January 2014

Genome-Wide Association Study Identifies Novel Ethnic-Specific Associations With Body Mass Index

Yasmmyn D. Salinas
Yale University, yasmmyn.salinas@yale.edu

GENOME-WIDE ASSOCIATION STUDY IDENTIFIES NOVEL ETHNIC-SPECIFIC ASSOCIATIONS WITH BODY MASS INDEX

By

Yasmmyn D. Salinas

A Thesis Presented to
The Faculty of the Yale School of Public Health
Yale University

In Candidacy for the Degree of
Master of Public Health

2014

ABSTRACT

Background. Obesity affects over one-third of the US population, and is a risk factor for various chronic conditions, including type 2 diabetes, heart disease, and stroke. The disease results from a combination of behavioral and environmental risk factors and genetic predisposition. To date, over 50 genetic polymorphisms have been associated with increased body mass index (BMI), but these associations explain only a small percentage of the heritable risk of obesity. Moreover, the majority of these associations have been identified in populations of European ancestry. We sought to identify novel associations with BMI and to evaluate their generalizability across ethnic groups, using subjects from the Multi-ethnic Study of Atherosclerosis (MESA).

Methods. Ethnic-specific genome-wide association analyses were conducted to identify single nucleotide polymorphisms (SNPs) associated with BMI among 1,257 Hispanic, 705 Asian, 1,551 African American, and 2,416 Caucasian MESA participants. We compared and contrasted findings across ethnic groups, and accounted for potential differences in linkage disequilibrium patterns by examining the ± 500kb flanking regions of the top SNPs in all four ethnic groups.

Results. We identified one genome-wide significant association with BMI in Hispanic subjects: rs12253976 near *KLF6* ($p=6.88x10^{-09}$). The top SNPs in each of the other ethnic groups—rs9961691 near *GATA6* in Asians ($p=1.53x10^{-06}$), rs7092615 near *LYZL2* in African Americans ($p=2.26x10^{-07}$), and rs6866721 near *SEMA6A* in Caucasians ($p=9.23x10^{-08}$)—may also be of interest. Each of these SNPs showed no evidence of an association with BMI in the other ethnic groups.

Conclusion. We present one of the first GWAS to examine BMI-associated variants across ethnic groups in the same study. The existence of ethnic-specific associations with BMI highlights the need for future investigations in larger multiethnic cohorts. Discovery of further ethnic-specific BMI-associated loci may contribute to personalized obesity interventions.

Acknowledgements.

I would like to thank Dr. Andrew Dewan and Dr. Michael Bracken for their mentorship throughout this project. I would also like to thank Dr. Leyao Wang for her help with the EIGENSTRAT analyses. Finally, I want to thank those who participated in the MESA study, without whom the present study would not have been possible.

MESA and the MESA SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169 and CTSA UL1-RR-024156.

Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetric Genome-Wide Human SNP Array 6.0.

Table of Contents.

List of Tables.

List of Figures.

Introduction.

Obesity is one of the most pressing health problems in the United States (U.S.). It affects over 35.7% of adults [1] and 16.9% of children [2], placing them at higher risk of metabolic syndrome, type 2 diabetes, cardiovascular disease, and other chronic conditions [3, 4]. Obesity-related co-morbidities place a huge financial burden on the healthcare system and consequently on the U.S. economy: in 2010, the medical costs of treating people for overweight and obesity were estimated to be $72 billion and $198 billion, respectively [5]. Understanding the etiology of obesity and developing interventions to prevent obesity-related co-morbidities are therefore important public health concerns.

The etiology of obesity involves multiple interactions among behavioral, environmental, and genetic factors [6]. Though the current obesity epidemic is commonly attributed to lifestyle and environmental changes[7], it is also recognized that individuals respond differently to obesogenic environments and that those differences are driven by genetic variation. In fact, family and twin studies have shown that genetic factors explain 40-70% of the variation in common obesity [8, 9].

In the last few years, genome-wide association studies (GWAS) have increased our understanding of the heritable risk of obesity by identifying approximately 50 obesity-susceptibility loci [10]. However, further investigations are warranted for several reasons. First, the known obesity loci with the largest effect sizes—variants in the *FTO* and *MC4R* genes—only account for an estimated 2% of the variation in body mass index (BMI) [11]. Second, the majority of GWAS conducted to date have primarily focused on populations of European ancestry [12]. While GWAS of Asian populations have begun to emerge, African Americans and Hispanics continue to be underrepresented in these studies, and it is precisely these populations who are disproportionately affected by overweight and obesity in the U.S [1, 13].

Finally, most obesity GWAS investigations have focused on one ethnic group in isolation and infrequently attempt to compare and contrast findings across ethnic groups.

In order to gauge the clinical and public health implications of obesity-associated variants, it is not sufficient to simply replicate findings in other Caucasian populations; we need to evaluate whether these associations are generalizable to individuals of other ethnicities [14]. In addition, GWAS in multiethnic populations may reveal additional loci that are not readily detectable in Caucasians due to allele frequency and haplotype structure differences [15]. For these reasons, in the present study, we examined genetic associations with BMI—a correlate of obesity [16]—in the Multi-ethnic Study of Atherosclerosis (MESA), which includes individuals of four ethnic groups: Hispanic, Asian, African American, and Caucasian. We compared and contrasted findings across ethnic groups and evaluated whether variants in the *FTO* gene are associated with BMI to a similar extent across ethnic groups.

Methods.

Study population

The MESA study is a multicenter, prospective cohort study of the characteristics of subclinical cardiovascular disease and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease. Recruitment has been described in detail elsewhere [17]. Briefly, 6,814 men and women aged 45 to 84 years were recruited from six U.S. field centers from July 2000 to July 2002. The study collected self-identified racial/ethnic group data via a standard questionnaire. Each field center recruited an equal number of men and women from 2 or more ethnic groups, and there were overlapping ethnic groups among field centers to minimize confounding of race/ethnicity by site [17]. Genotype information was available for 6,361 of the participants: 1,449 Hispanics, 775 Asians, 1,611 African Americans, and 2,526 Caucasians.

For the present study, the MESA population was divided into four ethnic-specific samples consisting of individuals with both phenotype and genotype information available. Genotype and phenotype information were acquired from the National Center for Biotechnology Information's database of Genotypes and Phenotypes (NCBI dbGaP study accession: phs000209.v7.p2 MESA SNP Health Association Resource (SHARe)). To be included in the GWAS analyses, individuals had to meet our quality control thresholds (described below) and had to have complete data for all ethnic-specific regression model covariates.

Genotyping

Details of sample preparation and genotyping have been reported previously [17]. Briefly, DNA was extracted from blood sample buffy coat, and genotyping was conducted by Affymetrix Research Services Lab using the Affymetrix 6.0 SNP array. Samples were required to have a call rate of at least 95%.

Phenotypic Data

All measurements used in this study were obtained at the first MESA study visit. The primary outcome was BMI (kg/m^2), calculated from objective height and weight measurements collected by trained staff at the six field centers. All genotyped participants of each ethnic group had BMI information from this baseline visit.

Information regarding demographic and lifestyle characteristics and medical history was obtained through standardized questionnaires administered at baseline. Physical activity levels were also assessed at baseline using the MESA Typical Week Physical Activity Survey, adapted from the Cross-Cultural Activity Participation Study [18] to determine the time and frequency spent in various physical activities during a typical week in the past month. The following variables were examined in our study due to their previously reported associations with BMI:

sex; age at baseline interview (years); education (categorical: highest level of schooling completed, grouped into 9 categories ranging from no schooling to graduate and professional schooling) [19]; income (categorical: total gross income earned by all family members, grouped into 13 categories ranging from < $5,000 to more than $100,000) [20]; smoking (categorical: never, former, current smokers) [21]; arthritis (yes/no) [22]; diabetes (yes/no) [23]; and moderate and vigorous physical activity (continuous: met-min/week) [24].

Ethnic specific linear regression models

Ethnic-specific bivariate associations between BMI and age, sex, education, income, smoking, arthritis, diabetes, and moderate and vigorous physical activity were examined in SAS 9.3 (SAS Institute, Cary NC). Ethnic-specific multivariate linear regression models were then built, and the most parsimonious models were selected using a backwards selection method, eliminating covariates with p-values greater than 0.05 until the model adjusted-R^2 values were maximized. Age and sex were retained in the reduced models regardless of the statistical significance of their associations with BMI. Individuals with missing values for any covariate included in the final ethnic-specific models were removed from the analysis (n=15 Hispanics, n=18 Asians, n=29 African Americans, and n=92 Caucasians).

For these analyses, the 9 education categories were collapsed into three: (1) < 12 years of schooling, (2) 12-15 years of schooling (high school graduates and some college), and (3) $\geq$ 16 years of schooling (college graduates and professional/graduate school graduates). Income was examined as a binary predictor, collapsing the 13 categories into two: high and low income. For this purpose, median household income was averaged over the 2000-2002 recruitment period ($41,448). Income categories $40-49,000 and above were defined as high income. Moderate and vigorous physical activity was examined as a categorical predictor since its distribution was strongly right-skewed, with possible outliers, and an F-test suggested that its association with

BMI is not linear ($p<0.05$). Additionally, a Kruskal-Wallis test revealed that the distribution of moderate and vigorous physical activity differed significantly across the four ethnic-groups ($p<0.0001$), so ethnic-specific cut points were chosen to divide participants into quartiles. Quartiles for moderate and vigorous physical activity were: 0 to 1665, 1666 to 4072.5, 4073 to 8280, and 8281 to 45060 met-min/week for Hispanics; 0 to 1305, 1306 to 2580, 2581 to 4770, and 4771 to 30240 met-min/week for Asians; 0 to 2115, 2116 to 4560, 4561 to 8625, and 8626 to 103320 met-min/week for African Americans; and 0 to 2280, 2281 to 4207.5, 4208 to 7220, and 7221 to 56550 met-min/week for Caucasians.

