# Fine Mapping and Identification of BMI Loci in African Americans

Jian Gong,[1,*] Fredrick Schumacher,[9] Unhee Lim,[8] Lucia A. Hindorff,[10] Jeff Haessler,[1] Steven Buyske,[4,5] Christopher S. Carlson,[1] Stephanie Rosse,[1] Petra Bůžková,[19] Myriam Fornage,[23] Myron Gross,[24] Nathan Pankratz,[24] James S. Pankow,[25] Pamela J. Schreiner,[25] Richard Cooper,[13] Georg Ehret,[11,12] C. Charles Gu,[17] Denise Houston,[18] Marguerite R. Irvin,[16] Rebecca Jackson,[6] Lew Kuller,[7] Brian Henderson,[9] Iona Cheng,[27] Lynne Wilkens,[8] Mark Leppert,[15] Cora E. Lewis,[14] Rongling Li,[10] Khanh-Dung H. Nguyen,[11] Robert Goodloe,[22] Eric Farber-Eger,[22] Jonathan Boston,[22] Holli H. Dilks,[22] Marylyn D. Ritchie,[20] Jay Fowke,[21,22] Loreall Pooler,[9] Misa Graff,[3] Lindsay Fernandez-Rhodes,[3] Barbara Cochrane,[26] Eric Boerwinkle,[23] Charles Kooperberg,[1] Tara C. Matise,[4] Loic Le Marchand,[8] Dana C. Crawford,[22] Christopher A. Haiman,[9] Kari E. North,[2,3] and Ulrike Peters[1,*]

Genome-wide association studies (GWASs) primarily performed in European-ancestry (EA) populations have identified numerous loci associated with body mass index (BMI). However, it is still unclear whether these GWAS loci can be generalized to other ethnic groups, such as African Americans (AAs). Furthermore, the putative functional variant or variants in these loci mostly remain under investigation. The overall lower linkage disequilibrium in AA compared to EA populations provides the opportunity to narrow in or fine-map these BMI-related loci. Therefore, we used the Metabochip to densely genotype and evaluate 21 BMI GWAS loci identified in EA studies in 29,151 AAs from the Population Architecture using Genomics and Epidemiology (PAGE) study. Eight of the 21 loci (*SEC16B, TMEM18, ETV5, GNPDA2, TFAP2B, BDNF, FTO*, and *MC4R*) were found to be associated with BMI in AAs at $5.8 \times 10^{-5}$. Within seven out of these eight loci, we found that, on average, a substantially smaller number of variants was correlated ($r^2 > 0.5$) with the most significant SNP in AA than in EA populations (16 versus 55). Conditional analyses revealed *GNPDA2* harboring a potential additional independent signal. Moreover, Metabochip-wide discovery analyses revealed two BMI-related loci, *BRE* (rs116612809, $p = 3.6 \times 10^{-8}$) and *DHX34* (rs4802349, $p = 1.2 \times 10^{-7}$), which were significant when adjustment was made for the total number of SNPs tested across the chip. These results demonstrate that fine mapping in AAs is a powerful approach for both narrowing in on the underlying causal variants in known loci and discovering BMI-related loci.

## Introduction

Obesity (MIM 601665) is a major risk factor for a number of chronic diseases, such as type 2 diabetes (MIM 125853), hyperlipidemia (MIM 144250), cardiovascular diseases, and several cancer types.[1,2] Worldwide obesity prevalence has nearly doubled since 1980, and in 2008 more than 1.4 billion adults worldwide were obese. In the United States, more than one-third of adults (35.7%) were obese in 2010.

Studies have shown that obesity is highly heritable; heritability is estimated to fall in the range of 40%–70%.[3,4] Genome-wide association studies (GWAS) have identified numerous loci associated with body mass index (BMI),[5–7] a common measure of obesity. However, most of these studies were performed among European-ancestry (EA) populations. It is still unclear whether previously identified GWAS loci are population specific or whether they can be generalized to other ethnic groups, such as African Americans (AAs). Furthermore, the overall lower linkage disequilibrium (LD) patterns in AA compared to EA populations can offer opportunities to narrow in or fine-map BMI-related loci.[8] This will help to reduce the number of variants for functional follow-up studies, which tend to

[1]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; [2]Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, NC 27514, USA; [3]Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA; [4]Department of Genetics, Rutgers University, Piscataway, NJ 08854, USA; [5]Department of Statistics and Biostatistics, Rutgers University, Piscataway, NJ 08854, USA; [6]Department of Internal Medicine, Ohio State Medical Center, Columbus, OH 43210, USA; [7]Department of Medicine, University of Pittsburgh, Pittsburgh, PA 15261, USA; [8]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI 96813, USA; [9]Department of Preventive Medicine, Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA; [10]Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA; [11]Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; [12]Division of Cardiology, Geneva University Hospital, Geneva, OH 44041, Switzerland; [13]Preventive Medicine and Epidemiology, Loyola University, Chicago, IL 60153, USA; [14]Department of Medicine, University of Alabama, Birmingham, AL 35294, USA; [15]Department of Human Genetics, University of Utah, Salt Lake City, UT 84112, USA; [16]Department of Biostatistics, University of Alabama, Birmingham, AL 35294, USA; [17]Department of Biostatistics, Washington University, St. Louis, MO 63110, USA; [18]Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA; [19]Department of Biostatistics, University of Washington, Seattle, WA 98195, USA; [20]Pennsylvania State University, State College, PA 16801, USA; [21]Vanderbilt Epidemiology Center, Vanderbilt University Medical Center, Nashville, TN 37203, USA; [22]Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, TN, 37203, USA; [23]The Human Genetics Center and Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; [24]Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455, USA; [25]Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA; [26]Baylor College of Medicine, Houston, TX 77030, USA; [27]Cancer Prevention Institute of California, Fremont, CA 94538, USA
*Correspondence: jgong@fhcrc.org (J.G.), upeters@fhcrc.org (U.P.)
http://dx.doi.org/10.1016/j.ajhg.2013.08.012. 

be time and labor intensive. In addition, dense genotyping of the GWAS loci could aid the discovery of additional independent signals within the GWAS loci.

In this study, we densely genotyped 21 BMI loci identified in EA studies in 29,151 AAs from the Population Architecture using Genomics and Epidemiology (PAGE) consortium by using the Metabochip.[9] We aimed to fine-map the 21 known BMI loci in the AA population and search for additional independent signals associated with BMI. For validated loci in AAs, we evaluated whether weaker LD patterns in AAs can help narrow in on the underlying potential causal variants. In addition, because Metabochip was developed to test putative association signals for BMI and many obesity-related metabolic and cardiovascular traits and to fine-map established loci,[9] we also conducted a Metabochip-wide discovery-oriented analysis to search for potential BMI-associated loci.

