	12	14	0.067	-0.130, 0.264	0.041	-0.113, 0.196
3	AGCA/AGCA	6	9	-0.124	-0.440, 0.193	-0.048	-0.244, 0.149
3	TACA/AGCA	8	15	-0.257	-0.519, 0.005	-0.122	-0.303, 0.059
3	TACA/TACA	7	10	-0.090	-0.383, 0.203	-0.038	-0.228, 0.152
3	TACA/TGCA	31	43	-0.069	-0.221, 0.083	-0.051	-0.181, 0.079
3	TATA/AGCA	36	47	-0.024	-0.165, 0.116	-0.019	-0.141, 0.104
3	TATA/TACA	30	40	-0.040	-0.192, 0.112	-0.029	-0.159, 0.101
3	TATA/TATA	40	52	-0.021	-0.156, 0.114	-0.016	-0.135, 0.103
3	TATA/TGTC	12	18	-0.124	-0.353, 0.106	-0.067	-0.236, 0.102
3	TGCA/AGCA	43	53	0.021	-0.106, 0.148	0.017	-0.097, 0.130
3	TGCA/TGCA	71	104	-0.108	-0.222, 0.007	-0.089	-0.193, 0.015
3	TGTC/AGCA	5	6	0.043	-0.264, 0.350	0.017	-0.177, 0.211
3	TGTC/TGCA	8	13	-0.175	-0.449, 0.099	-0.079	-0.264, 0.105
3	TATA/TGCA	98	124	REF		REF	

* N = 579, excludes 3 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 548, excludes 24 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 7 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Table A.18. Unadjusted prevalence differences of uterine fibroids (2+ cm) with *IGFBP-3* group-specific diplotypes among African Americans*

Group	Diplotypes	N (fibroids 2+ cm)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
1	Rare	27	49	-0.027	-0.199, 0.146	-0.018	-0.160, 0.124
1	TACGAGG/GCGGAGA	10	17	0.010	-0.245, 0.266	0.005	-0.173, 0.184
1	TACGAGG/GCGGCAA	9	15	0.022	-0.246, 0.290	0.010	-0.173, 0.193
1	TACGAGG/TACGAGG	30	51	0.010	-0.159, 0.180	0.007	-0.133, 0.147
1	TAGGAGA/TACGAGG	5	10	-0.078	-0.404, 0.248	-0.029	-0.227, 0.170
1	TAGGAGA/TAGGAGG	5	7	0.137	-0.213, 0.486	0.046	-0.157, 0.250
1	TAGGAGG/GCGGAGA	7	13	-0.039	-0.329, 0.250	-0.017	-0.206, 0.172
1	TAGGAGG/TAGGAGG	23	34	0.099	-0.089, 0.286	0.063	-0.087, 0.213
1	TCGCAGA/GCGGAGA	9	10	0.322	0.110, 0.534	0.187	0.026, 0.349
1	TCGCAGA/GCGGCAA	3	9	-0.244	-0.569, 0.080	-0.091	-0.289, 0.107
1	TCGCAGA/TACGAGG	36	71	-0.071	-0.225, 0.084	-0.051	-0.183, 0.080
1	TCGCAGA/TAGGAGA	4	8	-0.078	-0.439, 0.283	-0.025	-0.231, 0.180
1	TCGCAGA/TAGGAGG	25	44	-0.010	-0.188, 0.169	-0.006	-0.152, 0.139
1	TCGCAGA/TCGCAGA	18	25	0.142	-0.061, 0.346	0.086	-0.072, 0.243
1	TCGCAGA/TCGGAGG	13	20	0.072	-0.160, 0.305	0.039	-0.132, 0.209
1	TCGGAGA/TACGAGG	6	11	-0.032	-0.344, 0.279	-0.013	-0.208, 0.182
1	TCGGAGG/TACGAGG	14	31	-0.126	-0.329, 0.077	-0.076	-0.234, 0.081
1	TCGGAGG/TAGGAGG	9	19	-0.104	-0.351, 0.143	-0.053	-0.228, 0.123
1	TCGGAGG/TCGGAGG	2	7	-0.292	-0.642, 0.058	-0.099	-0.302, 0.105
1	TAGGAGG/TACGAGG	52	90	REF		REF	