SNP Analysis

Ethnic-specific genome-wide SNP analyses were performed using the PLINK software package (http://pngu.mgh.harvard.edu/purcell/plink/ [25]) . A total of 909,622 SNPs were genotyped in individuals of each ethnic group. SNPs were excluded based on low call rate (<98%), low minor allele frequency (MAF <0.01), and significant deviation from Hardy-Weinberg equilibrium (p-value $\leq 5.5 \times 10^{-8}$).

Genetic quality control procedures included assessments for cryptic relatedness and population stratification. Cryptic relatedness between subjects in each ethnic group was examined by pair-wise identity-by-descent (IBD) estimation in PLINK. Pairs showing $\hat{\pi}$ (estimated proportion of genomic variation shared IBD) > 0.2 were inspected. One subject from each family was included, and 52 Asians, 177 Hispanics, 31 African Americans, and 18 Caucasians were excluded from downstream SNP analyses. Population stratification was assessed by performing principal components analyses using EIGENSTRAT version 3.0 (http://genepath.med.harvard.edu/~reich/Software.htm).

Linear regression analyses were performed within PLINK to test the association between individual SNPs and BMI, with initial adjustment for the top two principal components (PCs)

identified from the EIGENSTRAT analyses, additional adjustment for age and sex, and complete adjustment for all ethnic-specific model covariates. Bonferroni-corrected genome-wide significance thresholds were applied to each ethnic group to maintain an overall study α of 0.05. Because the Bonferroni correction is conservative, only tested markers were included in the calculation of the corrected significance thresholds. Therefore the significance thresholds used were: 5.86×10^{-8} (0.05/853,278) for Hispanics; 7.31×10^{-8} (0.05/683,998) for Asians; 5.73×10^{-8} (0.05/871,948) for African Americans; and 6.67×10^{-8} (0.05/749,659) for Caucasians.

After identifying the top SNP candidates ($p < 5.5 \times 10^{-6}$) in each ethnic group, we evaluated the generalizability of these associations to other ethnic groups. To account for potential differences in linkage disequilibrium (LD) patterns across ethnic groups, we examined the ± 500kb flanking regions of each of the top SNPs in all four ethnic groups.

Results.

Our study included 1,257 Hispanics, 705 Asians, 1,551 African Americans, and 2,416 Caucasian subjects. The demographic, behavioral, and clinical characteristics of these MESA participants according to race/ethnicity are displayed in **Table 1**. African Americans had higher BMI (mean 30.13kg/m^2 ± 5.86) than Hispanics (mean 29.29kg/m^2 ± 5.10), Caucasians (mean 27.74kg/m^2 ± 5.07), and Asians (mean 24.03 kg/m^2 ± 3.30). African Americans also reported greater physical activity and were more likely to have arthritis than participants of all other ethnic groups. Caucasians were the most likely to have ever smoked cigarettes. Hispanics were the most likely to have low income, not to have graduated high school, and to report having diabetes.

Ethnic-specific regression models

The unadjusted associations between BMI and participant demographic, behavioral, and clinical risk factors are summarized in **Table 2**. Arthritis and diabetes were positively associated with BMI across all ethnic groups. Education and income were significant predictors of BMI only in Caucasians, where the mean BMI difference between participants who completed higher education and those who did not finish high school was -1.569kg/m^2 (p=0.001), and the mean BMI difference between low and high income participants was 0.712 kg/m^2 (p=0.001). Physical activity was associated with BMI in Caucasians, with a mean difference in BMI of -0.952 kg/m^2 (p=0.001) between those in the highest and lowest quartile. Smoking was only a significant predictor of BMI in African Americans, and the mean BMI difference between "ever" and "never" smokers was -0.641 kg/m^2 (p=0.0320). After adjustment for the independent effects of these factors in multivariate models, the effect of smoking became significant in Hispanics and Caucasians, and the overall effect of education became significant in Asians and African Americans. The final ethnic specific models are displayed in **Table 3**.

Population stratification

EIGENSTRAT analyses found only minor evidence of population stratification in Hispanics, Asians, African Americans, and Caucasians. PC 1 and PC 2 explained the majority of the variation in allelic frequencies within each of these populations and were adjusted for in linear regression models. As shown in **Figure 1**, across the four ethnic groups, the observed p-values did not deviate significantly from the expected p-values under the null hypothesis. Adjustment for population stratification did not greatly influence the magnitude of our observed p-values.

SNP Analysis

Following SNP quality control (**Table 4**), there were 683,998 SNPs in Asians, 853,278 in Hispanics, 871,948 in African Americans, and 749,659 in Caucasians. The top SNPs ($p < 5.5 \times 10^{-6}$) in each ethnic group are displayed in **Table 5**. Following linear regression analyses adjusted for all the covariates and assuming an additive mode of inheritance, the most significant SNP in Hispanic subjects was rs12253976 near *KLF6* (beta=5.542 kg/m^2 per allele, 95% CI: 3.680 to 7.404; $p=6.88 \times 10^{-9}$). This SNP was the only variant in the ethnic-specific analyses that reached genome-wide significance after Bonferroni adjustment for multiple comparisons. In Asian subjects, the most significant SNP was rs9961691 near *GATA6* (beta= -0.994 kg/m^2 per allele, 95% CI: -1.396 to -0.592; $p=1.53 \times 10^{-6}$). In African Americans, the most significant SNP was rs7092615 near *LYZL2* (beta = 1.077 kg/m^2 per allele, 95% CI: 0.671 to 1.483; $p=2.41 \times 10^{-7}$). In Caucasians, the most significant SNP was rs6866721 near *SEMA6A* (beta=0.764 kg/m^2 per allele, 95% CI: 0.484 to 1.043; $p=9.23 \times 10^{-8}$). As shown in **Table 6**, the strength of the associations and estimated per-allele effect sizes for rs12253976 in Hispanic subjects and rs9961691 in Asian subjects were relatively consistent across the unadjusted, minimally adjusted, and fully adjusted models. For rs7092615 in African Americans and rs6866721 in Caucasians, the associations were strengthened after adjustment for age and sex and complete adjustment for ethnic-specific model covariates.

Regional plots visualizing association results for the top SNP in each ethnic group and their respective ± 500 kb flanking region SNPs are shown in **Figures 2-5.** The chromosome 5 region of Caucasian subjects contains a set of SNPs with low p-values ($<10^{-5}$) and in strong linkage disequilibrium ($R^2 > 0.8$) with rs6866721. This region includes top SNP candidates rs1672492, rs1672491, and rs7704264 (**Table 5**). A similar pattern of association can be seen in the chromosome 18 region flanking rs9961691 in Asian subjects. In contrast, the chromosome 10 region flanking top SNP candidate rs7092615 in African Americans revealed only one other variant (rs156710) with low p-value ($<10^{-5}$) and in perfect LD with rs7092615 ($R^2 =1.0$).

Finally, the plot for top SNP candidate rs12253976 in Hispanics did not show evidence of association for SNPs in the chromosome 10 region flanking rs12253976.

Ethnic-specificity of the associations

The associations between the top SNP candidates (**Table 5**) and BMI were generally ethnic-specific. Two notable exceptions were rs12255372 (T/G) and rs7926805 (C/T). rs12255372 was associated with lower BMI in both Hispanics and African Americans. The association signal was stronger in Hispanics ($p=2.25x10^{-6}$) than in African Americans, where it was only nominally significant ($p = 0.008$), but the estimated per-allele effect sizes in each group did not differ significantly (beta = -1.117, 95% CI: -1.578 to -0.657 in Hispanics and beta = -0.594, 95% CI: -1.035 to -0.162 in African Americans). rs7926805 was associated with lower BMI in both Caucasians and African Americans. The association signal was stronger in Caucasians ($p = 2.35x10^{-6}$) than in African Americans, where it was only nominally significant ($p = 0.018$), but the estimated effect sizes in each group did not differ significantly (beta= -0.766, 95% CI: -1.084 to -0.450 in Caucasians and beta = -0.477, 95% CI: -0.871 to -0.084 in African Americans).

Association with FTO in Caucasians

The association between *FTO* SNP rs9939609 and BMI in our sample of MESA Caucasians approached the nominal significance level of 0.05 ($p=0.057$ in linear regression models adjusted for ethnic-specific model covariates). As shown in **Table 7**, this SNP did not appear to have a significant effect on the BMI of the other ethnic groups. Our estimated per-allele effect size of 0.279 kg/m^2 (95% CI: -0.008 to 0.566) in Caucasians is within the range of what has been reported for this SNP in previous studies of European populations [26]. In *ad hoc* analyses (data not shown), we investigated whether the relatively low association signal was due

to the older age of our participants (mean age in Caucasians = 62.54yrs ± 10.22), since a recent meta-analysis [27] suggested that rs9939609 might have greater effects on body weight in younger adults relative to older adults. However, we found no significant evidence of heterogeneity of the associations between *FTO* and BMI in our Caucasian population across age quartiles (p=0.660).

Discussion.

The most significant SNP associated with BMI in our study was rs12253976 (G/T), identified in Hispanic subjects. rs12253976 is a rare SNP whose minor allele frequency (MAF) has not yet been reported for Hispanic populations included in the HapMap project. dbSNP reports a MAF of 2.2% for the global 1000 Genome phase I population, and the HapMap project reports MAFs ranging from 7.3% to 11.9% in populations of African ancestry. HapMap also reports that the SNP is monomorphic in a Han Chinese population (CHB) and in a Utah population of Northern and Western European ancestry (CEU). This data corroborates what we observed in the present study. In our MESA population, rs12253976 was most common in African Americans (MAF = 6.7%), was extremely rare in Caucasians (MAF= 0.1%), and was monomorphic in Asians. In Hispanics, the MAF was 1.1%, and the association between this SNP and BMI was driven by only 28 heterozygous individuals with average BMI of 34.70 ± 7.59 kg/m^2 (**Table 8**). On average, these heterozygotes were 5.54 BMI units heavier than those homozygous for the ancestral "T" allele. Given the lack of data for rs12253976 in other Hispanic populations, we cannot discard the possibility that some of these heterozygotes may be a product of genotyping error.

