A Meta-Analysis Identifies New Loci Associated with Body Mass index in Individuals of African Ancestry

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation

Published Version
doi:10.1038/ng.2608

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:11879226

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
A Meta-Analysis Identifies New Loci Associated with Body Mass Index in Individuals of African Ancestry


1These authors jointly directed the study and are to whom correspondence should be addressed: Christopher Haiman at Department of Preventive Medicine, USC/Norris Comprehensive Cancer Center, Harlyne Norris Research Tower, 1450 Biggby Street, Room 1504, Los Angeles, CA 90033, USA. Tel: +1 3234427755, Fax: +1 3234427749, haiman@usc.edu. Kari North at Department of Epidemiology, UNC Gillings School of Global Public Health, and Carolina Center for Genome Sciences, 137 E Franklin St., Suite 306, Chapel Hill, NC 27514, USA. Tel: +1 9199662148, Fax: +1 9199669800, kari_north@unc.edu.

†A list of contributing members appears in the Supplementary Note.

∗These authors contributed equally.

†In Memoriam

Author Contributions


Genotyping: AB, UB, SJC, YIC, DD, SFAG, XG, DGH, JHH, JNH, TDH, TH, KCI, YL, YCL, WM, RN, JRP, NDP, SS, DJV

Phenotyping: AAA, DKA, MAA, EPB, RSC, ED, BIP, OG, SFAG, JH, TH, KCI, AK, CK, EKK, SL, JEM, MN, RN, AO, HO-B, BMP, JRP, SRP, CNR, ER, SB, DS, LS, BOT, TRY

Statistical Methods and Data analysis: AAA, DKA, LFB, CBWK, GK, GC, NEC, WC, GAC, YIC, JD, PD, TLE, CF, MFF, JPB, EMG, MG, OG, XG, HAO, MCI, MR, BJK, CK, EKK, SJK, CML, GL, G Li, HL, KL, LAL, RJFL, VL, YL, Youshang Liu, YCL, KL, BM, KLM, YAM, AN, KEN, MAN, MCYN, AND, CDF, JRP, EAR, SKR, BP, APR, LJR-T, DAS, EKS, E Schad, YVS, BOT, KCT, DRV-E, MKW, ZW, LKW, TW, LRY, LLY, JHZ, NAZ, JZ, JMG, WZ

Writing Group: GKC, TLE, MG, BEH, JNH, RFL, CAH, LAL, KLM, KEN, MCYN, CP, GP, APR, KCT


1The Center for Observational Research, Arogen, Inc. Thousand Oaks, CA, USA
2Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
3Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
4Department of Epidemiology and Population Health, University of Louisville, Louisville, KY, USA
5Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA
6Division of Genetics, Children’s Hospital, Boston, MA, USA
7Division of Endocrinology, Children’s Hospital, Boston, MA, USA
8Center for Human Genetics Research, Vanderbilt Epidemiology Center, Department of Medicine, Vanderbilt University, Nashville, TN, USA