## Subjects and Methods

### Study Population

The National Human Genome Research Institute funds the PAGE consortium to investigate the epidemiologic architecture of well-replicated genetic variants associated with human diseases or traits.[10] PAGE consists of a coordinating center and four consortia, Epidemiologic Architecture for Genes Linked to Environment (EAGLE), which is uses data from Vanderbilt University Medical Center's biorepository and links it to deidentified electronic medical records (BioVU); the Multiethnic Cohort Study (MEC); the Women's Health Initiative (WHI); and Causal Variants Across the Life Course (CALiCo), itself a consortium of five cohort studies—the Atherosclerosis Risk in Communities (ARIC) study, Coronary Artery Risk Development in Young Adults (CARDIA), the Cardiovascular Health Study (CHS), the Hispanic Community Health Study/Study of Latinos, and the Strong Heart Study.[10]

This PAGE Metabochip study included AA participants from the ARIC, BioVU, CHS, CARDIA, MEC, and WHI studies and from extended collaborations to two additional studies – GenNet and the Hypertension Genetic Epidemiology Network (HyperGen) (Table S1, in the Supplemental Data available with this article online). The detailed description of each study can be found in the Supplemental Data. We excluded underweight (BMI $<$ 18.5 kg/m$^2$) and extremely overweight (BMI $>$ 70 kg/m$^2$) individuals under the assumption that these extremes could be attributable to data-coding errors or an underlying rare condition outside the scope of this investigation. We also limited analysis to adults (defined as having an age $>$ 20 years). The CARDIA participants are young, and the BMI $<$ 18.5 exclusion criterion was not applied in this cohort. All studies were approved by institutional review boards at their respective sites, and all study participants provided informed consent.

### Anthropometric Measurements

For individuals from the ARIC, CHS, CARDIA, HyperGEN, GenNet, and WHI studies, BMI was calculated from height and weight measured at the time of study enrollment. For individuals from BioVU, the median height and weight across all visit years were used in BMI calculations. For individuals from MEC, self-reported height and weight were used for calculations of baseline BMI. A validation study within MEC has shown high validity of self-reported height and weight. Specifically, this study showed that BMI was underestimated on the basis of self-reports versus measured weight, but the difference was small ($<$ 1 BMI unit) and was comparable to the findings from national surveys.[11]

### Genotyping and Quality Control

Genotyping was performed with the Metabochip, whose design has been described elsewhere.[9] In brief, the Metabochip, a custom Illumina iSelect genotyping array of nearly 200,000 SNP markers, is designed to cost-effectively analyze putative association signals identified through GWAS meta-analyses of many obesity-related metabolic and cardiovascular traits and to fine-map established loci.[9] Metabochip SNPs were selected from the catalogs developed by the International HapMap and 1000 Genomes projects.[9] More than 122,000 SNPs were included for fine mapping of 257 GWAS loci of 23 traits (including 21 BMI loci).[9] For determination of the boundaries around each GWAS index SNP, all SNPs with $r^2 \geq 0.5$ with the index SNP were identified, and then initial boundaries were expanded by 0.02 cM in either direction through use of the HapMap-based genetic map. SNPs were excluded if (1) the Illumina design score was $<$0.5 or (2) SNPs within 15 bp in both directions of the SNP of interest could be found with an allele frequency of $>$0.02 among Europeans (CEU). SNPs annotated as nonsynonymous, essential splice site, or stop codon were included regardless of allele frequency, design score, or nearby SNPs in the primer.[9] Twenty-one BMI GWAS loci identified at the time at which the Metabochip was designed were represented for signal fine mapping (Table S2).

Samples were genotyped at the Human Genetics Center of the University of Texas, Houston (ARIC, CHS, CARDIA, GenNet, and HyperGEN), the Vanderbilt DNA Resources Core in Nashville (BioVU), the University of Southern California Epigenome Center (MEC), and the Translational Genomics Research Institute (WHI). Each center genotyped the same 90 HapMap YRI (Yoruba in Ibadan, Nigeria) samples to facilitate cross-study quality control (QC), as well as 2%–3% study-specific blinded replicates to assess genotyping quality. Genotypes were called separately for each study via GenomeStudio with the GenCall 2.0 algorithm. Study-specific cluster definitions (based on samples with call rate $>$ 95%; ARIC, BioVU, CHS, CARDIA, MEC, and WHI) or cluster definitions provided by Illumina (GenNet and HyperGEN) were used for sample calling, and samples were kept in the analysis if the call rate was $>$95%. We excluded SNPs with a GenTrain score $<$0.6 (ARIC, BioVU, CHS, CARDIA, MEC, and WHI) or $<$0.7 (GenNet and HyperGEN), a cluster separation score $<$0.4, a call rate $<$0.95, and a Hardy-Weinberg equilibrium $p < 1 \times 10^{-6}$. We utilized the common 90 YRI samples and excluded any SNP that had more than 1 Mendelian error (in 30 YRI trios), any SNP that had more than two replication errors with discordant calls when comparisons were made across studies in 90 YRI samples, and any SNP that had more than three discordant calls for 90 YRI genotyped in PAGE versus the HapMap database. SNPs were excluded from the meta-analyses if they were present in less than three studies.

For ARIC, BioVU, CHS, CARDIA, MEC, and WHI combined we identified related individuals by using PLINK to estimate identical-by-descent (IBD) statistics for all pairs. When apparent pairs of first-degree relatives were identified, we excluded from each pair the member with the lower call rate. We excluded from further analysis samples with an inbreeding coefficient (F) above

0.15 (ARIC, BioVU, CHS, CARDIA, MEC, and WHI).[12] We determined principal components of ancestry in each study separately by using EIGENSOFT[13,14] and excluded apparent ancestral outliers from further analysis as described elsewhere.[15]

### WHI SHARe Imputation

Of the WHI women genotyped on the Metabochip, 1,962 women were part of the group of 8,288 WHI subjects genotyped for the WHI SNP Health Association Resource (SHARe) GWAS via the Affymetrix 6.0 platform. To improve statistical power, we imputed the Metabochip SNPs in the remaining 6,326 SHARe subjects with Affymetrix 6.0 data. Details can be found elsewhere.[16] In brief, we first merged genotypes for the 1,962 subjects genotyped on both the Affymetrix 6.0 platform and the Metabochip and constructed haplotypes (study-specific reference panel). We then phased the haplotypes for samples genotyped on the Affymetrix 6.0 platform only and performed a haplotype-to-haplotype imputation on Metabochip SNPs for the 6,326 target individuals to estimate genotypes (as allele dosages). We used MACH for phasing and Minimac for final imputation. To evaluate the quality of each imputed SNP, we calculated the dosage $r^2$. We excluded imputed SNPs with $r^2 < 0.5$ for SNPs with allele frequency $< 1\%$ and with $r^2 < 0.3$ for SNPs with allele frequency $> 1\%$. Given the large reference panel and strict QC criteria, this resulted in high imputation quality.[16]

### Statistical Analysis

In each study, we evaluated the association between natural-log-transformed BMI and each SNP. Because the distribution of BMI was not normal (it was skewed toward higher BMI), we used natural-log-transformed BMI,[17] which reduces the influence of potential outlying observations on the analyses. Except for GenNet and HyperGen, linear regression models were used under the assumption of an additive genetic model and with the adjustment for age, sex, study site (as applicable), and ancestry principal components in each study. All models (except WHI) included the interaction term of sex and age so that possible effect modification by sex was accounted for. Family data from GenNet and HyperGen were analyzed with linear mixed models so that relatedness was accounted for. We used fixed-effect models with inverse variance weighting to pool the study-specific association results as implemented in METAL.[18] We used Q-statistics and $I^2$ to measure heterogeneity across studies. To determine whether the 21 previously identified GWAS loci were significantly associated with BMI in AA, we used the p value threshold $5.8 \times 10^{-5}$ as an approximate correction for an average of 866 SNPs at each locus (i.e., 0.05/866 SNPs).[19] To identify additional independent signals in any of the loci that were significantly associated with BMI in our study, we conducted conditional analyses. We performed linear regression models that included the most significant SNP (i.e., lead SNP) as a covariate and each of the other SNPs at the same locus (two SNPs in each model) to search for additional independent signals. If, after adjustment for the lead SNP, any SNP remained significant at the locus-specific Bonferroni-corrected significance level (i.e., 0.05/number of SNPs tested at a given locus), we defined the SNP as an additional independent signal. When testing all other SNPs on the Metabochip, we used a Bonferroni-adjusted significance level based on the total number of SNPs on the chip, $2.5 \times 10^{-7}$ (0.05/200,000 SNPs), to declare a BMI-associated SNP. No inflation was observed in any analysis (the inflation factor $\lambda = 1.00$) in our Metabochip-wide analysis. For sex-stratified analysis, GenNet and HyperGen were excluded because of the family-study design.