Regional association plots in Hispanics did not show evidence of association for SNPs in the ± 500kb flanking region of rs12253976 (**Figure 2**). While this may be indicative of a spurious association with BMI, we cannot abandon the possibility that rs12253976 may be in

weak linkage disequilibrium (LD) with nearby genetic markers. LD plots for this region in a HapMap African American population (data not shown) support this hypothesis. However, we cannot confirm whether a similar LD pattern exists in Hispanics, given the absence of data for this SNP in the Hispanic HapMap population.

The closest gene to rs12253976 is *KLF6*. To our knowledge, no GWAS has previously reported an association between SNPs in or near *KLF6* and any obesity-related trait. However, two family linkage studies have provided evidence for linkage of BMI to this region of chromosome 10 (10p15.1). Chagnon et al. [28] found suggestive evidence (LOD Score = 2.3-2.7, $p<0.001$) of linkage to markers D10S1435 and D10S189 in 10p15.3 and 10p15.1, respectively, using data from 522 Caucasian subjects (192 parents and 330 offspring from 99 families) in the HERITAGE Family Study cohort. Similarly, Lindsay et al. [29] reported moderate linkage of BMI to marker D10S189 (LOD Score= 1.7) in a sample of 1,338 Pima Indians (consisting of 332 nuclear families and 112 extended pedigrees). rs12253976 lies between markers D10S1435 and D10S189.

KLF6 encodes a transcriptional regulator that contains zinc-finger motifs. KLF6 has been shown to accelerate adipocyte differentiation by repressing the expression of adipogenesis inhibitors [30]. In addition, it promotes the transcription of adipocyte differentiation regulators, such as PPARγ [31]. PPARγ is considered the master regulator of adipogenesis [32], and the *PPARγ* gene has been linked to the development of obesity in numerous studies [33, 34]. Therefore, there exists a biologically plausible explanation for the association between variants near *KLF6* and obesity-related traits. Nevertheless, we cannot rule out a spurious association. Though we have employed rigorous QC procedures (see Methods), this is a secondary data analysis, so we do not have the ability to re-genotype our subjects.

The second most significant association observed in our study was for rs6866721 (C/A), identified in Caucasian subjects. rs6866721 is a relatively common variant, with a MAF of

43.7% in our Caucasian population, which approximates what has been reported for the HapMap population of European ancestry (CEU; MAF=38.9%). Regional association plots showed a strong association signal for SNPs in the vicinity of rs6866721 (**Figure 3**), including SNPs rs7704264, rs1672492, and rs1672491 (**Table 5**), which are in strong LD ($R^2 > 0.8$) with rs6866721. The association with rs6866721 was borderline significant at our strict Bonferroni-adjusted threshold of 6.67×10^{-8} for Caucasians. Individuals carrying one copy of the minor "C" allele were, on average, 0.764 BMI units heavier than those homozygous for the ancestral "A" allele.

To our knowledge, no other GWAS has directly implicated rs6866721 in the pathogenesis of obesity. However, the locus containing rs6866721 and its flanking SNPs (5q23.1) has been linked to obesity-related traits in family studies. Chen et al. [35] found suggestive linkage (LOD score 1.5-2.3, p= 0.0006-0.0043) of BMI between markers D5S1505 (5q23.1) and D5S1453 (5q21.1) in a study of 782 Caucasian siblings from 342 Louisiana families. A weak linkage of 5q23.1 to BMI (LOD score = 1.5) was also reported in the NHLBI Family Heart Study [36]. rs6866721 lies between markers D5S1453 and D5S1505.

The closest gene to rs6866721 is *SEMA6A*, which encodes a trans-membrane domain that plays an important role in cellular signaling and axon guidance [37]. This gene has not been implicated as a key player in the pathogenesis of obesity. However, downstream of rs6866721, there are also multiple binding sites for nuclear factor (NF)-κB, a primary regulator of inflammatory responses [38] whose activity regulates lypolysis in adipose tissue [39]. Downregulation of NF-κB lypolytic pathways has been associated with obesity in animal models [40]. Given the compelling evidence for locus 5q23.1, we hypothesize that our strong association signal for rs6866721 and its surrounding SNPs (**Figure 3**) might be due to variation in one of these NF-κB regulatory regions. Future studies are needed to validate our finding for rs6866721 and to further dissect locus 5q23.1 to identify a possible causal variant.

We also observed an association between BMI and rs7092615 (T/C) in African Americans. rs7092615 is a relatively common variant: a MAF of 37.5% in our African American population approximates what HapMap reports for its population of African Ancestry in the United States (ASW; MAF= 42.1%). Regional association plots revealed only one other variant (rs1567101, see **Table 5**) associated with BMI in the region containing rs7092615 (**Figure 4**). rs1567101 is in perfect LD with rs7092615 (R^2 =1.0), and LD plots for this region in the HapMap ASW population revealed that no other SNPs are in moderate to strong LD with these two SNPs (data not shown), providing supporting evidence for our findings.

The closest gene to rs7092615 is *LYZL2*, which encodes a lysozyme-like protein that plays a role in the immune response. However, rs7092615 is not in LD with markers within this gene. Therefore, genetic variation in *LYZL2* is not likely to be responsible for the association observed in this study. rs7092615 is also unlikely to be regulatory, since there are no predicted transcription factor binding sites in the small LD block containing this SNP. Nonetheless, we cannot discard the possibility that this variant may still play an important functional role. For instance, a recent study showed that obesity-associated non-coding sequences within the *FTO* gene are functionally connected, at megabase distances, with the homeobox gene *IRX3* [41]. The obesity-associated *FTO* region does not affect *FTO* expression, but actually interacts directly with the promoters of *IRX3* to affect *IRX3* expression in the human brain [41]. Currently, we have no evidence to support that rs7092615 plays such a role in the regulation of any obesity-related gene. For this reason, the association in this study must be interpreted with caution.

We also observed an association with rs9961691 (C/G) in Asian subjects. rs9961691 is unique among our top 4 SNP candidates because its minor allele appears to decrease BMI, while the ancestral allele appears to increase it. The MAF of 20.3% for this variant in our Asian population approximates what HapMap reports for its Chinese Han population (CHB; MAF=17.9%). Regional association plots revealed a moderate association signal for variants

flanking rs7092615 (**Figure 5**). Of these variants, the most strongly associated with BMI were rs7231159 and rs9950004 (**Table 5**), which are in high LD ($R^2 > 0.9$) with rs7092615. LD plots for this region in the HapMap CHB population (data not shown) revealed a pattern of inheritance that closely resembles the association pattern seen in **Figure 5**, providing strong supporting evidence for our findings.

To our knowledge, this study is the first to report the association of rs9961691 with BMI. The gene closest to rs9961691 is *GATA6*, which encodes a zinc finger transcription factor that is involved in terminal differentiation and cell proliferation [42], and whose function has not been directly linked to obesity. However, we hypothesize that rs9961691 may be a regulatory SNP, as it lies in a PAX5 binding site. PAX5 is a paired box transcription factor that is a key player in early development [43]. The *PAX5* gene has been previously associated with BMI and total fat mass in previous GWAS [44-46], and studies in the *PAX5* knockout mouse have suggested an important role for PAX5 in driving the phenotypic outcomes of diet-induced obesity, in terms of adipose burden, skeletal quality, and the balance of the immune system [47]. Though the association with rs9961691 in our study did not reach genome-wide significance, the biologically relevant role of PAX5 in the pathogenesis of obesity warrants further investigation in this genomic locus.

Our study detected a marginal association with the *FTO* obesity-associated variant rs9939609 only in Caucasian subjects. Though not genome-wide significant, the rs9939609 association in our Caucasian population is consistent with prior evidence from over 15 GWAS conducted in populations of European ancestry [12]. In our study, individuals with one copy of the minor "A" allele were, on average, 0.279 BMI units heavier than those homozygous for the ancestral "T" allele.

The positive relationship between rs9939609 and BMI was not generalizable to subjects of other ethnic groups. In Asians, our data suggests that this SNP may have a reverse effect on

BMI, though this effect was not statistically significant (**Table 7**). This finding adds to the set of inconsistent findings reported in other Asian populations [12]; earlier GWAS in Asian populations reported that *FTO* variants did not have a significant effect on obesity-related traits, while at least two newer studies have yielded positive findings [48, 49].

We did not observe an effect of *FTO* on the BMI of either African Americans or Hispanics. In African Americans, this is consistent with data from three previous GWAS [50-52]. However, our null finding in Hispanics contradicts the evidence available from previous candidate-gene studies of rs9939609 [10].

The conflicting results for Asians and Hispanics could be due to a number of reasons, including differential adjustment for confounders, varying degrees of statistical power across studies, and inherent differences in the study populations that may influence BMI and obesity risk. Because the MAFs for rs9939609 in our four MESA ethnic groups approximate those reported for Caucasian, Hispanic, Asian, and African American populations in the HapMap project (**Table 7**), we have evidence to suggest that our MESA population is representative of these ethnic populations. Therefore, we hypothesized that the null findings in Asians and Hispanics, as well as the relatively weak association signal in Caucasian subjects, may be due to low statistical power in our study. *Ad hoc* power calculations revealed that in Asians, a variant like rs9939609 with a MAF of 12.9% needed to have an effect size of 1.60 kg/m^2 to be detected in our study with 80% power. In Hispanics, a variant with a MAF of 30.9% needed to have an effect size of 1.40 kg/m^2 to be detected in our study with 80% power. In Caucasians, a variant with a MAF of 40.8% needed to have an effect size of 0.95 kg/m^2 to be detected with 80% power. Therefore, our study was underpowered to detect the effects of *FTO* rs9939609.