Nat Genet. Author manuscript; available in PMC 2013 December 01.
9Department of Genetics, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC, USA
10Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA
11Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, USA
12Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA
13Department of Family and Preventive Medicine, University of California San Diego, La Jolla, CA, USA
14Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA
15Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA
16Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, USA
17Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA
18Department of Epidemiology & Prevention, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA
19Thurston Arthritis Research Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
20The Charles Bronfman Institute of Personalized Medicine, The Icahn School of Medicine at Mount Sinai, New York, NY, USA
21Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
22Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA
23Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA
24Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
25Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN, USA
26The Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA
27Department of Thoracic Surgery, Vanderbilt University School of Medicine, Nashville, TN, USA
28Department of Surgery, College of Medicine, University of Ibadan, Ibadan, Nigeria
29Department of Epidemiology, Division of Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
30Department of Family and Community Medicine, Geissel School of Medicine, Dartmouth, Hanover, NH
31The Cancer Institute of New Jersey, New Brunswick, NJ, USA
32Karmanos Cancer Institute, Wayne State University, Detroit, MI, USA
33Department of Pediatrics, Section of Genomic Pediatrics, Medical College of Wisconsin, Milwaukee, WI, USA
34Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
35Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
36The Translational Genomics Research Institute, Phoenix, AZ, USA
37Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA
38Medical Genetics Institute, Cedars Sinai Medical Center, Los Angeles, CA, USA
39Department of Genetics, Harvard Medical School, Boston, MA, USA
40Department of Urology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
41Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN, USA
42Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA
43South Central VA Mental Illness, Research, and Clinical Center, Little Rock, AR, USA
44VA Geriatric Research, Education, and Clinical Center, Little Rock, AR, USA
45Division of Health Services Research, Department of Psychiatry, University of Arkansas for Medical Sciences, Little Rock, AR, USA
46Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA
47Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health Bethesda, MD, USA
48Laboratory of Epidemiology and Population Science, National Institutes on Aging, National Institutes of Health, Bethesda, MD, USA
49Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
50Department of Preventive Medicine, Stony Brook University, Stony Brook, NY, USA
51Chronic Disease Research Centre, University of the West Indies, Bridgetown, Barbados
52Faculty of Medical Sciences, University of the West Indies, Bridgetown, Barbados
53Ministry of Health, Bridgetown, Barbados
54Department of Health Disparities Research, Division of OVP, Cancer Prevention and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
55Center for Community-Engaged Translational Research, Duncan Family Institute, The University of Texas MD Anderson Cancer Center, Houston, TX, USA
56Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, USA
57MedStar Health Research Institute, Washington, DC, USA
58Georgetown-Howard Universities Center for Clinical and Translational Sciences, Washington, DC, USA

Nat Genet. Author manuscript; available in PMC 2013 December 01.
Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, TN, USA

Henri Begleiter Neurodynamics Laboratory, Department of Psychiatry and Behavioral Sciences, SUNY Downstate Medical Center, Brooklyn, NY, USA

Institute of Translational Medicine and Therapeutics, University of Pennsylvania, Philadelphia, PA, USA

Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh PA, USA

Sickle Cell Center, Department of Medicine, Georgia Health Sciences University, Augusta, GA, USA

Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA

Montreal Heart Institute, Université de Montréal, Montréal, Québec, Canada

Program in Genomics and Endocrinology, Division of Genetics, Children’s Hospital, Boston, MA, USA

Department of Pediatrics, Harvard Medical School, Boston, MA, USA

Department of Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA

Regional Center for Neurosensory Disorders, School of Dentistry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Broad Institute of Harvard and MIT, Metabolic Disease Initiative, Cambridge, MA, USA

Center for Population and Reproductive Health, College of Medicine, University of Ibadan, Ibadan, Nigeria

Department of Urology, Northwestern University, Chicago, IL, USA

Department of Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Department of Public Health Sciences, Henry Ford Hospital, Detroit, MI, USA

Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA

Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, USA

Department of Social and Preventive Medicine, University at Buffalo, Buffalo NY, USA

Department of Medicine, University of Ibadan, Ibadan, Nigeria

Department of Medicine, University of Chicago, Chicago, IL, USA

Slone Epidemiology Center at Boston University, Boston MA, USA

Department of Pathology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

Department of Human Genetics, Miller School of Medicine, University of Miami, FL, USA

Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
85Department of Epidemiology and Public Health, University of Miami Miller School of Medicine, Miami, FL, USA
86Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL, USA
87Department of Genetics and Genomic Sciences, The Icahn School of Medicine at Mount Sinai, New York, NY, USA
88Icahn Institute for Genomics and Multiscale Biology, The Icahn School of Medicine at Mount Sinai, New York, NY, USA
89Department of Epidemiology, University of Washington, Seattle, WA, USA
90Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA
91Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, MI, USA
92Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA
93Division of Gastroenterology. Massachusetts General Hospital, Boston, MA, USA
94The Broad Institute, Cambridge, MA, USA
95Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX, USA
96Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY, USA
97Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA
98Department of Preventive Medicine and Epidemiology, Loyola University Chicago Stritch School of Medicine, Maywood, IL, USA
99Center for Human Genetics Research, Vanderbilt Epidemiology Center, Vanderbilt University, Nashville, TN, USA
100Vanderbilt Epidemiology Center, Department of Obstetrics and Gynecology, Vanderbilt University, Nashville, TN, USA
101Core Genotype Facility, SAIC-Frederick, Inc., National Cancer Institute-Frederick, Frederick, MD, USA
102Institute for Human Genetics, University of California San Francisco
103Department of Neurological Surgery, University of California San Francisco
104Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany
105Institute for Human Genetics, Departments of Epidemiology and Biostatistics and Urology, University of California, San Francisco, San Francisco, CA, USA
106Department of Medicine, University of Vermont College of Medicine, Burlington, VT, USA
107MRC Epidemiology Unit, Institute of Metabolic Science, Addenbrooke’s Hospital, Cambridge, UK
108Department of Epidemiology, School of Public Health, University of Michigan, MI, USA
109Department of Medicine, Medical University of South Carolina, Charleston, SC, USA
110College of Nursing, Medical University of South Carolina, Charleston, SC, USA
Framingham Heart Study, Boston University School of Medicine, Boston, MA, USA
Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
Division of Cancer Etiology, Department of Population Science, Beckman Research Institute, City of Hope, Duarte, CA, USA
International Epidemiology Institute, Rockville, MD, USA
Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA
Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC, USA
Section on Gerontology and Geriatric Medicine, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA
Health Disparities Research Section, Laboratory of Population Science, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA
National Heart, Lung and Blood Institute’s Framingham Heart Study, Framingham, MA, USA
National Heart, Lung and Blood Institute’s Center for Population Studies, Framingham, MA, USA
Division of Endocrinology, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA, USA
Department Nutrition Sciences, University of Alabama at Birmingham and the Birmingham VA Medical Center, Birmingham, AL, USA
The Center for Applied Genomics, The Children’s Hospital of Philadelphia Research Institute, The Children’s Hospital of Philadelphia, PA, USA
Division of Human Genetics, The Children’s Hospital of Philadelphia Research Institute, The Children’s Hospital of Philadelphia, PA, USA
Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA
Institute of Diabetes, Obesity and Metabolism, Perelman School of Medicine at the University of Pennsylvania, Philadelphia PA, USA
Cancer Prevention Institute of California, Fremont, CA, USA
Division of Epidemiology, Department of Health Research & Policy, Stanford University School of Medicine, Stanford, CA, USA
Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA
Department of Health Studies, University of Chicago, IL, USA
Department of Medicine, University of Illinois at Chicago, Chicago, IL, USA
SWOG Statistical Center, Seattle, WA, USA
Glickman Urologic and Kidney Institute, Cleveland Clinic, Cleveland, OH
Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA
Department of Epidemiology, School of Public Health, University of California at Los Angeles, Los Angeles, CA, USA
Department of Biostatistics, University of Washington, Seattle, WA, USA
Abstract