LD in the AA sample was calculated in 500 kb sliding windows via PLINK.[20] Likewise, the Malmo Diet and Cancer Study on 2,143 controls from a Swedish population[21] provided Metabochip LD and frequency information in Europeans to facilitate the LD pattern comparisons between AA and EA populations. We used LocusZoom plots[22] to graphically display the fine-mapping results. SNP positions from NCBI build 37 were used, and recombination rates were estimated from 1000 Genomes Project data.

### Functional Annotation

To inform the discussion about the underlying potential functional variants, we made functional hypotheses for each of our most significantly BMI-associated variants by compiling a list of correlated SNPs ($r^2 > 0.5$) genotyped in our AA study populations and annotating each list for potential regulatory evidence consistent with enhancers, promoters, insulators, silencers and other effects related to gene expression. Because our lead SNPs and the SNPs in strong LD with our lead SNPs were in noncoding regions, we hypothesized that the underlying biology behind the signal was likely to impact gene expression through some unknown regulatory mechanism. For each list we aligned correlated SNPs with a combined browser view of all currently available ENCODE tracks in the UCSC Genome Browser and compared each allelic region for altered transcription factor binding site (TFBS) motifs by using JASPAR and ConSite. Given that methylation patterns are highly variable, it is useful to look for, in addition to BMI-relevant tissues, histone modifications in many cell lines to identify regions that are actively regulated, meaning that histone marks are present in some but not all cell lines. Thus, in addition to adipocytes, hepatocytes, and neurons, we used ENCODE's histone modification tracks to query a wide variety of cell lines for all available histone marks to identify SNPs falling in various regulatory regions. The DNase hypersensitivity track provided a more precise demarcation of open chromatin loci, and the ChIP-Seq TFBS track provided evidence for the binding of specific proteins. Although less specific than ChIP-Seq, JASPAR, ConSite, and HaploReg databases were used for querying a larger number of conserved TFBS and predicting alterations in predicted motifs between reference and alternate alleles. A 46-way PhastCons track in the UCSC Genome Browser was used as secondary evidence for a regulatory region, but lack of conservation did not rule out a functional candidate. The most likely functional-candidate SNPs for each locus were evaluated for statistical significance in association with BMI. The detailed list of functional annotation data sets we used is shown in Table S3.

## Results

This study consisted of 29,151 AAs from eight studies. Study participants had an average age of 51.2 years (Table S1). Approximately 80% of study participants were women. Within and across studies, men tended to have a lower mean BMI than women. The obesity rate (BMI $\geq$ 30 kg/m$^2$) ranged from 16%–46% in men and 26%–64% in women. After quality control, we tested 18,187 genetic variants across 21 BMI loci and 177,663 variants across the Metabochip.

**Table 1. Association Results for the Most Significant SNP and All Previously Identified GWAS SNPs for Eight BMI-Related Loci with Significant[a] Results in African Americans**

| Region | Lead AA SNP on Metabochip | GWAS Index SNP | Position | Number of SNPs | Candidate Gene | $R^2$ Lead SNP with GWAS SNP in AAs | $R^2$ Lead SNP with GWAS SNP in EA Populations | CA[b] | CAF[c] in AAs | CAF in EA Populations | Effect | p Value | Rsq[d] | Het P[e] | $I^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1q25.2 | rs543874 | rs543874 | 177889480 | 765 | *SEC16B* | - | - | T | 0.75 | 0.81 | −0.0110 | $2.4 \times 10^{-9}$ | NA | 0.04 | 0.52 |
| 2p25.3 | rs6548240 | - | 636929 | 1,123 | *TMEM18* | - | - | A | 0.87 | 0.83 | 0.0130 | $1.1 \times 10^{-7}$ | 0.98 | 0.96 | 0 |
| - | - | rs6548238 | 634905 | - | *TMEM18* | 0.86 | 0.99 | A | 0.12 | 0.17 | −0.0128 | $3.8 \times 10^{-7}$ | 1.00 | 0.82 | 0 |
| - | - | rs2867125 | 622827 | - | *TMEM18* | 0.80 | 0.96 | A | 0.12 | 0.17 | −0.0114 | $5.6 \times 10^{-6}$ | 1.00 | 0.79 | 0 |
| - | - | rs7561317 | 644953 | - | *TMEM18* | 0.49 | 0.99 | A | 0.24 | 0.17 | −0.0054 | $4.8 \times 10^{-3}$ | 1.00 | 0.89 | 0 |
| 3q27.2 | rs7647305 | rs7647305 | 185834290 | 369 | *ETV5* | - | - | A | 0.41 | 0.23 | −0.0069 | $3.2 \times 10^{-5}$ | 0.95 | 0.29 | 0.18 |
| - | - | rs9816226 | 185834499 | - | *ETV5* | 0.37 | 0.85 | A | 0.80 | 0.81 | 0.0067 | $9.1 \times 10^{-4}$ | 0.96 | 0.57 | 0 |
| 4p12 | rs10938397 | rs10938397 | 45182527 | 342 | *GNPDA2* | - | - | A | 0.75 | 0.57 | −0.0099 | $1.7 \times 10^{-7}$ | 0.99 | 0.24 | 0.23 |
| 6p12.3 | rs2744475 | - | 50784880 | 1,685 | *TFAP2B* | - | - | C | 0.67 | 0.71 | −0.0082 | $2.8 \times 10^{-6}$ | 0.98 | 0.39 | 0.05 |
| - | - | rs987237 | 50803050 | - | *TFAP2B* | 0.23 | 0.50 | A | 0.90 | 0.82 | −0.0093 | $5.2 \times 10^{-4}$ | NA | 0.07 | 0.46 |
| 11p14.1 | rs1519480 | - | 27675712 | 688 | *BDNF* | - | - | A | 0.25 | 0.68 | −0.0095 | $7.8 \times 10^{-7}$ | 1.00 | 0.90 | 0 |
| - | - | rs6265 | 27667202 | - | *BDNF* | 0.14 | 0.12 | A | 0.05 | 0.18 | −0.0172 | $1.8 \times 10^{-5}$ | NA | 0.64 | 0 |
| - | - | rs925946 | 27679916 | - | *BDNF* | 0.12 | 0.97 | T | 0.27 | 0.30 | −0.0005 | $7.9 \times 10^{-1}$ | 1.00 | 0.23 | 0.24 |
| - | - | rs10767664[f] | 27724745 | - | *BDNF* | 0.12 | 0.13 | C | 0.93 | 0.76 | 0.0111 | $3.7 \times 10^{-4}$ | 0.99 | 0.28 | 0.18 |
| 16q12.2 | rs62048402 | - | 53803223 | 1,814 | *FTO* | - | - | A | 0.11 | 0.41 | 0.0120 | $5.1 \times 10^{-6}$ | 1.00 | 0.14 | 0.34 |
| - | - | rs1421085 | 53800954 | - | *FTO* | 1.00 | 1.00 | T | 0.89 | 0.59 | −0.0119 | $6.5 \times 10^{-6}$ | NA | 0.15 | 0.33 |
| - | - | rs9930506 | 53830465 | - | *FTO* | 0.44 | 0.81 | T | 0.79 | 0.57 | −0.0082 | $3.9 \times 10^{-5}$ | NA | 0.21 | 0.27 |
| - | - | rs9941349 | 53825488 | - | *FTO* | 0.52 | 0.91[g] | A | 0.19 | 0.41 | 0.0082 | $1.1 \times 10^{-4}$ | NA | 0.56 | 0 |
| - | - | rs1558902 | 53803574 | - | *FTO* | 0.98 | 1.00 | A | 0.88 | 0.58 | −0.0132 | $4.5 \times 10^{-4}$ | 1.00 | 0.43 | 0 |
| - | - | rs8050136 | 53816275 | - | *FTO* | 0.15 | 0.94 | A | 0.43 | 0.41 | 0.0027 | $1.0 \times 10^{-1}$ | NA | 0.05 | 0.48 |
| - | - | rs9939609 | 53820527 | - | *FTO* | 0.13 | 0.94 | A | 0.47 | 0.41 | 0.0027 | $1.0 \times 10^{-1}$ | 1.00 | 0.10 | 0.40 |
| - | - | rs1121980 | 53809247 | - | *FTO* | 0.14 | 0.90 | T | 0.47 | 0.43 | 0.0021 | $1.9 \times 10^{-1}$ | NA | 0.26 | 0.21 |
| - | - | rs6499640 | 53769677 | - | *FTO* | 0.00 | 0.09 | A | 0.65 | 0.60 | −0.0013 | $4.6 \times 10^{-1}$ | 1.00 | 0.52 | 0 |
| 18q21.32 | rs6567160 | - | 57829135 | 1275 | *MC4R* | - | - | A | 0.81 | 0.75 | −0.0096 | $4.7 \times 10^{-6}$ | 0.98 | 0.03 | 0.52 |
| - | - | rs17782313 | 57851097 | - | *MC4R* | 0.06 | 0.99 | T | 0.72 | 0.77 | −0.0068 | $1.5 \times 10^{-4}$ | 1.00 | 0.50 | 0 |
| - | - | rs10871777 | 57851763 | - | *MC4R* | 0.05 | 0.97 | A | 0.71 | 0.75 | −0.0061 | $6.7 \times 10^{-4}$ | 1.00 | 0.40 | 0.04 |