* N = 579, excludes 3 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 541, excludes 32 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 6 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Group	Diplotypes	N (fibroids 2+ cm)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
2	AC/AC	5	7	0.140	-0.201, 0.481	0.049	-0.153, 0.251
2	CC/AC	13	28	-0.110	-0.306, 0.086	-0.068	-0.222, 0.086
2	CC/CC	37	58	0.064	-0.076, 0.204	0.049	-0.074, 0.171
2	CT/AC	34	60	-0.007	-0.149, 0.134	-0.006	-0.129, 0.118
2	CT/CT	107	195	-0.025	-0.121, 0.071	-0.022	-0.112, 0.068
2	CT/CC	124	216	REF		REF	

* N = 579, excludes 3 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 564, excludes 8 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 7 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Group	Diplotypes	N (fibroids 2+ cm)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
3	Rare	8	14	-0.082	-0.354, 0.191	-0.037	-0.222, 0.147
3	AGCA/AGCA	3	9	-0.320	-0.639, -0.001	-0.122	-0.318, 0.075
3	TACA/AGCA	5	15	-0.320	-0.573, -0.067	-0.158	-0.336, 0.020
3	TACA/TACA	7	10	0.047	-0.249, 0.343	0.019	-0.172, 0.210
3	TACA/TGCA	28	43	-0.002	-0.167, 0.163	-0.001	-0.139, 0.136
3	TATA/AGCA	27	47	-0.079	-0.243, 0.086	-0.055	-0.192, 0.082
3	TATA/TACA	26	40	-0.003	-0.173, 0.167	-0.002	-0.143, 0.138
3	TATA/TATA	29	52	-0.096	-0.254, 0.063	-0.068	-0.202, 0.066
3	TATA/TGTC	9	18	-0.153	-0.399, 0.092	-0.078	-0.253, 0.097
3	TGCA/AGCA	36	53	0.026	-0.125, 0.177	0.019	-0.110, 0.148
3	TGCA/TGCA	50	104	-0.172	-0.300, -0.045	-0.137	-0.250, -0.023
3	TGTC/AGCA	4	6	0.013	-0.373, 0.400	0.004	-0.206, 0.214
3	TGTC/TGCA	5	13	-0.269	-0.546, 0.009	-0.120	-0.306, 0.065
3	TATA/TGCA	81	124	REF		REF	

* N = 579, excludes 3 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 548, excludes 24 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 7 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Table A.19. Unadjusted prevalence differences of uterine fibroids (any size) with *IGF-I* group-specific diplotypes among Caucasians*

Group†	Diplotypes	N (any fibroids)‡	total N‡	Prevalence Differences‖	95% Confidence Intervals‖	Posterior Medians	95% Posterior Limits
1	A/G, A/A	59	114	0.008	-0.101, 0.118	0.007	-0.093, 0.107
1	G/G	139	273	REF		REF	

* N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.
† rs35767 represents entire group because it could not be combined with block 1 SNPs for diplotype estimation.
‡ N = 387, excludes 4 women missing SNP genotype and 8 women missing uterine fibroid status.
‖ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Group	Diplotypes	N (any fibroids)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
2	Rare	5	10	-0.028	-0.352, 0.296	-0.010	-0.208, 0.188
2	AAT/ACT	3	10	-0.228	-0.527, 0.071	-0.094	-0.285, 0.098
2	AAT/GCC	10	22	-0.073	-0.302, 0.155	-0.040	-0.209, 0.129
2	AAT/GCT	25	50	-0.028	-0.195, 0.140	-0.019	-0.158, 0.120
2	ACT/GCC	14	26	0.011	-0.203, 0.224	0.006	-0.156, 0.169
2	ACT/GCT	20	34	0.060	-0.130, 0.251	0.038	-0.113, 0.190
2	GCC/GCC	17	28	0.079	-0.125, 0.283	0.048	-0.110, 0.206
2	GCT/GCT	46	98	-0.058	-0.195, 0.078	-0.045	-0.165, 0.075
2	GCT/GCC	57	108	REF		REF	