As observed for *FTO*, the associations detected between the top SNP candidate in each ethnic group (**Table 6**) and BMI were ethnic-specific. To discard the possibility that these findings were due to differences in LD patterns across ethnic groups, we examined the ± 500kb

flanking regions of each of these SNPs across all four ethnic groups (data not shown). Regional association plots confirmed that no variants in the vicinity of these top SNP candidates were associated with BMI in other ethnic groups.

Our study is not the first to suggest the ethnic specificity of genetic associations with obesity-related traits. For instance, variants in the intronic region of the *SIM1* gene have been strongly associated with BMI and obesity risk in Pima Indians, but not in a French European population [53]. Similarly, a functional coding variant (W64R) in the *ADRB3* gene was associated with BMI in East Asian subjects but not in European subjects in a large meta-analysis of 44,833 subjects [54]. The existence of ethnic-specific genetic associations with obesity suggests that parts of a common obesity pathway may be activated differently across ethnic groups.

Our study is one of the first multi-ethnic GWAS attempting to compare and contrast findings across ethnic groups. Conducting ethnic-specific genome-wide scans allowed us to identify four novel associations with BMI. Stratifying the MESA population by race/ethnicity, however, also significantly limited our statistical power to detect variants with small effect sizes. Low statistical power may explain why only one of the associations in our study reached genome-wide significance, why we detected fewer and weaker associations in Asian subjects, and why we were unable to replicate associations with previously identified candidate genes for obesity such as *MC4R* and *BDNF*.

Besides low power, our study had other limitations. First, we acknowledge that BMI is not the most precise measure of obesity [55]. Obesity indicates an excess of adipose tissue, not an excess of body weight. Future analyses will include measures of waist-to-hip ratio, which has been shown to more adequately represent excess body fat [56]. Second, the older age of the MESA participants may have limited our ability to detect genetic associations with BMI. The distillation of the genetic component of some complex traits is easier in children, where the

relative environmental exposure time is far less [12]. It follows that studies focusing on *FTO* obesity-associated variants have suggested that most of the effect of genetic variants on BMI gain occurs during childhood, adolescence and young adulthood [57-59]. In *ad hoc* analyses for *FTO* rs9939609*,* however, our study found that age did not act as an effect modifier of the association between *FTO* and BMI in our Caucasian population. Nevertheless, it is possible that some other existing associations with BMI were masked in our study population.

Notwithstanding these limitations, our study was able to identify four novel ethnic-specific associations with BMI: rs12253976 in locus 10p15.1, rs6866721 in locus 5q23.1, rs7092615 in locus 10p11.23, and rs9961691 in locus 18q11.2. The average per-allele effects of these variants were greater than those reported for all previously identified BMI-associated variants, including those in the *FTO* gene, even after adjustment for various demographic and environmental determinants of BMI. The validity and functional significance of these novel associations needs further investigation. If replicated in larger multi-ethnic cohorts and meta-analyses, these top SNP candidates may be useful in improving current obesity-risk prediction models and may even encourage the derivation of ethnic-specific risk models for use in culturally tailored obesity interventions.

References.

1. Flegal KM, Carroll MD, Kit BK, Ogden CL: **Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010**. *JAMA: the journal of the American Medical Association* 2012, **307**(5):491-497.
2. Ogden CL, Carroll MD, Kit BK, Flegal KM: **Prevalence of obesity and trends in body mass index among US children and adolescents, 1999-2010**. *JAMA: the journal of the American Medical Association* 2012, **307**(5):483-490.
3. Bray GA: **Medical consequences of obesity**. *Journal of Clinical Endocrinology & Metabolism* 2004, **89**(6):2583-2589.
4. Adams KF, Schatzkin A, Harris TB, Kipnis V, Mouw T, Ballard-Barbash R, Hollenbeck A, Leitzmann MF: **Overweight, obesity, and mortality in a large prospective cohort of persons 50 to 71 years old**. *New England Journal of Medicine* 2006, **355**(8):763-778.
5. Behan DF, Cox SH, Lin Y, Pai J, Pedersen HW, Yi M: **Obesity and its relation to mortality and morbidity costs**. *Society of Actuaries Consulted on August* 2010, **5**:2011.
6. Grundy SM: **Multifactorial causation of obesity: implications for prevention**. *The American journal of clinical nutrition* 1998, **67**(3):563S-572S.
7. Rennie KL, Johnson L, Jebb SA: **Behavioural determinants of obesity**. *Best Practice & Research Clinical Endocrinology & Metabolism* 2005, **19**(3):343-358.
8. Maes HH, Neale MC, Eaves LJ: **Genetic and environmental factors in relative body weight and human adiposity**. *Behavior genetics* 1997, **27**(4):325-351.
9. Clément K, Sorensen TI: **Obesity: genomics and postgenomics**: CRC Press; 2010.
10. Jacobsson JA, Schiöth H, Fredriksson R: **The impact of intronic single nucleotide polymorphisms and ethnic diversity for studies on the obesity gene FTO**. *Obesity Reviews* 2012, **13**(12):1096-1109.
11. Bogardus C: **Missing heritability and GWAS utility**. *Obesity* 2009, **17**(2):209-210.
12. Xia Q, Grant SF: **The genetics of human obesity**. *Annals of the New York Academy of Sciences* 2013.
13. Pan L, Galuska D, Sherry B, Hunter A, Rutledge G, Dietz W, Balluz L: **Differences in prevalence of obesity among black, white, and hispanic adults-United States, 2006-2008**. *Morbidity and Mortality Weekly Report* 2009, **58**(27):740-744.
14. XI B, MI J: **Genome-wide Association Studies of Common Obesity: Now and Future**. *Biomedical and Environmental Sciences* 2013, **26**(10):787-791.
15. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M: **The structure of haplotype blocks in the human genome**. *Science* 2002, **296**(5576):2225-2229.
16. Daniels SR, Khoury PR, Morrison JA: **The utility of body mass index as a measure of body fatness in children and adolescents: differences by race and gender**. *Pediatrics* 1997, **99**(6):804-807.
17. Bild DE, Bluemke DA, Burke GL, Detrano R, Roux AVD, Folsom AR, Greenland P, JacobsJr DR, Kronmal R, Liu K: **Multi-ethnic study of atherosclerosis: objectives and design**. *American Journal of Epidemiology* 2002, **156**(9):871-881.
18. Ainsworth BE, Irwin ML, Addy CL, Whitt MC, Stolarczyk LM: **Moderate physical activity patterns of minority women: the Cross-Cultural Activity Participation Study**. *Journal of women's health & gender-based medicine* 1999, **8**(6):805-813.
19. Molarius A, Seidell JC, Sans S, Tuomilehto J, Kuulasmaa K: **Educational level, relative body weight, and changes in their association over 10 years: an international**

perspective from the WHO MONICA Project. *American Journal of Public Health* 2000, **90**(8):1260-1268.
20. Chang VW, Lauderdale DS: **Income disparities in body mass index and obesity in the United States, 1971-2002**. *Archives of Internal Medicine* 2005, **165**(18):2122-2128.
21. Albanes D, Jones DY, Micozzi MS, Mattson ME: **Associations between smoking and body weight in the US population: analysis of NHANES II**. *American journal of public health* 1987, **77**(4):439-444.
22. Davison KK, Ford ES, Cogswell ME, Dietz WH: **Percentage of body fat and body mass index are associated with mobility limitations in people aged 70 and older from NHANES III**. *Journal of the American Geriatrics Society* 2002, **50**(11):1802-1809.
23. Meigs JB, Wilson PW, Fox CS, Vasan RS, Nathan DM, Sullivan LM, D'Agostino RB: **Body mass index, metabolic syndrome, and risk of type 2 diabetes or cardiovascular disease**. *Journal of Clinical Endocrinology & Metabolism* 2006, **91**(8):2906-2912.
24. Dipietro L: **Physical activity, body weight, and adiposity: an epidemiologic perspective**. *Exercise and sport sciences reviews* 1994, **23**:275-303.
25. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ: **PLINK: a tool set for whole-genome association and population-based linkage analyses**. *The American Journal of Human Genetics* 2007, **81**(3):559-575.
26. Peng S, Zhu Y, Xu F, Ren X, Li X, Lai M: **FTO gene polymorphisms and obesity risk: a meta-analysis**. *BMC medicine* 2011, **9**(1):71.
27. Graff M, Gordon-Larsen P, Lim U, Fowke JH, Love S-A, Fesinmeyer M, Wilkens LR, Vertilus S, Ritchie MD, Prentice RL: **The Influence of Obesity-Related Single Nucleotide Polymorphisms on BMI Across the Life Course The PAGE Study**. *Diabetes* 2013, **62**(5):1763-1767.
28. Chagnon YC, Rice T, Pérusse L, Borecki IB, Ho-Kim M-A, Lacaille M, Paré C, Bouchard L, Gagnon J, Leon AS: **Genomic scan for genes affecting body composition before and after training in Caucasians from HERITAGE**. *Journal of Applied Physiology* 2001, **90**(5):1777-1787.
29. Lindsay RS, Kobes S, Knowler WC, Bennett PH, Hanson RL: **Genome-wide linkage analysis assessing parent-of-origin effects in the inheritance of type 2 diabetes and BMI in Pima Indians**. *Diabetes* 2001, **50**(12):2850-2857.
30. Rosen ED, MacDougald OA: **Adipocyte differentiation from the inside out**. *Nature Reviews Molecular Cell Biology* 2006, **7**(12):885-896.
31. Li D, Yea S, Li S, Chen Z, Narla G, Banck M, Laborda J, Tan S, Friedman JM, Friedman SL: **Krüppel-like factor-6 promotes preadipocyte differentiation through histone deacetylase 3-dependent repression of DLK1**. *Journal of Biological Chemistry* 2005, **280**(29):26941-26952.
32. Rosen ED, Hsu C-H, Wang X, Sakai S, Freeman MW, Gonzalez FJ, Spiegelman BM: **C/EBPα induces adipogenesis through PPARγ: a unified pathway**. *Genes & development* 2002, **16**(1):22-26.
33. Ristow M, Müller-Wieland D, Pfeiffer A, Krone W, Kahn CR: **Obesity associated with a mutation in a genetic regulator of adipocyte differentiation**. *New England Journal of Medicine* 1998, **339**(14):953-959.