(*Continued on next page*)

**Table 1. *Continued***

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| - | - | rs12970134 | 57884750 | - | *MC4R* | 0.24 | 0.77 | A | 0.14 | 0.28 | | 0.0065 | $5.9 \times 10^{-3}$ | 1.00 | 0.21 | 0.26 |
| - | - | rs571312 | 57839769 | - | *MC4R* | 0.03 | 0.99 | A | 0.34 | 0.25 | | 0.0009 | $6.1 \times 10^{-1}$ | NA | 0.99 | 0 |

Results for all 21 loci are shown in Table S2.
[a]Significance level: 0.05 divided by 866 (average number of SNPs across 21 BMI loci).
[b]CA: coded allele.
[c]CAF: coded allele frequency.
[d]Measurement of imputation accuracy, ranging from 0 (low) to 1 (high); NA indicates that there was no imputation in a subset of 6,326 WHI samples and that all other samples have directly genotyped data.
[e]Het p: heterogeneity test p value.
[f]SNP failed in quality control; SNP proxy (rs988748, LD $r^2 = 1.0$ in HapMap YRI) was substituted.
[g]SNP failed in genotyping in our EA population, and $r^2$ is calculated from 1000 Genomes Project European populations.

### Fine Mapping BMI Loci

Among the 21 BMI GWAS loci identified in EA studies, eight loci (*SEC16B* [MIM 612855], *TMEM18* [MIM 613220], *ETV5* [MIM 601600], *GNPDA2* [MIM 613222], *TFAP2B* [MIM 601601], *BDNF* [MIM 113505], *FTO* [MIM 610966], and *MC4R* [MIM 155541]) displayed SNPs with significant evidence of association (Table 1; see also Table S2). The lead SNP (the most significant SNP in AAs) in each of these eight loci had a minor-allele frequency >0.05 and showed little evidence of heterogeneity. In these eight loci most GWAS index SNPs previously identified in EA GWASs (20 out of 23) had a consistent direction of the effects reported in original EA studies. Because the results of the *FTO* locus have been described previously,[23] we focused on the other seven regions. Among those seven loci, the lead SNPs, rs543874 in *SEC16B* ($p = 1.5 \times 10^{-9}$), rs7647305 in *ETV5* ($p = 3.2 \times 10^{-5}$), and rs10938397 in *GNPDA2* ($p = 1.7 \times 10^{-7}$), were consistent with the previously identified GWAS SNP (the most significant SNP highlighted in the previous GWAS) in EA populations. For rs543874 and rs10938397, the observed effects on BMI were slightly stronger in AAs than in EA individuals (change in BMI per coded allele: 1.1% in AAs and 0.9% in EA individuals for rs543874; 1.0% in AAs and 0.8% in EAs for rs10938397[5]), whereas they were slightly weaker in AAs than in EA individuals for rs7647305 (change in BMI per coded allele: 0.7% in AAs and 0.9% in EA individuals[6]). The minor-allele frequency (MAF) was higher for rs543874 and rs7647305 and lower for rs10938397 in AAs than in EA individuals (Table 1). In the other four loci, the lead SNPs in AAs differed from the GWAS SNPs from EA populations. The lead SNPs in all four loci (rs6548240 in *TMEM18*, rs2744475 in *TFAP2B*, rs1519480 in *BDNF*, and rs6567160 in *MC4R*) were modestly to strongly correlated ($r^2$ ranged from 0.5–1.0) with at least one of the GWAS SNPs on the basis of LD in EA populations. However, when LD was based on AA populations, the correlation was weaker, in several cases substantially weaker (Table 1). For the 13 BMI loci that did not replicate in our AA analysis, most GWAS index SNPs (13 out of 17) from previous GWASs involving EA individuals showed effects in the same direction when results from our AA samples were compared with results from the previous EA studies.

We investigated the question of whether LD patterns in AA studies can narrow previous association signals from EA studies and found that at seven out of the eight significant loci, AA LD patterns assisted with narrowing association signals (Table 2; see also Figure S1). One of the most extreme examples was for *MC4R*. Among EA individuals, 107 and 119 SNPs were correlated ($r^2 > 0.5$) with the lead SNP in our analysis (rs6567160) and with the GWAS index SNPs, respectively; these SNPs represent a region spanning 184 kb and 230 kb, respectively. However, among AAs only five SNPs were correlated with rs6567160 at $r^2 > 0.5$, and these SNPs represented a region spanning 71 kb. In *SEC16B*, *ETV5*, *TFAP2B*, *BDNF*, and *FTO*, the number of