* N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.
† N = 386, excludes 5 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing uterine fibroid status.
‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Group	Diplotypes	N (any fibroids)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
3	Rare	11	22	-0.120	-0.351, 0.112	-0.064	-0.234, 0.105
3	GCTCT/GCTGC	4	8	-0.120	-0.480, 0.241	-0.039	-0.244, 0.167
3	GGCCC/AGTGC	11	21	-0.096	-0.331, 0.140	-0.051	-0.222, 0.121
3	GGCCC/GCTCT	14	25	-0.060	-0.278, 0.159	-0.034	-0.198, 0.131
3	GGCCC/GCTGC	14	35	-0.220	-0.410, -0.029	-0.139	-0.290, 0.012
3	GGCCC/GGCCC	36	76	-0.146	-0.296, 0.004	-0.107	-0.236, 0.021
3	GGCCT/AGTGC	6	16	-0.245	-0.502, 0.013	-0.119	-0.298, 0.060
3	GGCCT/GCTCT	16	24	0.047	-0.166, 0.260	0.027	-0.135, 0.189
3	GGCCT/GCTGC	11	23	-0.141	-0.368, 0.086	-0.077	-0.246, 0.091
3	GGCCT/GGCCT	12	33	-0.256	-0.448, -0.064	-0.161	-0.313, -0.009
3	GGCCT/GGTGT	2	7	-0.334	-0.683, 0.015	-0.113	-0.316, 0.090
3	GGCCC/GGCCT	57	92	REF		REF	

* N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.

† N = 382, excludes 9 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Group	Diplotypes	N (any fibroids)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
4	CA/TA, Rare	9	16	0.019	-0.236, 0.273	0.009	-0.169, 0.187
4	CA/CA	8	22	-0.180	-0.395, 0.034	-0.104	-0.267, 0.059
4	CG/CA	62	118	-0.018	-0.135, 0.099	-0.015	-0.121, 0.091
4	CG/TA	26	59	-0.103	-0.250, 0.044	-0.077	-0.203, 0.050
4	CG/CG	93	171	REF		REF	

* N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.

† N = 386, excludes 5 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Table A.20. Unadjusted prevalence differences of uterine fibroids (2+ cm) with *IGF-I* group-specific diplotypes among Caucasians*

Group†	Diplotypes	N (fibroids 2+ cm)‡	total N‡	Prevalence Differences‖	95% Confidence Intervals‖	Posterior Medians	95% Posterior Limits
1	A/G, A/A	40	114	0.018	-0.086, 0.121	0.015	-0.081, 0.111
1	G/G	91	273	REF		REF	

* N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.
† rs35767 represents entire group because it could not be combined with block 1 SNPs for diplotype estimation.
‡ N = 387, excludes 4 women missing SNP genotype and 8 women missing uterine fibroid status.
‖ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Group	Diplotypes	N (fibroids 2+ cm)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
2	Rare	3	10	-0.024	-0.321, 0.273	-0.010	-0.201, 0.181
2	AAT/ACT	2	10	-0.124	-0.387, 0.139	-0.059	-0.240, 0.122
2	AAT/GCC	8	22	0.040	-0.180, 0.259	0.022	-0.143, 0.187
2	AAT/GCT	16	50	-0.004	-0.161, 0.152	-0.003	-0.136, 0.130
2	ACT/GCC	10	26	0.061	-0.146, 0.267	0.036	-0.123, 0.195
2	ACT/GCT	13	34	0.058	-0.127, 0.244	0.038	-0.111, 0.187
2	GCC/GCC	11	28	0.069	-0.133, 0.270	0.042	-0.115, 0.199
2	GCT/GCT	32	98	0.002	-0.126, 0.131	0.002	-0.112, 0.116
2	GCT/GCC	35	108	REF		REF	

* N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.
† N = 386, excludes 5 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing uterine fibroid status.
‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Group	Diplotypes	N (fibroids 2+ cm)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
3	Rare	9	22	0.029	-0.199, 0.257	0.016	-0.153, 0.184
3	GCTCT/GCTGC	2	8	-0.130	-0.446, 0.186	-0.050	-0.246, 0.146
3	GGCCC/AGTGC	10	21	0.096	-0.140, 0.331	0.051	-0.121, 0.222
3	GGCCC/GCTCT	9	25	-0.020	-0.233, 0.192	-0.012	-0.174, 0.150
3	GGCCC/GCTGC	10	35	-0.095	-0.274, 0.085	-0.062	-0.208, 0.083
3	GGCCC/GGCCC	26	76	-0.038	-0.184, 0.107	-0.029	-0.154, 0.097
3	GGCCT/AGTGC	2	16	-0.255	-0.445, -0.065	-0.162	-0.313, -0.011
3	GGCCT/GCTCT	9	24	-0.005	-0.223, 0.212	-0.003	-0.167, 0.161
3	GGCCT/GCTGC	7	23	-0.076	-0.289, 0.137	-0.044	-0.206, 0.118
3	GGCCT/GGCCT	8	33	-0.138	-0.315, 0.039	-0.092	-0.236, 0.052
3	GGCCT/GGTGT	1	7	-0.238	-0.515, 0.040	-0.106	-0.292, 0.079
3	GGCCC/GGCCT	35	92	REF		REF	

* N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.