Genome-wide association studies (GWAS) have identified 36 loci associated with body mass index (BMI), predominantly in populations of European ancestry. We conducted a meta-analysis to examine the association of >3.2 million SNPs with BMI in 39,144 men and women of African ancestry, and followed up the most significant associations in an additional 32,268 individuals of African ancestry. We identified one novel locus at 5q33 (GALNT10, rs7708584, $p=3.4 \times 10^{-11}$) and another at 7p15 when combined with data from the Giant consortium (MIR148A/NFE2L3, rs10261878, $p=1.2 \times 10^{-10}$). We also found suggestive evidence of an association at a third locus at 6q16 in the African ancestry sample (KLHL32, rs974417, $p=6.9 \times 10^{-8}$). Thirty-two of the 36 previously established BMI variants displayed directionally consistent effect estimates in our GWAS (binomial $p=9.7 \times 10^{-7}$), of which five reached genome-wide significance. These findings provide strong support for shared BMI loci across populations as well as for the utility of studying ancestrally diverse populations.
There are notable racial and ethnic disparities in the prevalence of obesity in the United States; nearly 50% of African American adults are classified as obese compared to 35% of non-Hispanic whites. Genome-wide association studies (GWAS) have identified 36 BMI loci at statistically significant levels (p<5.0×10^{−8})\(^2–13\), of these, 32 were identified in individuals of European ancestry\(^3–8\) and four in East Asian populations\(^9,10\). Large GWAS of BMI in populations of African ancestry are lacking, and will be important for identifying genetic variants that are unique or of greater importance to this population\(^14–17\). In this study, we conducted a large GWAS meta-analysis of BMI in men and women of African ancestry to search for novel loci, and tested associations with common variation at the 36 known loci to better understand their relevance in African ancestry populations.

Thirty-six GWAS, totaling 39,144 men and women of African ancestry, were included in the Stage 1 meta-analysis of as many as 3,283,202 (minor allele frequency >1%) genotyped and imputed single nucleotide polymorphisms (SNPs) (Online Methods, Supplementary Tables 1–3, Supplementary Note). After applying both study-specific and overall Stage 1 genomic-control corrections (Supplementary Table 2), 11 SNPs at five loci achieved genome-wide significance (p<5×10^{−8}) (Table 1, Figure 1, Supplementary Figure 1). Four of these loci are known BMI loci (1q25, SEC16B; 4p12, GNPDA2; 16q12, FTO; and 18q21, MC4R). The fifth locus, at 5q33 (rs7708584, approximately 27 kb upstream of GALNT10, p=8.02×10^{−9}), has not been previously associated with BMI at genome-wide significant levels in any population.

We subsequently selected the 1,500 most significantly associated SNPs from Stage 1 (p<1.19×10^{−3}) and examined associations with BMI in an independent sample of 6,817 men and women of African ancestry from seven additional studies (Stage 2) (Online Methods, Supplementary Tables 1–3, Supplementary Note). Of these 1,500 SNPs, 179 replicated at nominal significance (p<0.05) and had effects directionally consistent with Stage 1 (Supplementary Table 4). A meta-analysis of Stages 1 and 2 revealed a second novel locus, 6q16 (rs974417, located in an intronic region of KHLHL32; Stage 2 p=3.5×10^{−3}; Stage 1+2 p=2.2×10^{−8}) and confirmed our finding at rs7708584 at 5q33 near GALNT10 (Stage 2 p=9.4×10^{−3}; Stage 1+2 p=2.2×10^{−10}). We further examined the associations of these two variants in a third stage composed of 25,451 individuals of African ancestry from an additional 12 studies. Support for an association was noted with both variants, although the strength of the association was greater for rs7708584 (GALNT10, p=7.1×10^{−3}) than for rs974417 (KHLHL32, p=0.09). In combining results across all three stages (n=71,412), rs7708584 (GALNT10) was significantly associated with BMI (p=3.4×10^{−11}) (Table 1, Figure 2a, b).