34. Masud S, Ye S: **Effect of the peroxisome proliferator activated receptor-γ gene Pro12Ala variant on body mass index: a meta-analysis**. *Journal of medical genetics* 2003, **40**(10):773-780.
35. Chen W, Li S, Cook N, Rosner B, Srinivasan S, Boerwinkle E, Berenson G: **An autosomal genome scan for loci influencing longitudinal burden of body mass index from childhood to young adulthood in white sibships: The Bogalusa Heart Study**. *International journal of obesity* 2004, **28**(4):462-469.
36. Feitosa MF, Borecki IB, Rich SS, Arnett DK, Sholinsky P, Myers RH, Leppert M, Province MA: **Quantitative-trait loci influencing body-mass index reside on chromosomes 7 and 13: the National Heart, Lung, and Blood Institute Family Heart Study**. *The American Journal of Human Genetics* 2002, **70**(1):72-82.
37. Nakamura F, Kalb RG, Strittmatter SM: **Molecular basis of semaphorin-mediated axon guidance**. *Journal of neurobiology* 2000, **44**(2):219-229.
38. Ghosh S, May MJ, Kopp EB: **NF-κB and Rel proteins: evolutionarily conserved mediators of immune responses**. *Annual review of immunology* 1998, **16**(1):225-260.
39. Laurencikiene J, Van Harmelen V, Nordström EA, Dicker A, Blomqvist L, Näslund E, Langin D, Arner P, Rydén M: **NF-κB is important for TNF-α-induced lipolysis in human adipocytes**. *Journal of lipid research* 2007, **48**(5):1069-1077.
40. Carlsen H, Haugen F, Zadelaar S, Kleemann R, Kooistra T, Drevon CA, Blomhoff R: **Diet-induced obesity increases NF-κB signaling in reporter mice**. *Genes & nutrition* 2009, **4**(3):215-222.
41. Smemo S, Tena JJ, Kim K-H, Gamazon ER, Sakabe NJ, Gómez-Marín C, Aneas I, Credidio FL, Sobreira DR, Wasserman NF: **Obesity-associated variants within FTO form long-range functional connections with IRX3**. *Nature* 2014.
42. Beuling E, Baffour–Awuah NYA, Stapleton KA, Aronson BE, Noah TK, Shroyer NF, Duncan SA, Fleet JC, Krasinski SD: **GATA factors regulate proliferation, differentiation, and gene expression in small intestine of mature mice**. *Gastroenterology* 2011, **140**(4):1219-1229. e1212.
43. Nutt SL, Urbanek P, Rolink A, Busslinger M: **Essential functions of Pax5 (BSAP) in pro-B cell development: difference between fetal and adult B lymphopoiesis and reduced V-to-DJ recombination at the IgH locus**. *Genes & development* 1997, **11**(4):476-491.
44. Fall T, Ingelsson E: **Genome-wide association studies of obesity and metabolic syndrome**. *Molecular and cellular endocrinology* 2014, **382**(1):740-757.
45. Melka MG, Bernard M, Mahboubi A, Abrahamowicz M, Paterson AD, Syme C, Lourdusamy A, Schumann G, Leonard GT, Perron M: **Genome-wide scan for loci of adolescent obesity and their relationship with blood pressure**. *The Journal of Clinical Endocrinology & Metabolism* 2011, **97**(1):E145-E150.
46. Pei Y-F, Zhang L, Liu Y, Li J, Shen H, Liu Y-Z, Tian Q, He H, Wu S, Ran S: **Meta-analysis of genome-wide association data identifies novel susceptibility loci for obesity**. *Human molecular genetics* 2014, **23**(3):820-830.
47. Chan ME, Adler BJ, Green DE, Rubin CT: **Bone structure and B-cell populations, crippled by obesity, are partially rescued by brief daily exposure to low-magnitude mechanical signals**. *The FASEB Journal* 2012, **26**(12):4855-4863.

48. Wen W, Cho Y-S, Zheng W, Dorajoo R, Kato N, Qi L, Chen C-H, Delahanty RJ, Okada Y, Tabara Y: **Meta-analysis identifies common variants associated with body mass index in east Asians**. *Nature genetics* 2012, **44**(3):307-311.
49. Okada Y, Kubo M, Ohmiya H, Takahashi A, Kumasaka N, Hosono N, Maeda S, Wen W, Dorajoo R, Go MJ: **Common variants at CDKAL1 and KLF9 are associated with body mass index in east Asian populations**. *Nature genetics* 2012, **44**(3):302-306.
50. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A, Styrkarsdottir U, Gretarsdottir S, Thorlacius S, Jonsdottir I: **Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity**. *Nature genetics* 2009, **41**(1):18-24.
51. Kang SJ, Chiang CW, Palmer CD, Tayo BO, Lettre G, Butler JL, Hackett R, Adeyemo AA, Guiducci C, Berzins I: **Genome-wide association of anthropometric traits in African-and African-derived populations**. *Human molecular genetics* 2010, **19**(13):2725-2738.
52. Ng MC, Hester JM, Wing MR, Li J, Xu J, Hicks PJ, Roh BH, Lu L, Divers J, Langefeld CD: **Genome-Wide Association of BMI in African Americans**. *Obesity* 2012, **20**(3):622-627.
53. Holder JL, Butte NF, Zinn AR: **Profound obesity associated with a balanced translocation that disrupts the SIM1 gene**. *Human molecular genetics* 2000, **9**(1):101-108.
54. Kurokawa N, Young EH, Oka Y, Satoh H, Wareham NJ, Sandhu M, Loos RJ: **The ADRB3 Trp64Arg variant and BMI: a meta-analysis of 44 833 individuals**. *International Journal of Obesity* 2008, **32**(8):1240-1249.
55. Romero-Corral A, Somers VK, Sierra-Johnson J, Thomas RJ, Collazo-Clavell M, Korinek J, Allison TG, Batsis J, Sert-Kuniyoshi F, Lopez-Jimenez F: **Accuracy of body mass index in diagnosing obesity in the adult general population**. *International journal of obesity* 2008, **32**(6):959-966.
56. Taylor RW, Jones IE, Williams SM, Goulding A: **Evaluation of waist circumference, waist-to-hip ratio, and the conicity index as screening tools for high trunk fat mass, as measured by dual-energy X-ray absorptiometry, in children aged 3–19 y**. *The American Journal of Clinical Nutrition* 2000, **72**(2):490-495.
57. Hardy R, Wills AK, Wong A, Elks CE, Wareham NJ, Loos RJ, Kuh D, Ong KK: **Life course variations in the associations between FTO and MC4R gene variants and body size**. *Human molecular genetics* 2010, **19**(3):545-552.
58. Kaakinen M, Läärä E, Pouta A, Hartikainen A-L, Laitinen J, Tammelin TH, Herzig K-H, Sovio U, Bennett AJ, Peltonen L: **Life-course analysis of a fat mass and obesity-associated (FTO) gene variant and body mass index in the Northern Finland Birth Cohort 1966 using structural equation modeling**. *American journal of epidemiology* 2010, **172**(6):653-665.
59. Hertel JK, Johansson S, Sonestedt E, Jonsson A, Lie RT, Platou CG, Nilsson PM, Rukh G, Midthjell K, Hveem K: **FTO, Type 2 Diabetes, and Weight Gain Throughout Adult Life A Meta-Analysis of 41,504 Subjects From the Scandinavian HUNT, MDC, and MPP Studies**. *Diabetes* 2011, **60**(5):1637-1644.

Table 1. Description of MESA population by ethnicity[a]

Characteristic [b]	Hispanics (n=1257) [c,d]	Asians (n=705) [c,d]	African Americans (n=1551) [c,d]	Caucasians (n=2416) [c,d]
Age	61.70 ± 10.34	62.89 ± 10.40	62.41 ± 10.11	62.54 ± 10.22
Sex				
Female	637 (50.7)	359 (50.9)	830 (53.5)	1256 (52.0)
Male	620 (49.3)	346 (49.1)	721 (46.5)	1160 (48.0)
Income				
Low	921 (75.0)	471 (67.3)	741 (51.6)	768 (31.8)
High	307 (25.0)	229 (32.7)	695 (48.4)	1648 (68.2)
Education				
<12 yrs	597 (47.5)	178 (25.3)	189 (12.2)	117 (4.8)
12-15 yrs	537 (42.7)	257 (36.5)	841 (54.2)	1085 (44.9)
≥ 16 yrs	123 (9.8)	270 (38.3)	521 (33.6)	1214 (50.3)
Smoking				
Ever	574 (45.7)	178 (25.3)	852 (54.9)	1350 (55.9)
Never	683 (54.3)	527 (74.8)	699 (45.1)	1066 (44.1)
Physical Activity (met-min/wk)	6024.52 ± 6156.59	3663.04 ± 3777.82	6538.28 ± 6913.56	5717.60 ± 5406.72
Diabetes				
Yes	216 (17.2)	81 (11.5)	243 (15.7)	126 (5.2)
No	1041 (82.8)	624 (88.5)	1308 (84.3)	2290 (94.8)
Arthritis				
Yes	431 (34.3)	177 (25.1)	674 (43.5)	853 (35.3)
No	826 (65.7)	528 (74.9)	877 (56.5)	1563 (64.7)
Body Mass Index (kg/m^2)	29.29 ± 5.10	24.03 ± 3.30	30.13 ± 5.86	27.74 ± 5.07

[a] Table values are mean ± SD for continuous variables and n (column %) for categorical variables.
[b] Characteristics included are covariates in at least one ethnic-specific linear regression model.
[c] Sample size represents number of individuals that passed QC and have complete data for all ethnic-specific linear regression model covariates.
[d] Percentages may not sum to 100% due to rounding and n's may not sum to sample size due to missing values for variables not included as covariates in ethnic-specific regression models.