**Table 2. Comparison of Correlation between African American and European Populations for Eight BMI-Related Loci that had Significant[a] Lead SNPs**

| Region | Gene | Lead SNP in AAs of PAGE | Region Size | Number of SNPs | AA Populations | | EA Populations | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Number of SNPs with $r^2 > 0.5$ with Lead SNP | Region size for SNPs with $r^2 > 0.5$ with Lead SNP (bp) | Number of SNPs with $r^2 > 0.5$ with Lead SNP | Region Size for SNPs with $r^2 > 0.5$ with Lead SNP (bp) | Number of SNPs with $r^2 > 0.5$ with GWAS Index SNPs | Region size for SNPs $r^2>0.5$ with GWAS index SNPs (bp) |
| 1q25.2 | *SEC16B* | rs543874 | 180 kb | 765 | 2 | 4,242 | 39 | 95,660 | 39 | 95,660 |
| 2p25.3 | *TMEM18* | rs6548240 | 250 kb | 1,123 | 81 | 40,820 | 104 | 53,290 | 106 | 75,430 |
| 3q27.2 | *ETV5* | rs7647305 | 110 kb | 369 | 5 | 10,348 | 28 | 49,410 | 29 | 49,620 |
| 4p12 | *GNPDA2* | rs10938397 | 90 kb | 342 | 6 | 20,482 | 5 | 20,482 | 5 | 20,482 |
| 6p12.3 | *TFAP2B* | rs2744475 | 560 kb | 1,685 | 1 | 1,128 | 44 | 15,320 | 29 | 152,570 |
| 11p14.1 | *BDNF* | rs1519480 | 300 kb | 688 | 7 | 21,942 | 42 | 22,4200 | 42[b] | 224,200[b] |
| 16q12.2 | *FTO* | rs62048402 | 650 kb | 1,814 | 19 | 44,529 | 74 | 47,575 | 76[c] | 73,850[c] |
| 18q21.32 | *MC4R* | rs6567160 | 360 kb | 1,275 | 5 | 71,160 | 107 | 183,730 | 119 | 229610 |
| Average for number of SNPs or region size | | | | | 16 | 26,831 | 55 | 86,208 | 56 | 115,178 |

[a]Significance level: 0.05 divided by 866 (average number of SNPs across 21 BMI loci).
[b]Only included the GWAS index SNP rs925946 because the other two GWAS SNPs, rs10767664 and rs6265, have $r^2 < 0.15$ with the lead SNP and rs925946 in AA and EA populations.
[c]GWAS index SNP rs6499640 was not included because it has $r^2 < 0.1$ with the lead SNP and the other GWAS index SNPs in AA and EA populations.

SNPs correlated at $r^2>0.5$ with the lead SNP and the spanning region size were also substantially reduced and to a limited degree for *TMEM18* (Table 2). Only for *GNPDA2* did the number of correlated SNPs and the spanning region size not reduce, but the number of correlated SNPs in Europeans was already small.

To further refine the regions associated with BMI in AAs, we performed conditional analyses for each of the eight significant loci by including the lead SNP in a locus as a covariate to search for additional independent signals. rs186117327 in *GNPDA2* was borderline significantly associated with BMI when the locus-specific significance level was taken into account (Table 3). No evidence of heterogeneity was observed across studies. In *GNPDA2*, rs186117327 was significantly associated with BMI in the marginal analysis ($p = 3.7 \times 10^{-7}$),; this association became less significant but remained marginally so when adjustment was made for the lead SNP ($p = 1.7 \times 10^{-4}$). rs186117327 was weakly correlated with the lead SNP (rs10938397) as well as with the GWAS SNP at this locus ($r^2 = 0.11$ in AAs; 0.16 in EA individuals).

### Identification of Two BMI-Related Loci in AAs by Metabochip-wide Analysis

In the Metabochip-wide analysis (excluding the SNPs in the 21 BMI-related loci), we identified five SNPs in two loci (these SNPs were rs116612809, rs114584581, rs74941130, and rs79329695 in 2p23.2/*BRE* [MIM 610497] and rs4802349 in 19q13.32/*DHX34*) as being associated with BMI in AAs at a Metabochip-wide significance level ($p < 2.5 \times 10^{-7}$). Two of the SNPs in 2p23.2/*BRE* (rs116612809 and rs79329695) reached the conventional genome-wide significance level ($p < 5.0 \times 10^{-8}$) (Table 4 and Figure 1). Furthermore, with $p = 3.7 \times 10^{-7}$, rs57813622 at the 7p21.2/*DGKB* [MIM 604070] locus approached the Metabochip-wide significance level.

In 2p23.2/*BRE*, all four SNPs are rare variants in Europeans (MAF = 0.1% in the 1000 Genomes Project) but common in AAs (MAF = 10%). The most significant SNP, rs116612809, is located in the intronic region of *BRE* and is highly correlated with the three other significant SNPs ($r^2$ ranges from 0.98–1.00, Figure 2). In the conditional analysis for all other SNPs at the *BRE* locus, we did not observe evidence for an additional independent signal associated with BMI after conditioning on rs116612809, suggesting that all four SNPs point to the same functional variant. The 2p23.2/*BRE* was included on the Metabochip as a part of a ~1.3-Mb-long region so that a GWAS locus associated with blood triglyceride (TG) concentrations could be fine-mapped.[24,25] Within this region, the four BMI-associated SNPs were more than 500 kb away from the TG GWAS index SNPs (rs1260326, rs1260333, rs780093, and rs780094; $r^2$ ranged from 0.02–0.05 in AAs). To examine the association between rs116612809 and blood TG concentrations in our population, we used ARIC and WHI, for which the TG individual data were available (n = 11,680). We did not observe a significant

**Table 3. Conditional Analysis Showing a Locus with Evidence of Second Independent BMI Association Signals in African Americans**

| Region | Gene | SNP | Position | Number of Tested SNPs | Coded Allele | CAF | Marginal Results[a] | | | Conditional Results[b] | | | $r^2$ with Lead SNP in AAs |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Effect | p Value | Het p[d] | Effect | p Value | Het p | |
| 4p12 | *GNPDA2* | rs186117327 | 45,101,187 | 285 | A | 0.75 | −0.0095 | $3.7 \times 10^{-7}$ | 0.54 | −0.0074 | $1.7 \times 10^{-4}$ | 0.91 | 0.11 |
| - | - | rs10938397[c] | 45,182,527 | - | A | 0.75 | −0.0099 | $1.7 \times 10^{-7}$ | 0.24 | −0.0068 | $1.6 \times 10^{-3}$ | 0.37 | - |

[a]Marginal results represent results when only the single variant was in the model.
[b]Conditional analysis represents the result for the SNP when adjustment was made for the most significant lead SNP and vice versa.
[c]The most significant SNP in a locus.
[d]Het p: heterogeneity-test p value.

**Table 4. Two BMI Loci Identified in African Americans on the Basis of Metabochip-wide Analysis**

| Region | SNP | Position | Candidate Gene | Coded Allele | CAF in AAs | CAF in EA Individuals | Effect | p value | Rsq[a] | Het p | $r^2$ with Lead SNP in AAs |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2p23.2 | rs116612809 | 28,301,171 | *BRE* | A | 0.90 | 0.001 | −0.0151 | $3.6 \times 10^{-8}$ | 0.99 | 0.48 | - |
| - | rs114584581 | 28,304,380 | *BRE* | A | 0.10 | 0.001 | 0.0148 | $5.9 \times 10^{-8}$ | 0.99 | 0.43 | 0.98 |
| - | rs74941130 | 28,306,293 | *BRE* | A | 0.10 | 0.001 | 0.0148 | $6.9 \times 10^{-8}$ | 0.99 | 0.44 | 1.00 |
| - | rs79329695 | 28,319,874 | *BRE* | A | 0.10 | 0.001 | 0.0151 | $3.7 \times 10^{-8}$ | 0.99 | 0.56 | 0.99 |
| 19q13.32 | rs4802349 | 47,874,510 | *DHX34* | A | 0.48 | 0.12 | −0.0087 | $1.2 \times 10^{-7}$ | 1.00 | 0.46 | - |

[a]Measurement of imputation accuracy, ranging from 0 (low) to 1 (high).