To identify additional novel loci that may be of importance across populations, we examined the 1,500 most significant SNPs from Stage 1 in publicly available data from the GIANT consortium of ~124,000 individuals European ancestry\(^7\) (Online methods). While rs7708584 (GALNT10) was significantly associated with BMI in European ancestry populations (effect allele frequency [EAF]=0.42; p=1.2×10^{−5}), rs974417 (KHLHL32) was not (EAF=0.85; p=0.45), although it was directionally consistent. Through a meta-analysis of European and African ancestry individuals, we identified an additional novel variant at 7p15 (rs10261878) that was also associated with BMI in European ancestry populations (GIANT: EAF=0.94, p=2.2×10^{−5}). SNP rs10261878 at 7p15 is located in an intergenic region 39 kb upstream of microRNA 148a (MIR148A) and approximately 241 kb upstream of the gene NFE2L3. This variant was positively associated with BMI in Stages 1 (p=1.7×10^{−4}) and 3 (p=1.0×10^{−3}) in the African ancestry GWAS, with a directionally consistent yet non-significant association noted in the smaller Stage 2 (p=0.33) (Figure 2c, Supplementary
Table 5). In combining results across studies of African (Stages 1, 2 and 3) and European ancestry (combined n=194,247), both SNPs rs7708584 (GALNT10, p=5.1×10^{-14}) and rs10261878 (MIR148a/NFE2L3, p=1.2×10^{-10}) were significantly associated with BMI; SNP rs974417 (KLHL32) failed to meet the genome-wide significance threshold (p=5.7×10^{-6}). In individuals of East Asian descent from the AGEN and RIKEN consortia (n=27,715 and 26,620, respectively) (Figure 3, Supplementary Table 6, Online Methods) rs7708584 (GALNT10, p=0.002) and rs974417 (KLHL32, p=0.023) were directionally consistent and significantly associated with BMI, while rs10261878 (MIR148A/NFE2L3) was neither directionally consistent nor statistically significantly associated with BMI (p = 0.053). Lastly, we examined the associations with BMI in children of African ancestry (n=3,751) (Online Methods) and for all three SNPs, the associations were directionally consistent, but did not reach statistical significance (p>0.05) (Supplementary Table 7).

To further understand differences by ancestral background as well as characterize the functional and genetic epidemiologic architecture of the two novel BMI loci (5q33, GALNT10, 7p15, MIR148A/NFE2L3) and the suggestive locus at 6q16 (KLHL32), we performed several additional analyses. Local ancestry adjustment (in 69% of the Stage 1 sample; Online Methods) resulted in numerically similar effect estimates (Supplementary Table 8) and we did not detect evidence of significant effect heterogeneity in analyses stratified by local ancestry (Supplementary Table 9). We found that the three BMI loci were associated with waist circumference (among n~20,000, of which many individuals overlap those studied here), but not with BMI-adjusted waist circumference, waist-to-hip ratio, or height (Supplementary Table 10), suggesting that the three loci are associated with overall body size, rather than with fat distribution. We found no evidence of pleiotropy with adiposity-related metabolic traits using GWAS data provided by trait-specific consortia in men and women predominantly of European ancestry (Supplementary Table 11).

We examined the associations with BMI in our African ancestry Stage 1 sample of the index SNPs reported for the 36 previously established BMI loci in the European and Asian populations (Figure 3, Supplementary Table 12). The associations were directionally consistent with the effects reported in the original papers for 32 of the 36 established BMI loci (p-value for binomial test of direction=9.7×10^{-7}) of which 16 variants associated with BMI at p<0.01 (p-value for binomial test <1.0×10^{-15}) (Supplementary Table 12).

Using the results from the Stage 1 meta-analysis, we searched for common variants within the established loci that better captured the association of the index SNP reported in the European and Asian populations. Seven regions (PTBP2, TMEM18, RBJ, NUDT3, BDNF, FTO, MC4R) harbored at least one variant that was correlated with the index SNP in the referent population (r^2 ≥0.4) and was associated with BMI in the African ancestry GWAS at a significance level that was at least one order of magnitude greater than that observed for the index SNP (Online Methods, Supplementary Table 13, Supplementary Figure 2a–g). These variants were also associated with BMI in GIANT (Supplementary Table 13) and are likely to be better markers of the biologically functional allele, at least in populations of African ancestry. We also interrogated the evidence for possible independent secondary signals by visual inspection of all p-values of SNP – BMI associations for SNPs with r^2 < 0.2 within the 1 Mb region of the index SNP. We did not detect evidence of independent secondary signals at any of the known BMI loci (at p<6.7×10^{-6}; see Online Methods). As illustrated in Supplementary Figure 3, for most loci, the genetic data from African ancestry populations may assist in refining the location of the risk variant as there are fewer markers correlated with the strongest signals and/or a more narrowed region in which proxies reside.