Table 2. Bivariate associations with BMI by ethnicity

	Hispanics		Asians		African Americans		Caucasians	
Characteristic	**Beta (SE)**	**p**	**Beta (SE)**	**p**	**Beta (SE)**	**p**	**Beta (SE)**	**p**
Age	-0.041 (0.014)	0.003	-0.019 (0.012)	0.113	-0.087 (0.015)	<0.001	-0.036 (0.010)	<0.001
Sex								
Female	Reference	--	Reference	--	Reference	--	Reference	--
Male	-1.304 (0.285)	<0.001	0.200 (0.249)	0.421	-2.577 (0.291)	<0.001	0.432 (0.206)	0.036
Income								
Low	0.410 (0.337)	0.225	-0.419 (0.266)	0.116	-0.013 (0.312)	0.9670	0.712 (0.221)	0.001
High	Reference	--	Reference	--	Reference	--	Reference	--
Education								
<12 yrs	Reference	--	Reference	--	Reference	--	Reference	--
12-15 yrs	0.334 (0.303)	0.270	-0.417 (0.322)	0.195	0.401 (0.471)	0.395	-0.640 (0.491)	0.193
≥ 16 yrs	-0.720 (0.504)	0.154	0.117 (0.318)	0.713	-0.509 (0.497)	0.306	-1.569 (0.489)	0.001
Smoking								
Ever	0.296 (0.289)	0.305	0.223 (0.286)	0.436	-0.641 (0.299)	0.0320	-0.231 (0.208)	0.266
Never	Reference	--	Reference	--	Reference	--	Reference	--
Physical Activity								
Quartile 1	Reference	--	Reference	--	Reference	--	Reference	--
Quartile 2	-0.773 (0.402)	0.054	0.307 (0.349)	0.379	-0.032 (0.421)	0.939	-1.083 (0.292)	<0.001
Quartile 3	-0.368 (0.405)	0.363	0.298 (0.354)	0.400	-0.310 (0.423)	0.464	-0.798 (0.291)	0.006
Quartile 4	-0.873 (0.407)	0.032	0.179 (0.353)	0.613	-0.209 (0.423)	0.621	-0.952 (0.292)	0.001
Diabetes								
Yes	2.389 (0.375)	<0.001	0.903 (0.389)	0.021	1.535 (0.408)	<0.001	3.639 (0.458)	<0.001
No	Reference	--	Reference	--	Reference	--	Reference	--
Arthritis								
Yes	1.274 (0.301)	<0.001	0.632 (0.286)	0.028	1.656 (0.298)	<0.001	1.058 (0.215)	<0.001
No	Reference	--	Reference	--	Reference	--	Reference	--

Abbreviation: SE=standard error

Table 3. Final ethnic-specific multivariate linear regression models

	Hispanics		Asians		African Americans		Caucasians	
Characteristic	**Beta (SE)**	**p**	**Beta (SE)**	**p**	**Beta (SE)**	**p**	**Beta (SE)**	**p**
Age	-0.074 (0.013)	< 0.001	-0.028 (0.012)	0.020	-0.114 (0.014)	< 0.001	-0.079 (0.011)	< 0.001
Sex								
Female	Reference	--	Reference	--	Reference	--	Reference	--
Male	-1.437 (0.275)	< 0.001	0.219 (0.248)	0.377	-2.169 (0.280)	< 0.001	0.996 (0.211)	< 0.001
Income								
Low							0.713 (0.239)	0.003
High							Reference	--
Education								
0-11yrs of schooling			Reference	--	Reference	--	Reference	--
12-15yrs of schooling			-0.468 (0.312)	0.134	0.199 (0.449)	0.658	-0.479 (0.486)	0.324
College graduates			0.197 (0.324)	0.543	-0.870 (0.474)	0.066	-1.545 (0.502)	0.002
Smoking								
Ever	0.739 (0.273)	0.007					-0.414 (0.203)	0.041
Never	Reference	--					Reference	--
Physical Activity								
Quartile 1							Reference	--
Quartile 2							-1.045 (0.283)	< 0.001
Quartile 3							-0.813 (0.282)	0.004
Quartile 4							-1.397 (0.288)	< 0.001
Diabetes								
Yes	2.579 (0.355)	<0.001	0.855 (0.384)	0.026	1.871 (0.383)	<0.001	3.340 (0.452)	< 0.001
No	Reference	--	Reference	--	Reference	--	Reference	--
Arthritis								
Yes	1.138 (0.290)	<0.001	0.787 (0.293)	0.008	1.968 (0.292)	<0.001	1.327 (0.220)	< 0.001
No	Reference	--	Reference	--	Reference	--	Reference	--

Abbreviation: SE=standard error

Table 4. Summary of Individual and SNP Quality Control Procedures

	Hispanics	Asians	African Americans	Caucasians
Individual QC summary				
Individuals with available genotypic information	1,449	775	1,611	2,526
Individuals excluded due to low call rate (< 95%)	0	0	0	0
Related Individuals	177	52	31	18
Individuals excluded due to missing data for $\geq$ 1 covariate in ethnic-specific regression models	15	18	29	92
Individuals analyzed	1,257	705	1,551	2,416
SNP QC summary				
SNPs genotyped	909,622	909,622	909,622	909,622
SNPs excluded due to low call rate (< 98%)	23,307	27,333	18,841	22,359
SNPs excluded due to significant deviation from Hardy-Weinberg equilibrium ($p< 5.5 \times 10^{-8}$)	1,817	633	1,651	1,663
SNPs excluded due to low MAF (< 0.01)[a]	31,220	197,658	17,182	135,941
SNPs analyzed	853,278	683,998	871,948	749,659

[a] SNPs with MAF < 0.01 include monomorphic SNPs

Table 5. Top SNPs ($p< 5.5x10^{-6}$) across the four MESA ethnic groups[a]

					Hispanics		Asians		African Americans		Caucasians	
	SNP	Risk Allele	Chr	Bp	beta	p[b]	beta	p[b]	beta	p[b]	beta	p[b]
Hispanics	rs16883086	G	6	20101418	3.518	$4.44x10^{-06}$	monomorphic	monomorphic	0.139	0.696	n/a[c]	n/a[c]
	rs16909322	C	8	82614447	1.086	$4.96x10^{-06}$	-0.422	0.041	-0.457	0.034	-0.115	0.509
	rs1480000	T	8	145672815	-0.910	$5.43x10^{-06}$	0.060	0.738	0.140	0.505	-0.074	0.599
	rs12253976	G	10	3825732	5.542	**$6.88x10^{-09}$**	monomorphic	monomorphic	-0.322	0.419	n/a[c]	n/a[c]
	rs1409874	T	10	6690432	-1.028	$3.52x10^{-06}$	0.129	0.511	-0.440	0.164	-0.070	0.701
	rs1249269	G	10	28660149	0.924	$4.32x10^{-06}$	-0.348	0.216	-0.054	0.798	-0.112	0.433
	rs12255372	T	10	114798892	-1.117	$2.25x10^{-06}$	n/a[c]	n/a[c]	-0.594	0.008	-0.218	0.159
	rs16943469	T	17	54799958	3.759	$3.86x10^{-06}$	monomorphic	monomorphic	0.221	0.582	n/a[c]	n/a[c]
	rs6038725	A	20	7081039	-0.915	$3.62x10^{-06}$	0.543	0.009	0.122	0.559	0.226	0.131
	rs179747	G	20	7107995	-0.951	$1.41x10^{-06}$	0.323	0.110	0.213	0.318	0.214	0.151
	rs6085916	G	20	7112725	-0.916	$2.84x10^{-06}$	0.301	0.130	0.261	0.232	0.233	0.118
	rs2326897	C	20	7112922	-0.937	$1.84x10^{-06}$	0.300	0.131	0.285	0.193	0.230	0.123
	rs13036410	T	20	7122578	-0.934	$1.73x10^{-06}$	0.301	0.130	0.216	0.315	0.215	0.151
Asians	rs4610207	G	3	10974494	0.081	0.813	1.779	$4.63x10^{-06}$	0.417	0.098	-0.048	0.842
	rs9961691	C	18	17966386	0.147	0.506	-0.994	$1.53x10^{-06}$	-0.181	0.436	0.027	0.896
African Americans	rs11884662	C	2	19313837	0.237	0.389	-0.468	0.236	-0.984	$2.78x10^{-06}$	-0.177	0.283
	rs7602754	C	2	19315168	0.263	0.338	-0.468	0.236	-1.004	$1.43x10^{-06}$	-0.179	0.276
	rs10170855	C	2	19315301	0.219	0.429	-0.468	0.236	-0.968	$4.15x10^{-06}$	-0.178	0.278
	rs4852003	G	2	240414673	-0.157	0.764	-0.188	0.615	3.470	$5.01x10^{-06}$	0.303	0.281
	rs6739663	A	2	240878904	0.089	0.651	0.092	0.602	-0.933	$3.81x10^{-06}$	0.057	0.694
	rs7763896	C	6	132321497	0.121	0.772	monomorphic	monomorphic	2.920	$3.05x10^{-06}$	0.014	0.948
	rs1567101	A	10	30901089	-0.025	0.934	-0.025	0.902	1.059	$2.92x10^{-07}$	0.239	0.288
	rs7092615	T	10	30902186	-0.023	0.940	-0.012	0.952	1.077	$2.26x10^{-07}$	0.205	0.359
Caucasians	rs17526301	T	2	118513303	-0.186	0.522	-0.635	0.106	-0.351	0.449	1.112	$1.87x10^{-06}$
	rs7585972	T	2	118515382	-0.109	0.721	-0.636	0.108	-0.508	0.283	1.080	$3.60x10^{-06}$
	rs6899277	G	5	78881387	0.046	0.823	-0.023	0.895	0.293	0.276	-0.692	$2.14x10^{-06}$
	rs6866721	C	5	116240680	-0.013	0.948	-0.101	0.585	-0.122	0.558	0.764	$9.23x10^{-08}$
	rs1672492	G	5	116250736	0.036	0.855	-0.197	0.340	0.358	0.076	0.696	$2.07x10^{-06}$
	rs1672491	C	5	116250866	-0.032	0.869	-0.192	0.352	0.358	0.076	0.702	$1.69x10^{-06}$
	rs7704264	A	5	116252052	0.060	0.765	-0.257	0.136	0.116	0.633	0.713	$5.55x10^{-07}$
	rs4868301	T	5	173056118	-0.022	0.913	0.051	0.805	-0.368	0.110	-0.665	$1.76x10^{-06}$
	rs4868303	A	5	173057108	0.036	0.858	0.056	0.785	-0.349	0.129	-0.667	$1.51x10^{-06}$
	rs9359264	G	6	78107510	0.014	0.943	0.210	0.259	-0.017	0.938	0.707	$4.65x10^{-06}$
	rs2497155	T	6	84920682	-0.427	0.257	0.004	0.991	-0.075	0.906	-1.056	$2.56x10^{-06}$
	rs7926805	C	11	17715163	0.156	0.516	0.301	0.200	-0.477	0.018	-0.766	$2.35x10^{-06}$