† N = 382, excludes 9 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Group	Diplotypes	N (fibroids 2+ cm)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
4	CA/TA, Rare	8	16	0.126	-0.130, 0.381	0.061	-0.117, 0.240
4	CA/CA	4	22	-0.192	-0.369, -0.016	-0.128	-0.273, 0.016
4	CG/CA	39	118	-0.044	-0.155, 0.068	-0.036	-0.138, 0.065
4	CG/TA	16	59	-0.103	-0.238, 0.032	-0.080	-0.198, 0.039
4	CG/CG	64	171	REF		REF	

* N = 399, excludes 3 women missing more than 50% of *IGF-I* SNPs included in diplotype estimation.

† N = 386, excludes 5 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 8 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Table A.21. Unadjusted prevalence differences of uterine fibroids (any size) with *IGF-I* group-specific diplotypes among African Americans*

Group	Diplotypes	N (any fibroids)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
1	Rare	6	9	-0.092	-0.406, 0.223	-0.036	-0.231, 0.160
1	AGG/AGG	81	101	0.044	-0.056, 0.143	0.038	-0.055, 0.130
1	AGG/GCG	34	51	-0.092	-0.235, 0.052	-0.069	-0.193, 0.056
1	AGG/GGA	22	30	-0.025	-0.195, 0.145	-0.017	-0.158, 0.124
1	GCG/GCG	6	8	-0.008	-0.315, 0.298	-0.003	-0.197, 0.190
1	GGG/GCG	32	55	-0.176	-0.321, -0.032	-0.132	-0.257, -0.007
1	GGG/GGA	19	24	0.033	-0.141, 0.207	0.023	-0.120, 0.165
1	GGG/GGG	75	106	-0.051	-0.157, 0.056	-0.043	-0.141, 0.055
1	AGG/GGG	138	182	REF		REF	

* N = 580, excludes 2 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 566, excludes 7 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 7 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Group	Diplotypes	N (any fibroids)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
2	Rare	28	38	0.025	-0.132, 0.183	0.018	-0.115, 0.152
2	ATCCC/TCTTT	11	13	0.135	-0.074, 0.344	0.079	-0.081, 0.240
2	ATTCC/ATCCC	28	37	0.045	-0.111, 0.202	0.033	-0.100, 0.165
2	ATTCC/ATTTT	7	10	-0.011	-0.305, 0.282	-0.005	-0.195, 0.185
2	ATTCC/TCTCC	4	8	-0.211	-0.565, 0.143	-0.070	-0.275, 0.134
2	ATTCC/TCTCT	13	16	0.101	-0.104, 0.306	0.061	-0.098, 0.219
2	ATTCC/TCTTT	88	119	0.028	-0.079, 0.135	0.024	-0.075, 0.122
2	ATTCC/TTTCT	23	34	-0.035	-0.208, 0.138	-0.024	-0.166, 0.119
2	ATTCT/ATCCC	4	6	-0.045	-0.429, 0.339	-0.013	-0.223, 0.196
2	ATTCT/ATTCC	32	40	0.089	-0.055, 0.232	0.067	-0.058, 0.191
2	ATTCT/TCTCT	5	6	0.122	-0.185, 0.429	0.049	-0.145, 0.242
2	ATTCT/TCTTT	14	20	-0.011	-0.225, 0.202	-0.007	-0.169, 0.156
2	ATTCT/TTTCT	7	8	0.164	-0.077, 0.404	0.085	-0.088, 0.258
2	TCTCT/TCTTT	6	10	-0.111	-0.424, 0.201	-0.044	-0.239, 0.152
2	TCTTT/TCTTT	27	31	0.160	0.021, 0.298	0.122	0.001, 0.243
2	TTTCT/TCTTT	6	14	-0.283	-0.552, -0.014	-0.131	-0.314, 0.052
2	ATTCC/ATTCC	106	149	REF		REF	