To direct us to positional candidate genes, we examined the cis-associations between the index SNP and expression of gene transcripts within the flanking 1Mb-region (500 kb each
side) in human brain, subcutaneous and omental adipose tissue, and liver\textsuperscript{24–27} (Online Methods, Supplementary Table 14). SNP rs7708584 near \textit{GALNT10} showed nominally significant (p<0.05) associations with \textit{GALNT10} expression (for two of the three transcripts available) in liver, omental, and subcutaneous fat (p=0.048, 0.00010, and 0.00017, respectively). Furthermore, we found suggestive \textit{cis}-associations for rs10261878 near \textit{NFE2L3} with \textit{NFE2L3} expression in the same three tissues (p=0.039, 0.015, and 0.036 for liver, omental, and subcutaneous fat, respectively). However, despite the consistent associations observed for our lead SNPs in the \textit{GALNT10} and \textit{NFE2L3} loci, other nearby SNPs showed stronger association with the expression levels for the respective transcripts (Supplementary Figure 4). Subsequent conditional analyses adjusting for the most significant eQTL SNP in the region abolished the \textit{cis}-associations between the BMI-associated SNPs and the respective transcript expression levels (Supplementary Table 15). Taken together, these eQTL analyses could not confirm that the identified BMI-SNPs affect \textit{GALNT10} and \textit{NFE2L3} expression directly.

We did not find non-synonymous SNPs in \textit{GALNT10}, \textit{NFE2L3} or \textit{KLHL32} that were correlated (r\textsuperscript{2}> 0.2) with the most significant SNPs in the 1000 Genomes Project African ancestry populations (AFR). However, we did detect a number of correlated SNPs (r\textsuperscript{2}>0.5) in regulatory sequences determined based on overlapping chromatin marks in multiple cell types, including brain and adipose tissue (Online Methods). Many of these SNPs (or good proxies in the 1000 Genomes Project AFR, r\textsuperscript{2} range 0.59–1.0), which are located in putative enhancer and promoter regions, had only marginally weaker associations in Stage 1 than the most significant SNPs reported in these regions (Supplementary Tables 16–19, Supplementary Figure 5a–c). Together these data suggest that the biologically relevant variants in all three regions may be regulatory in function.

The variant rs7708584 at chromosome 5q33 is located upstream of the gene galactosamine:polypeptide N-acetylgalactosaminyltransferase 10 (\textit{GALNT10}), which catalyzes the first step in the synthesis of mucin-type oligosaccharides (Supplementary Note). The protein is highly expressed in the small intestine and at intermediate levels in the stomach, pancreas, ovary, thyroid gland and spleen\textsuperscript{28}. Suggestive associations between BMI and \textit{GALNT10} have been observed in a smaller sample of African Americans\textsuperscript{14} that are included in the present Stage 1 meta-analysis, although the lead SNP differed (rs2033195) and displayed only moderate LD (r\textsuperscript{2} = 0.27) with the lead SNP discovered herein. The variant at 7p15, rs10261878, is intergenic and located 39 kb from a microRNA gene (\textit{MIR148A}), which has been found to be significantly up-regulated during adipogenesis\textsuperscript{29} as well as in human adipocytes\textsuperscript{30}. In addition, human miR-148a has been shown to regulate \textit{CCKBR} (cholecystokinin B receptor), which has been reported to play a regulatory role in the control of food intake\textsuperscript{31}. The next closest gene (241 kb from rs10261878) is the nuclear factor (erythroid-derived 2)-like 3 gene (\textit{NFE2L3}), a transcription factor that binds to antioxidant response elements of target genes and appears to play a role in differentiation, inflammation, and carcinogenesis\textsuperscript{32}.