Abbreviations: Chr = Chromosome; Bp = base pair; SE = standard error

[a] Shaded values indicate the top SNPs within each ethnic group; p-value in bold indicates genome-wide significant result

[b] p-values adjusted for top two principal components and all covariates in ethnic-specific linear regression models (see Table 4)

[c] n/a indicates SNP had a MAF < 1% in that particular ethnic group

Table 6. Top SNP candidates in each ethnic group

		Unadjusted model		+ PC 1 and 2		+ Age and sex		Full model[a]	
Ethnicity	**SNP**	**Beta (95% CI)**	**p**	**Beta (95% CI)**	**p**	**Beta (95% CI)**	**p**	**Beta (95% CI)**	**p**
Hispanics	rs12253976	5.536 (3.650, 7.421)	$1.09x10^{-08}$	5.688 (3.763, 7.614)	$8.92x10^{-09}$	5.423 (3.512, 7.335)	$3.28x10^{-08}$	5.542 (3.680, 7.404)	$6.88x10^{-09}$
Asians	rs9961691	-0.935 (-1.341, -0.530)	$7.17x10^{-06}$	-0.931 (-1.335 -0.526)	$7.78x10^{-06}$	-0.934 (-1.339, -0.528)	$7.44x10^{-06}$	-0.994 (-1.396, -0.592)	$1.53x10^{-06}$
African Americans	rs7092615	0.979 (0.556, 1.401)	$6.11x10^{-06}$	0.947 (0.518, 1.375)	$1.60x10^{-05}$	1.059 (0.645, 1.473)	$5.91x10^{-07}$	1.077 (0.671, 1.483)	$2.26x10^{-07}$
Caucasians	rs6866721	0.707 (0.418, 0.997)	$1.71x10^{-06}$	0.747 (0.458, 1.037)	$4.53x10^{-07}$	0.751 (0.463, 1.040)	$3.65x10^{-07}$	0.764 (0.484, 1.043)	$9.23x10^{-08}$

Abbreviations: PC = principal components; CI= Confidence Interval
[a] Full model includes: PC 1 and 2, age, sex, and ethnic-specific model covariates (see Table 4)

Table 7. Associations with *FTO* rs9939609 across the four MESA ethnic groups

	Caucasians		Hispanics		Asians		African Americans	
	Beta (95% CI)	**p**	**Beta (95% CI)**	**p**	**Beta (95% CI)**	**p**	**Beta (95% CI)**	**p**
Unadjusted model	0.2658 (-0.030, 0.562)	0.078	0.102 (-0.317, 0.520)	0.635	-0.347 (-0.862, 0.168)	0.187	0.133 (-0.288, 0.553)	0.537
+ PC 1 and 2	0.2933 (-0.003, 0.590)	0.053	0.191 (-0.237, 0.619)	0.381	-0.407 (-0.923, 0.108)	0.122	0.127 (-0.294, 0.547)	0.556
+ age and sex	0.2766 (-0.019, 0.572)	0.067	0.257 (-0.167, 0.681)	0.234	-0.406 (-0.921, 0.110)	0.123	0.073 (-0.333, 0.480)	0.724
Full Model[a]	0.279 (-0.008, 0.566)	0.057	0.233 (-0.180, 0.647)	0.269	-0.389 (-0.903, 0.125)	0.138	0.114 (-0.285, 0.512)	0.577
MESA MAF	0.4084		0.3096		0.1286		0.4791	
HAPMAP MAF	0.46 (CEU)		0.228 (MEX)		0.15 (CHB)		0.50 (ASW)	

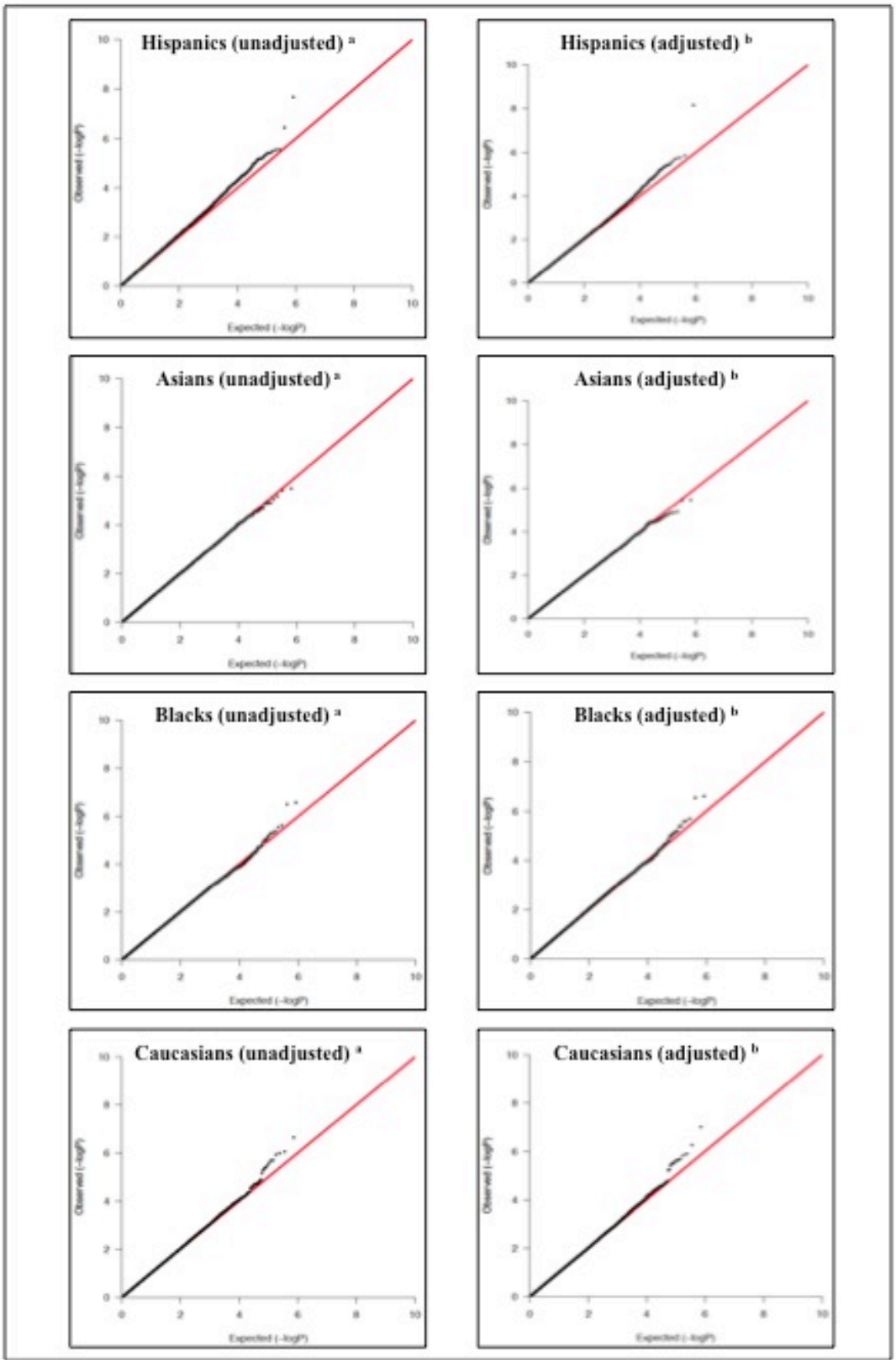
Abbreviations: PC = principal components; CI=confidence interval; ASW = African ancestry in Southwest USA population; CEU = Utah residents with Northern and Western European ancestry from the CEPH collection; CHB = Han Chinese in Beijing, China; and MEX = Mexican ancestry in Los Angeles, California
[a] Full model includes: PC 1 and 2, age, sex, and ethnic-specific model covariates (see Table 4)