* N = 580, excludes 2 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 559, excludes 14 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 7 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Group	Diplotypes	N (any fibroids)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
3	Rare	5	9	-0.214	-0.549, 0.121	-0.077	-0.277, 0.124
3	AACC/AACC	6	8	-0.020	-0.331, 0.291	-0.008	-0.203, 0.187
3	AACC/AACT	5	7	-0.056	-0.400, 0.289	-0.019	-0.222, 0.183
3	AACC/AGGT	6	14	-0.341	-0.613, -0.069	-0.156	-0.340, 0.028
3	AACC/TACC	7	10	-0.070	-0.366, 0.226	-0.029	-0.220, 0.162
3	AACC/TACT	39	53	-0.034	-0.179, 0.110	-0.026	-0.151, 0.100
3	AACT/AGGT	3	7	-0.341	-0.717, 0.034	-0.105	-0.313, 0.103
3	AACT/TACT	7	13	-0.232	-0.515, 0.052	-0.101	-0.289, 0.086
3	AAGC/AACC	23	36	-0.131	-0.308, 0.046	-0.087	-0.232, 0.057
3	AAGC/AACT	6	9	-0.103	-0.422, 0.215	-0.039	-0.236, 0.157
3	AAGC/AAGC	29	35	0.059	-0.091, 0.208	0.043	-0.085, 0.172
3	AAGC/AGGT	18	24	-0.020	-0.212, 0.172	-0.013	-0.165, 0.140
3	AAGC/TACC	22	27	0.045	-0.123, 0.213	0.031	-0.109, 0.170
3	AAGC/TACT	64	86	-0.026	-0.150, 0.098	-0.021	-0.132, 0.090
3	AGGT/AGGT	8	10	0.030	-0.231, 0.291	0.014	-0.166, 0.195
3	AGGT/TACT	38	47	0.039	-0.101, 0.178	0.029	-0.092, 0.151
3	TACC/TACC	4	6	-0.103	-0.489, 0.283	-0.031	-0.240, 0.179
3	TACC/TACT	31	45	-0.081	-0.240, 0.077	-0.058	-0.192, 0.076
3	TACT/TACT	77	100	REF		REF	

* N = 580, excludes 2 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 546, excludes 27 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 7 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Group	Diplotypes	N (any fibroids)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
4	CA/CA	71	93	0.071	-0.034, 0.176	0.060	-0.037, 0.157
4	CA/TA	32	40	0.107	-0.030, 0.245	0.082	-0.038, 0.203
4	CG/CG	108	147	0.042	-0.051, 0.135	0.037	-0.050, 0.124
4	CG/TA	35	46	0.068	-0.069, 0.205	0.052	-0.068, 0.173
4	TA/TA	7	9	0.085	-0.193, 0.363	0.038	-0.148, 0.224
4	CA/CG	160	231	REF		REF	

* N = 580, excludes 2 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 566, excludes 7 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 7 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Table A.22. Unadjusted prevalence differences of uterine fibroids (2+ cm) with *IGF-I* group-specific diplotypes among African Americans*

Group	Diplotypes	N (fibroids 2+ cm)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
1	Rare	5	9	-0.043	-0.376, 0.289	-0.016	-0.215, 0.184
1	AGG/AGG	64	101	0.035	-0.083, 0.153	0.028	-0.078, 0.135
1	AGG/GCG	25	51	-0.109	-0.263, 0.046	-0.079	-0.210, 0.053
1	AGG/GGA	17	30	-0.032	-0.223, 0.159	-0.020	-0.172, 0.131
1	GCG/GCG	3	8	-0.224	-0.567, 0.119	-0.078	-0.280, 0.124
1	GGG/GCG	23	55	-0.181	-0.329, -0.032	-0.134	-0.261, -0.006
1	GGG/GGA	15	24	0.026	-0.180, 0.232	0.016	-0.144, 0.175
1	GGG/GGG	60	106	-0.033	-0.151, 0.085	-0.027	-0.134, 0.080
1	AGG/GGG	109	182	REF		REF	