The most significant SNP at chromosome 6q16 (rs974417) is intronic in the kelch-like 32 gene (\textit{KLHL32}). Kelch-like genes have propeller domains that bind substrate proteins, promoting substrate ubiquitination, which modulates protein function. We also detected evidence of recent positive selection in and downstream of \textit{KLHL32} (Supplementary Figures 6–9, Supplementary Note).

In the largest GWAS meta-analysis of African ancestry populations to date, we identified two novel loci and one highly suggestive locus influencing BMI. The most informative SNPs in each of these three loci explain 0.10% of the variance in BMI in African ancestry populations compared to 0.05% in Europeans and 0.03% in Asians (Table 1, Supplementary
Table 6). Using the most significant ancestry-specific markers from each locus, the 36 known BMI loci explain 1.30% of the variance in BMI in men and women of African ancestry compared with 1.67% and 1.25% in European and Asian ancestry populations, respectively (Supplementary Tables 12 and 13). We provide evidence for a shared genetic influence on BMI across populations, as directionally consistent associations were observed with the majority of known BMI risk variants. This observation suggests that the biologically functional alleles are ancient and likely arose before migrations out of Africa. In addition, we were able to refine the window of association of some of the previously established BMI loci, which may eventually help identify the biologically functional variant(s). In this study, we did not identify common variants for BMI that are likely to contribute to population differences in the prevalence of obesity. The ability to map novel loci and replicate signals at established loci found in other populations reflects differences in allele frequency and effect size, which are influenced by population differences in recent demographic history and linkage disequilibrium with the functional variant as well as genetic and environmental modifying factors. Further studies will be needed to test the biologically functional alleles at the known loci as well as the contribution of less common variation that has yet to adequately surveyed by genome-wide SNP arrays. Taken together, these findings demonstrate the importance of conducting genetic studies in diverse populations in order to identify novel susceptibility loci for common traits.

Online Methods

Study Design

We utilized a three-stage design consisting of a GWAS meta-analysis (Stage 1), a follow-up of 1,500 SNPs (Stage 2), and a focused follow-up of the three novel loci (Stage 3). Stage 1 included results from 36 GWAS of 39,144 men and women of African ancestry (37,956 African American and 1,188 African; Supplementary Table 1). We took forward the 1,500 most significantly associated SNPs (p-value <0.0003) for examination in 6,817 additional men and women of African ancestry from seven GWAS (Stage 2, all African American). The three SNPs that reached genome-wide significance (p<5×10^{-8}) after the meta-analysis of Stage 1 and Stage 2 results were taken forward for further confirmation in 25,451 additional African ancestry subjects from twelve studies. All participants in these studies provided written informed consent for the research, and approval for the study was obtained from the ethics review boards at all institutions. A description of each participating study as well as details regarding the measurement and collection of height and weight data are provided in the Supplementary Note.

Genotyping and Quality Control

Genotyping in each study was conducted using Illumina or Affymetrix genome-wide SNP arrays. The size of each study ranged from 50 to 8,421 individuals. The details of the array, genotyping quality control procedures, and sample exclusions for each study that contributed data are listed in Supplementary Table 1 and Supplementary Table 2.