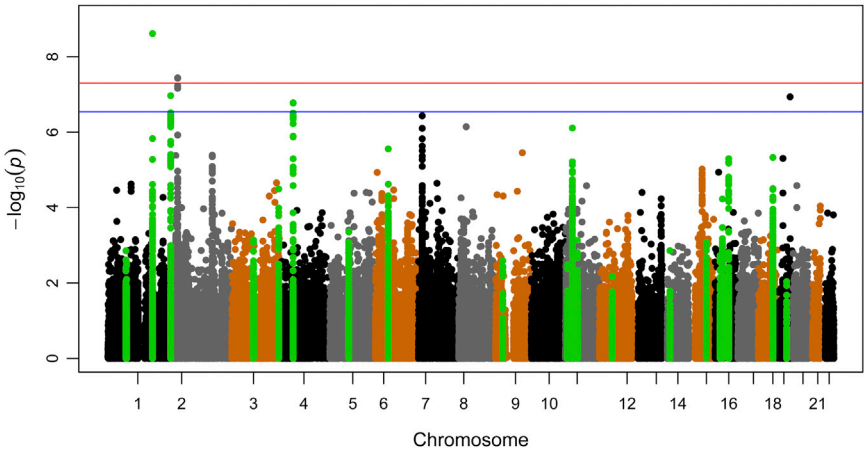


**Figure 1. Manhattan Plot for Metabochip-Wide Analysis of BMI in African Americans**

The –log10 of p values for each SNP on the Metabochip is plotted against physical chromosomal positions. Green dots represent the SNPs in 21 BMI-related loci that were previously identified in European populations and fine-mapped on the Metabochip. Blue line: $p = 2.5 \times 10^{-7}$; red line: $p = 5 \times 10^{-8}$.

association (p = 0.71) with blood TG concentrations, which did not change after adjustment for BMI. Furthermore, the association between rs116612809 and BMI did not change after adjustment for TG (Table S4).

In 19q13.32/*DHX34*, rs4802349 is located in an intron of *DHX34* and has a MAF of 0.48 in AAs and 0.12 in EA individuals. Only ten SNPs surrounding rs4802349 (±200 kb) were genotyped; none was correlated with rs4802349 in Aas, and none was associated with BMI. Previously, rs4802349 was reported to be a putative association signal for high-density lipoprotein (HDL) cholesterol at a moderate p value of 0.005.[26] We examined the association between rs4802349 and blood HDL concentrations with and without BMI adjustment in ARIC and WHI (n = 11,680). We observed that the p value changed from 0.25 to 0.04 after BMI adjustment. Furthermore, the association between rs4802349 and BMI did not change after adjustment for HDL (Table S4).

### Sex-Stratification Analyses

We did not observe any additional loci when we stratified the analysis by sex (Tables S5 and S6 and Figures S2 and S3). Also, the results for the SNPs that were significant in the fine-mapping analyses and the two BMI loci results were consistent across sex (p-heterogeneity ≥ 0.29, Table S7). However, as a result of the relatively small sample of men in this study, we have limited power to detect a difference-of-sex effect.

### Functional Annotation

In silico analysis assessed whether each BMI-associated locus was in an "open," transcriptionally permissive conformation, as would be expected of a functional locus. Using ENCODE data sets, we found each signal to be in a region consistent with regulatory evidence, such as active histone marks, open chromatin structure (DNase hypersensitivity), or regions experimentally shown to bind one or more transcription factors (Table S8). Bioinformatics analyses revealed that the lead SNP, rs6548240 in the *TMEM18* locus, was the strongest functional candidate. rs6548240 is in a region of open chromatin that binds multiple transcription factors, and elevated levels of active histone marks associated with promoters were detected in several cell lines (Figure S4). The functional properties of TMEM18 are obscure, although a recent study indicated that TMEM18 plays a regulatory role in adipocyte differentiation and biology.[27] For *GNPDA2*, the potential additional independent signal rs186117327 could be tagging rs7659184 ($r^2 = 0.6$ with rs186117327 in AAs), which falls in a region of open chromatin, and relative to the reference allele, the alternate allele reduces the binding affinity of GATA2 and EN1 transcription factors. In 2p23.2/*BRE*, bioinformatics analyses indicated that the lead SNP, rs116612809, tags another intronic SNP, rs78003529 ($r^2 = 0.65$ in AAs), which falls in a region of open chromatin with elevated enhancer histone marks and is in a region that binds both Pol2 and c-Jun. Furthermore, evidence from scans of positional weight matrices (PWMs) suggests that the alternate allele of rs78003527 has a much higher binding affinity for Nkx3 and Nkx2. In 19q13.32/*DHX34*, the *DHX34* signal rs4802349 and two SNPs highly correlated with it are located within putative regulatory regions. rs4802349 falls within histone-modification marks associated with enhancer activity and alters the Mtf1 transcription-factor binding motif. In addition, two other functional candidates, rs2547369 and rs2341878 ($r^2 = 0.7$ and 0.6, respectively in African populations) were highly correlated with this signal and fall in regions of open chromatin having methylation patterns associated with promoter or enhancer activity. Furthermore, rs25476369 ($r^2 = 0.65$ with rs4802349 in African populations) falls in a region that binds six transcription factors, whereas rs2341878 is of particular interest because it falls in a strong promoter region in a number of relevant tissues. The function of GNPDA2, BRE, and DHX34 is unknown, which makes it difficult to link the underlying biological mechanisms of these genetic variants to BMI.

## Discussion

In this study, encompassing close to 30,000 AAs, we used the Metabochip to systematically evaluate 21 BMI loci

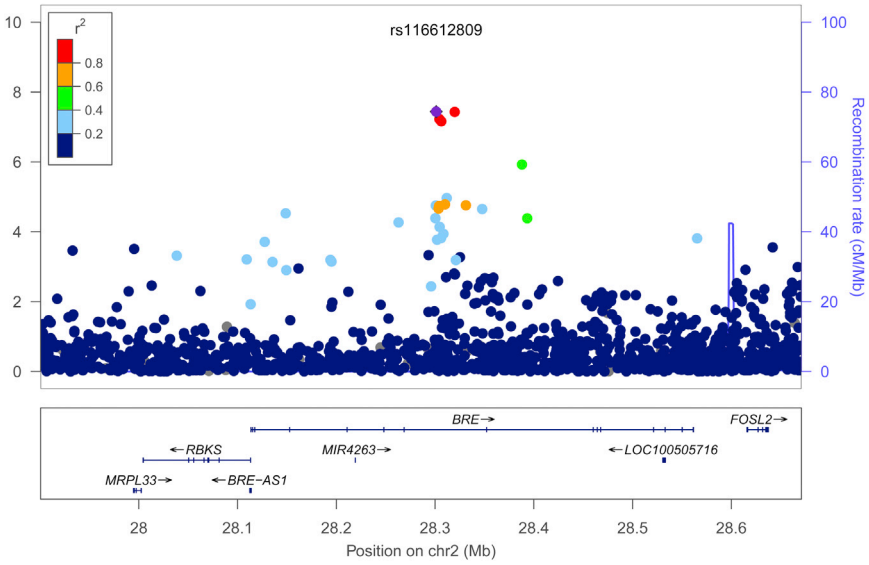


**Figure 2. Regional-Association Plot for the New BMI-Related Locus at 2p23.2/*BRE***
The –log10 of p values (left *y* axis) is plotted against the SNP genomic position based on NCBI build 37 (*x* axis); the estimated recombination rate from the 1000 Genomes Project for African populations is on the right *y* axis and is plotted in blue. The most significant SNP is denoted with a purple diamond. SNPs are colored to reflect correlation with the most significant SNP. Gene annotations are from the UCSC Genome Browser.