* N = 580, excludes 2 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 566, excludes 7 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 7 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Group	Diplotypes	N (fibroids 2+ cm)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
2	Rare	23	38	0.062	-0.113, 0.236	0.041	-0.102, 0.185
2	ATCCC/TCTTT	8	13	0.072	-0.205, 0.348	0.032	-0.153, 0.218
2	ATTCC/ATCCC	16	37	-0.111	-0.290, 0.067	-0.074	-0.219, 0.072
2	ATTCC/ATTTT	7	10	0.156	-0.139, 0.451	0.065	-0.125, 0.256
2	ATTCC/TCTCC	4	8	-0.044	-0.399, 0.312	-0.014	-0.219, 0.190
2	ATTCC/TCTCT	9	16	0.019	-0.237, 0.275	0.009	-0.170, 0.188
2	ATTCC/TCTTT	68	119	0.028	-0.092, 0.147	0.023	-0.085, 0.131
2	ATTCC/TTTCT	19	34	0.015	-0.170, 0.200	0.010	-0.139, 0.159
2	ATTCT/ATCCC	3	6	-0.044	-0.452, 0.364	-0.012	-0.225, 0.201
2	ATTCT/ATTCC	24	40	0.056	-0.115, 0.228	0.038	-0.103, 0.180
2	ATTCT/TCTCT	5	6	0.290	-0.019, 0.598	0.115	-0.080, 0.309
2	ATTCT/TCTTT	11	20	0.006	-0.226, 0.239	0.003	-0.167, 0.174
2	ATTCT/TTTCT	6	8	0.206	-0.104, 0.517	0.081	-0.114, 0.276
2	TCTCT/TCTTT	5	10	-0.044	-0.364, 0.276	-0.017	-0.214, 0.180
2	TCTTT/TCTTT	21	31	0.134	-0.049, 0.317	0.087	-0.061, 0.235
2	TTTCT/TCTTT	6	14	-0.115	-0.386, 0.156	-0.053	-0.237, 0.131
2	ATTCC/ATTCC	81	149	REF		REF	

* N = 580, excludes 2 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 559, excludes 14 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 7 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Group	Diplotypes	N (fibroids 2+ cm)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
3	Rare	3	9	-0.207	-0.530, 0.116	-0.077	-0.275, 0.120
3	AACC/AACC	3	8	-0.165	-0.514, 0.184	-0.056	-0.259, 0.147
3	AACC/AACT	3	7	-0.111	-0.491, 0.268	-0.034	-0.242, 0.175
3	AACC/AGGT	4	14	-0.254	-0.510, 0.002	-0.124	-0.303, 0.055
3	AACC/TACC	6	10	0.060	-0.259, 0.379	0.023	-0.174, 0.220
3	AACC/TACT	34	53	0.102	-0.060, 0.263	0.072	-0.064, 0.207
3	AACT/AGGT	2	7	-0.254	-0.603, 0.094	-0.086	-0.290, 0.117
3	AACT/TACT	6	13	-0.078	-0.367, 0.210	-0.034	-0.223, 0.155
3	AAGC/AACC	18	36	-0.040	-0.230, 0.150	-0.025	-0.177, 0.126
3	AAGC/AACT	5	9	0.016	-0.323, 0.355	0.005	-0.196, 0.207
3	AAGC/AAGC	24	35	0.146	-0.036, 0.328	0.095	-0.052, 0.242
3	AAGC/AGGT	14	24	0.043	-0.177, 0.263	0.024	-0.141, 0.190
3	AAGC/TACC	18	27	0.127	-0.076, 0.330	0.076	-0.081, 0.234
3	AAGC/TACT	51	86	0.053	-0.090, 0.196	0.040	-0.084, 0.164
3	AGGT/AGGT	7	10	0.160	-0.140, 0.460	0.065	-0.127, 0.258
3	AGGT/TACT	30	47	0.098	-0.070, 0.267	0.068	-0.072, 0.207
3	TACC/TACC	4	6	0.127	-0.263, 0.516	0.037	-0.173, 0.247
3	TACC/TACT	22	45	-0.051	-0.227, 0.125	-0.034	-0.178, 0.110
3	TACT/TACT	54	100	REF		REF	

* N = 580, excludes 2 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 546, excludes 27 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 7 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.

Group	Diplotypes	N (fibroids 2+ cm)†	total N†	Prevalence Differences‡	95% Confidence Intervals‡	Posterior Medians	95% Posterior Limits
4	CA/CA	56	93	0.061	-0.057, 0.179	0.050	-0.057, 0.157
4	CA/TA	25	40	0.084	-0.079, 0.247	0.059	-0.078, 0.195
4	CG/CG	82	147	0.017	-0.086, 0.120	0.014	-0.081, 0.109
4	CG/TA	26	46	0.024	-0.133, 0.181	0.017	-0.116, 0.150
4	TA/TA	6	9	0.126	-0.189, 0.440	0.049	-0.147, 0.244
4	CA/CG	125	231	REF		REF	

* N = 580, excludes 2 women missing more than 50% of *IGFBP-3* SNPs included in diplotype estimation.

† N = 566, excludes 7 women with less than 90% posterior probability for best diplotype assignment in PHASE software and 7 women missing uterine fibroid status.

‡ Estimated in SASV9.1 using Poisson regression with robust standard errors.