Table 8. Average BMI by Genotype for the top SNP in each ethnic group

Race/Ethnicity	SNP	Risk Allele	MAF	Genotype	Counts	BMI[a]
Hispanics	rs12253976	G	0.011	GG	0	--
				GT	28	34.70 ± 7.59
				TT	1229	29.16 ± 4.96
Asians	rs9961691	C	0.203	CC	39	23.24 ± 2.76
				CG	208	23.21 ± 2.95
				GG	458	24.47 ± 3.42
African Americans	rs7092615	T	0.375	TT	220	30.96 ± 5.89
				TC	716	30.61 ± 5.98
				CC	606	29.25 ± 5.60
Caucasians	rs6866721	C	0.437	CC	450	28.69 ± 5.38
				CA	1206	27.70 ± 5.14
				AA	756	27.21 ± 4.65

[a] BMI values are mean ± SD

Figure 1. Quantile-Quantile p-value plots, pre- and post- adjustment for population stratification



[a] p-values adjusted only for ethnic-specific model covariates
[b] p-values adjusted for ethnic-specific model covariates and for top two principal components

Figure 2. Regional association plot for top SNP rs12253976 in Hispanic subjects

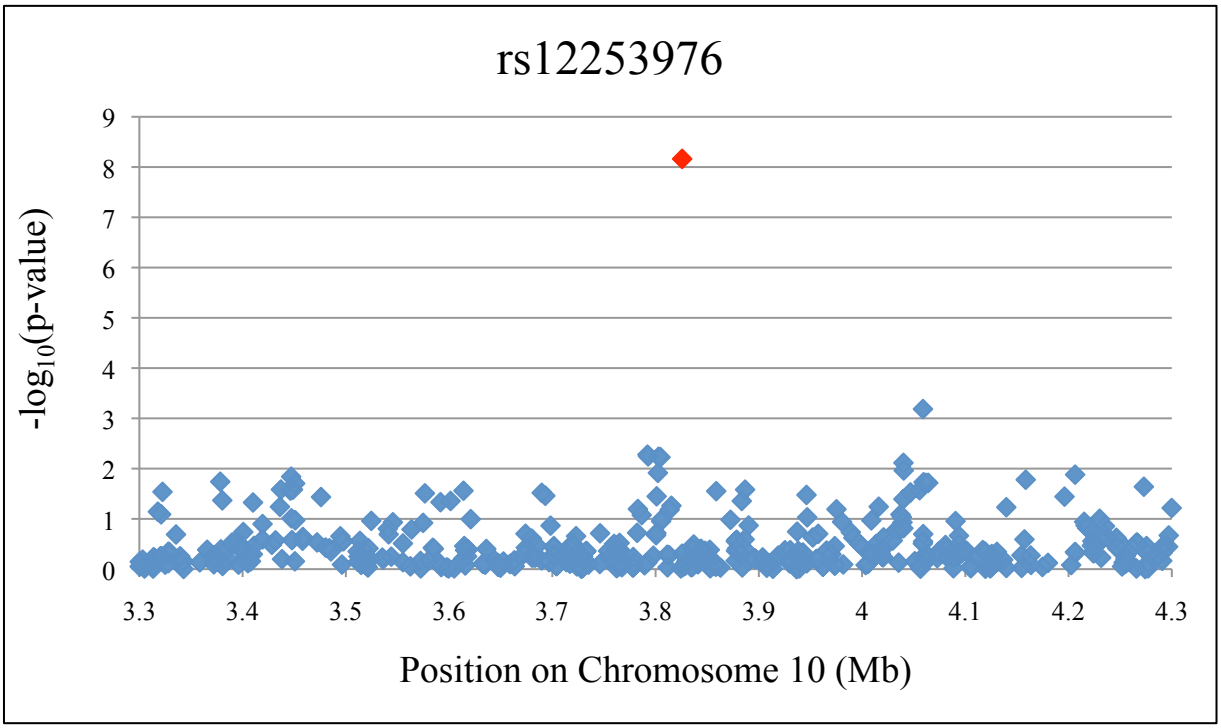


Figure 3. Regional association plot for top SNP rs6866721 in Caucasian subjects

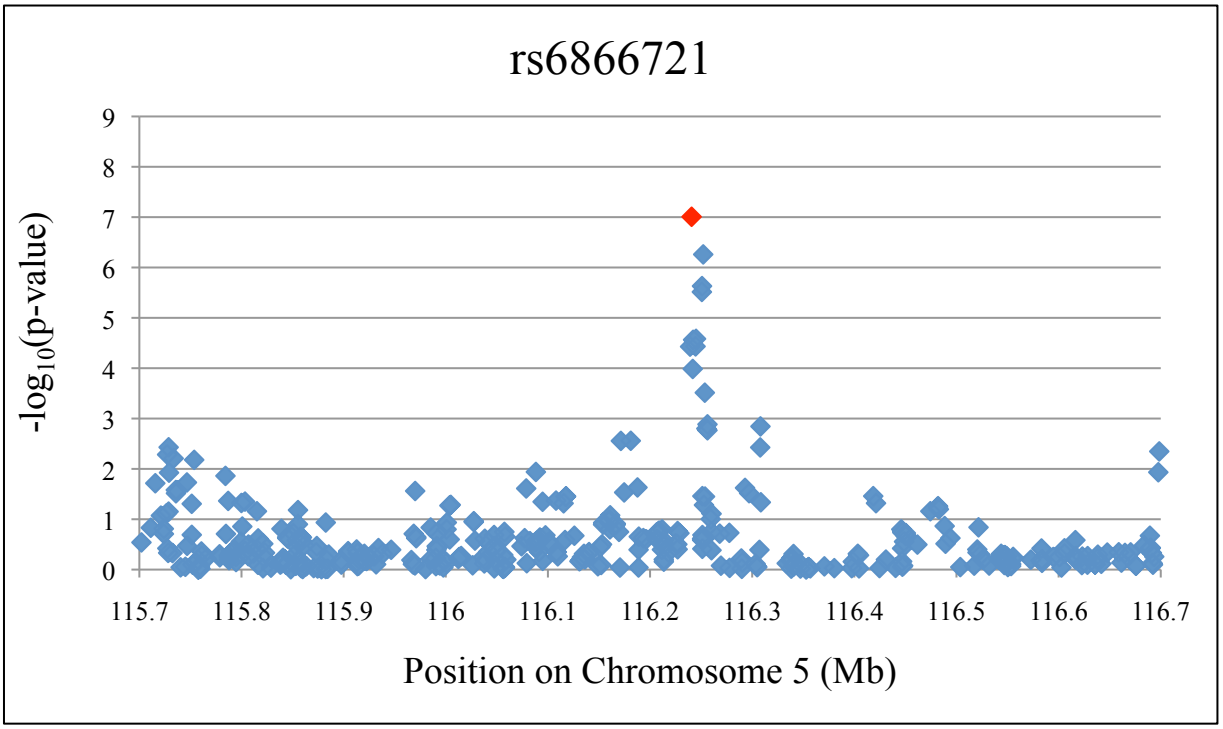


Figure 4. Regional Association plot for top SNP rs7092615 in African American subjects

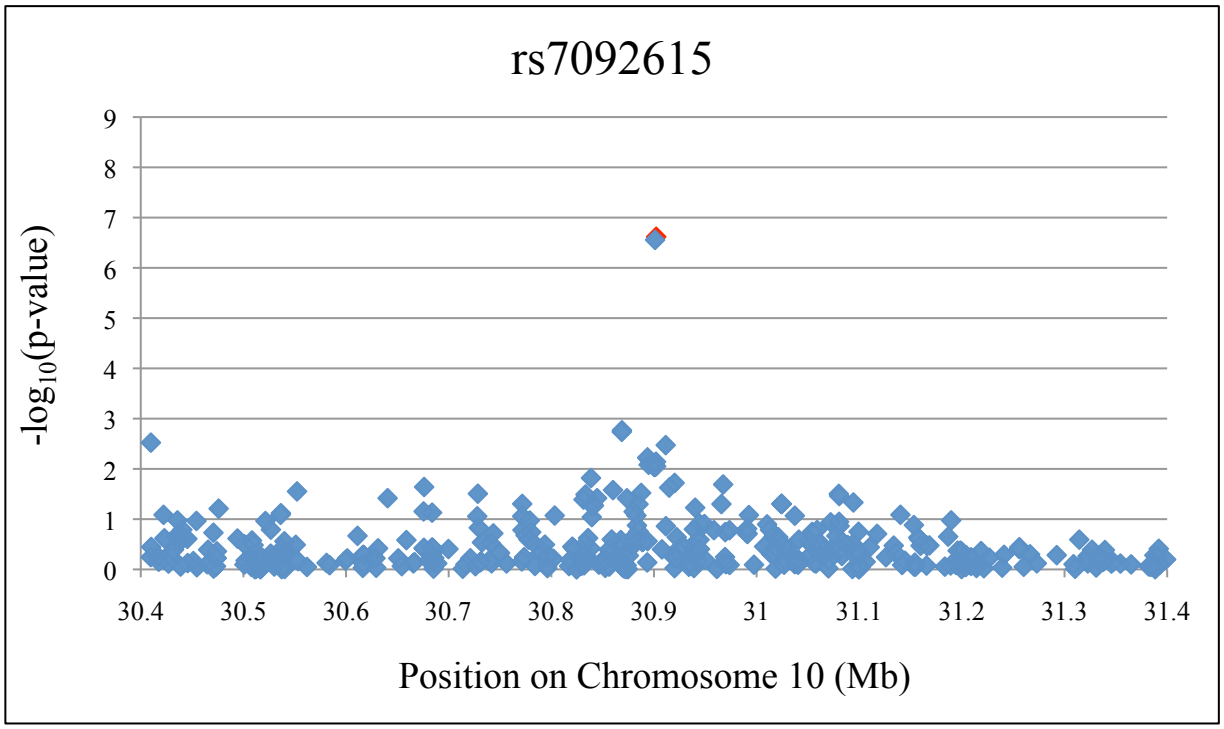


Figure 5. Regional Association plot for top SNP rs9961691 in Asian subjects