discovered among European descent populations in previous GWASs and to search for potential BMI loci. Eight of the 21 loci (*SEC16B*, *TMEM18*, *ETV5*, *GNPDA2*, *TFAP2B*, *BDNF*, *FTO*, and *MC4R*) were found to be associated with BMI in our AA study population. Further conditional analyses indicated that *GNPDA2* contained an additional independent signal. Moreover, Metabochip-wide analyses revealed two BMI-associated loci: the *BRE* locus at genome-wide significance ($p < 5 \times 10^{-8}$) and the *DHX34* locus at Metabochip-wide significance ($p < 2.5 \times 10^{-7}$).

Among the eight BMI loci that were significant in the AA population, the lead SNPs in *SEC16B, ETV5*, and *GNPDA2* were the same as the GWAS SNPs identified in previous EA studies, which provided further support for the idea that the three SNPs (rs543874, rs7647305, and rs10938397) are proxies of causal variants influencing BMI. The most significant (lead) SNPs in the other five loci differed from the previously reported GWAS SNPs, although, as expected, all were moderately to highly correlated with the GWAS SNPs in European populations ($r^2$ ranged from 0.5 to 1.0). In *TFAP2B* and *MC4R*, the GWAS SNPs were not significantly associated with BMI in AAs, suggesting that the lead SNPs in these two loci are better proxies for the underlying functional variants. We showed that the weaker LD patterns in AA populations than in EA populations substantially reduced the number of functional-variant proxies in six of the eight BMI loci that were significant in AAs (and to a limited extent in *TMEM18*). Our results illustrate the important contribution of AAs to systematic fine-mapping of GWAS loci originally reported in EA populations. In addition, the lead SNPs that were either consistent or correlated with the GWAS SNPs at these replicated loci in EA populations might indicate that EA populations and AA populations share the underlying causal variants at these loci.

Bioinformatics analyses revealed that the lead SNP, rs6548240 in the *TMEM18* locus, was the strongest functional candidate. Although rs6548240 is located in an intergenic region 31 kb downstream of *TMEM18*, ChIP-seq evidence indicates that CTCF looping might anchor this distant putative enhancer to the promoter of *TMEM18*. Taken together, these pieces of evidence make rs6548240 an interesting functional candidate for future laboratory follow-up. This example shows that combining fine mapping with bioinformatics analysis can help to narrow in on the putative functional variants for further follow-up studies.

There are multiple reasons that 13 BMI loci originally identified in EA populations were not significantly associated with BMI in this sample of AAs. Limited statistical power could be an important reason for this observation. Statistical power is impacted by the variance of BMI, MAF, effect size, and sample size. Compared to populations in some large European-focused studies, our AA population had larger variance in BMI[7] (standard deviation: 6.2 kg/m$^2$ versus 4.2 kg/m$^2$), which reduces the statistical power. Also compared with primarily very large European-focused studies with sample sizes from ~90,000 to ~250,000,[5–7] our study was relatively small. Nonreplication might arise because of different causal variants between EA individuals and AAs, a weak AA LD pattern, which leads to weak correlation between causal variants and marker SNPs on the Metabochip, and limited statistical power. All of this might explain that we did not observe all loci significantly associated with BMI in AAs, and it emphasizes the need for a larger sample size. We used a uniform p value threshold of $5.8 \times 10^{-5}$ (0.05/average number of SNPs per locus) as an approximate correction for the average of 866 SNPs across these loci. However, if we use Bonferroni correction (i.e., 0.05/the number of SNPs at a given locus) or correction by the effective number of SNPs at each locus after accounting for LD patterns, one more locus, *SH2B1* [MIM 608937], would be indicated to be significant in AAs (Table S2).

Conditional analysis indicated that the *GNPDA2* locus contained a potential additional independent signal. The potential additional independent signal, rs186117327, is physically relatively close (80 kb) to the lead SNP,

rs10938397, but is only weakly correlated ($r^2 = 0.11$) with it.

We identified two BMI-related loci, 2p23.2/*BRE* and 19p13.32/*DHX34*, in the Metabochip-wide analysis. We evaluated whether African ancestry plays role in these two BMI-related loci among 8,310 AAs from a GWAS in WHI. We didn't observe a significant effect of ancestry at these two loci ($p > 0.5$; Table S9). At 2p23.3/*BRE*, the most significant SNP, rs116612809, reached the conventional genome-wide significance level ($5 \times 10^{-8}$) and was surrounded (3–18 kb) by three highly correlated SNPs ($r^2 = 0.98$–$1.00$) showing very similar results. All four SNPs are common (MAF = 0.1), but they all are rare variants in EA populations (MAF = 0.001)[28] and hence it is unlikely that these SNPs would be identified in EA GWASs unless they have much stronger effects than the moderate effects observed in our study or are tested in very large sample sets. The lead SNP in 2p23.3/*BRE* (rs116612809) is located within an intron of *BRE*, which is stress responsive and highly expressed in brain and reproductive organs.[29] No previous study reported an association with BMI or any other trait; however, in support of our finding, the GIANT consortium reported that 82 of 200 SNPs within *BRE* had a p value $< 0.05$ (min. $p = 2.5 \times 10^{-4}$) for a BMI association in EA populations.[5] Accordingly, our results showed that varying allele frequency in different ancestral groups significantly contributes to the statistical power and that studying different ancestral groups helps to identify potentially functional loci.[30]

We found another potential BMI-related locus at 19q13.32/*DHX34*. The SNP rs4802349, located in an intron *DHX34*, was marginally significantly associated with BMI at a Metabochip-wide significance level ($p = 1.2 \times 10^{-7}$), but because the finding did not reach the conventional genome-wide significance level of $5 \times 10^{-8}$, additional replication studies in AAs are warranted. *DHX34* is a putative RNA helicase and has not been reported to be associated with any diseases or traits in humans before, except for a suggestive association with HDL cholesterol ($p = 0.005$).[26] A recent GWAS in EA populations showed an association with BMI in a neighboring region, 19q13.32/*ZC3H4*; rs3810291 was the most significant SNP.[5] However, rs4802349 is physically 300 kb away from rs3810291, and a recombination hotspot lies between *ZC3H4* and *DHX34* in AAs and Europeans (HapMap phase II YRI and CEU), resulting in a low correlation between both SNPs in African populations ($r^2 = 0.04$) and Europeans ($r^2 < 0.001$). This observation suggests that these two associations are independent from each other. Our results along with the evidence from functional annotation warrant additional validation studies for this locus.

In conclusion, we observed that eight BMI-associated GWAS loci identified from EA populations were significantly associated with BMI in AAs and identified a potential additional independent signal in one locus. In addition, we discovered two potential BMI-related loci through Metabochip-wide analysis; one of these loci which reached the conventional genome-wide significance level. Importantly, our study demonstrated that fine mapping in AA populations in combination with bioinformatics analyses is a valuable and effective way to narrow in on the underlying causal variants in GWAS loci discovered in EA populations and that studying minority populations can contribute to loci discovery.

## Supplemental Data

Supplemental data include four figures, nine tables, and supplemental text, including acknowledgments and can be found with this article online at http://www.cell.com/AJHG/.

## Acknowledgments

See Supplemental Data.

Received: May 6, 2013
Revised: August 7, 2013
Accepted: August 16, 2013
Published: October 3, 2013

## Web Resources

The URLs for data presented herein are as follows:

Center for Disease Control and Prevention (CDC), Adult Obesity Facts, http://www.cdc.gov/obesity/data/adult.html
GIANT consortium data files, http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files
MACH, http://www.sph.umich.edu/csg/abecasis/mach
Minimac, http://genome.sph.umich.edu/wiki/Minimac
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org
The Population Architecture using Genomics and Epidemiology (PAGE) Study, http://www.pagestudy.org
World Health Organization (WHO) (2011) Obesity and Overweight Fact Sheet, http://www.who.int/mediacentre/factsheets/fs311/en/index.html

## References

1. Kopelman, P.G. (2000). Obesity as a medical problem. Nature *404*, 635–643.
2. Miller, W.M., Nori-Janosz, K.E., Lillystone, M., Yanez, J., and McCullough, P.A. (2005). Obesity and lipids. Curr. Cardiol. Rep. *7*, 465–470.
3. Maes, H.H., Neale, M.C., and Eaves, L.J. (1997). Genetic and environmental factors in relative body weight and human adiposity. Behav. Genet. *27*, 325–351.
4. Hjelmborg, Jv., Fagnani, C., Silventoinen, K., McGue, M., Korkeila, M., Christensen, K., Rissanen, A., and Kaprio, J. (2008). Genetic influences on growth traits of BMI: a longitudinal study of adult twins. Obesity (Silver Spring) *16*, 847–852.
5. Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G., Jackson, A.U., Lango Allen, H., Lindgren, C.M., Luan, J., Mägi, R., et al.; MAGIC; Procardis Consortium. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat. Genet. *42*, 937–948.

6. Thorleifsson, G., Walters, G.B., Gudbjartsson, D.F., Steinthorsdottir, V., Sulem, P., Helgadottir, A., Styrkarsdottir, U., Gretarsdottir, S., Thorlacius, S., Jonsdottir, I., et al. (2009). Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat. Genet. *41*, 18–24.
7. Willer, C.J., Speliotes, E.K., Loos, R.J., Li, S., Lindgren, C.M., Heid, I.M., Berndt, S.I., Elliott, A.L., Jackson, A.U., Lamina, C., et al.; Wellcome Trust Case Control Consortium; Genetic Investigation of ANthropometric Traits Consortium. (2009). Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nat. Genet. *41*, 25–34.
8. McCarthy, M.I., and Hirschhorn, J.N. (2008). Genome-wide association studies: potential next steps on a genetic journey. Hum. Mol. Genet. *17*(R2), R156–R165.
9. Voight, B.F., Kang, H.M., Ding, J., Palmer, C.D., Sidore, C., Chines, P.S., Burtt, N.P., Fuchsberger, C., Li, Y., Erdmann, J., et al. (2012). The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. PLoS Genet. *8*, e1002793.
10. Matise, T.C., Ambite, J.L., Buyske, S., Carlson, C.S., Cole, S.A., Crawford, D.C., Haiman, C.A., Heiss, G., Kooperberg, C., Marchand, L.L., et al.; PAGE Study. (2011). The Next PAGE in understanding complex traits: design for the analysis of Population Architecture Using Genetics and Epidemiology (PAGE) Study. Am. J. Epidemiol. *174*, 849–859.
11. Connor Gorber, S., and Tremblay, M.S. (2010). The bias in self-reported obesity from 1976 to 2005: a Canada-US comparison. Obesity (Silver Spring) *18*, 354–361.
12. Weale, M.E. (2010). Quality control for genome-wide association studies. Methods Mol. Biol. *628*, 341–372.
13. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet. *2*, e190.
14. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.
15. Buyske, S., Wu, Y., Carty, C.L., Cheng, I., Assimes, T.L., Dumitrescu, L., Hindorff, L.A., Mitchell, S., Ambite, J.L., Boerwinkle, E., et al. (2012). Evaluation of the metabochip genotyping array in African Americans and implications for fine mapping of GWAS-identified loci: the PAGE study. PLoS ONE *7*, e35651.
16. Liu, E.Y., Buyske, S., Aragaki, A.K., Peters, U., Boerwinkle, E., Carlson, C., Carty, C., Crawford, D.C., Haessler, J., Hindorff, L.A., et al. (2012). Genotype imputation of Metabochip SNPs using a study-specific reference panel of ~4,000 haplotypes in African Americans from the Women's Health Initiative. Genet. Epidemiol. *36*, 107–117.
17. Fesinmeyer, M.D., North, K.E., Ritchie, M.D., Lim, U., Franceschini, N., Wilkens, L.R., Gross, M.D., Bůžková, P., Glenn, K., Quibrera, P.M., et al. (2013). Genetic risk factors for BMI and obesity in an ethnically diverse population: results from the population architecture using genomics and epidemiology (PAGE) study. Obesity (Silver Spring) *21*, 835–846.
18. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.
19. Wu, Y., Waite, L.L., Jackson, A.U., Sheu, W.H., Buyske, S., Absher, D., Arnett, D.K., Boerwinkle, E., Bonnycastle, L.L., Carty, C.L., et al. (2013). Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. PLoS Genet. *9*, e1003379.
20. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
21. Berglund, G., Elmstähl, S., Janzon, L., and Larsson, S.A. (1993). The Malmo Diet and Cancer Study. Design and feasibility. J. Intern. Med. *233*, 45–51.
22. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics *26*, 2336–2337.
23. Peters, U., North, K.E., Sethupathy, P., Buyske, S., Haessler, J., Jiao, S., Fesinmeyer, M.D., Jackson, R.D., Kuller, L.H., Rajkovic, A., et al. (2013). A systematic mapping approach of 16q12.2/FTO and BMI in more than 20,000 African Americans narrows in on the underlying functional variation: results from the Population Architecture using Genomics and Epidemiology (PAGE) study. PLoS Genet. *9*, e1003171.
24. Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T., et al. (2009). Common variants at 30 loci contribute to polygenic dyslipidemia. Nat. Genet. *41*, 56–65.
25. Aulchenko, Y.S., Ripatti, S., Lindqvist, I., Boomsma, D., Heid, I.M., Pramstaller, P.P., Penninx, B.W., Janssens, A.C., Wilson, J.F., Spector, T., et al.; ENGAGE Consortium. (2009). Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat. Genet. *41*, 47–55.
26. Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707–713.
27. Bernhard, F., Landgraf, K., Klöting, N., Berthold, A., Büttner, P., Friebe, D., Kiess, W., Kovacs, P., Blüher, M., and Körner, A. (2013). Functional relevance of genes implicated by obesity genome-wide association study signals for human adipocyte biology. Diabetologia *56*, 311–322.
28. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A.; 1000 Genomes Project Consortium. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.
29. Gu, C., Castellino, A., Chan, J.Y., and Chao, M.V. (1998). BRE: A modulator of TNF-alpha action. FASEB J. *12*, 1101–1108.
30. Pulit, S.L., Voight, B.F., and de Bakker, P.I. (2010). Multiethnic genetic association studies improve power for locus discovery. PLoS ONE *5*, e12600.