Statistical Analysis

In all GWAS, imputation to phased haplotype data from the founders of the CEU and YRI HapMap Phase 2 samples (build 21) was performed using MACH \(^1\), IMPUTE2 \(^2\) or BEAGLE \(^3\). SNPs with lower imputation quality scores (r\(^2\)<0.3) (Supplementary Table 2) as well as SNPs with a small number of allele counts after stratifying by sex and case-control status were excluded from analyses. Local ancestry, defined as the number of European chromosomes (continuous between 0–2), was estimated for the majority of the Stage 1 African ancestry studies (Supplemental Table 8), using HAPMIX \(^4\). To evaluate the effect of
admixture on the allele distribution between African and European segments we stratified
the analysis of each variant by local ancestry at each locus (Supplementary Table 9).

**Stage 1**—GWA analyses were performed by each of the participating studies. BMI was
regressed on age, age\(^2\), and study site (if needed) to obtain residuals, separately by sex and
case-control status, if needed. Residuals were inverse-normally transformed to obtain a
standard normal distribution with a mean of 0 and a SD of 1. For studies with unrelated
subjects, each SNP was tested for additive association with BMI by regressing the
transformed residuals on the number of copies of the SNP effect allele, adjusting for
population structure as measured by the first ten eigenvectors calculated for each study.
Analyses were stratified by sex and case-control status (if needed). For studies that included
related individuals, family-based association tests were conducted that take into
consideration the genetic relationships among the individuals. Study-specific lambda values
ranged from 0.95 to 1.08 (Supplementary Table 2). We applied genomic control (GC) in the
Stage 1 analysis (i.e. divided by the median of all \(\chi^2\) statistics for each study) to eliminate
any remaining over dispersion before combining the GWAS in the meta-analysis. In Stage 1,
we conducted a fixed effect meta-analysis using the inverse variance weighted method
implemented in the program METAL\(^5\). We performed a second GC correction of the Stage 1
meta-analysis results (lambda = 1.136) before selecting SNPs for follow-up.

**Stages 2 and 3**—The 1,500 most significant SNPs from Stage 1 were examined in an
additional 6,817 individuals, with each SNP being analyzed as described for Stage 1 and
meta-analyzed using the inverse-variance method using METAL. As in Stage 1, each SNP
was tested for association with BMI by regressing the transformed residuals on the number
of copies of the SNP effect allele, adjusting for population structure as measured by the first
ten eigenvectors calculated for each study. Further testing of the 3 novel variants was
conducted in an additional 25,451 individuals (Stage 3). Results from all stages were meta-
analyzed using the inverse-variance method in METAL.

**Examination in individuals of European ancestry**—We also examined the 1,500
most statistically significant SNPs from Stage 1 in the GIANT consortium (n=123,706
individuals of European ancestry)\(^6\). Of these, 1,390 were genotyped or imputed in GIANT
and 1,328 had data for n>50,000 and a MAF>1%. We conducted a meta-analysis of Stages
1+2+3+GIANT in the same manner as described above. The three novel variants were also
examined in the AGEN and RIKEN consortia\(^7,8\) and the Pediatric Research Consortium
(PeRC) (see Supplementary Note).

**Estimation of Variance Explained**

The total fraction of variance explained was calculated using the formula \(2f(1-f)a^2\), where
f is the frequency of the variant and a is the additive effect of the variant\(^9\). When calculating
percent variance explained in the African ancestry sample, for the previously-discovered
BMI variants that were not genome-wide significant in Stage 1, we used data from the Stage
1 sample; for those that were genome-wide significant we used data from the Stage 2
sample; and for the novel BMI variants we used data from the Stage 2+3 samples to avoid
inflating the estimates due to the winner’s curse. When summing percent variance explained
for the 36 previously-discovered BMI variants (Supplemental Table 12), we utilized the
more informative SNP discovered through fine-mapping at the seven loci (listed in
Supplemental Table 13). However, for these seven variants Stage 1 results were used and
estimates may be biased; Stage 2 and 3 studies only participated in the look-up of the top
SNPs from preceding Stages.
Bioinformatic Analysis of the Novel BMI Loci

In an attempt to identify functionality in non-coding regions at the three loci, we utilized FunciSNP version 0.99\(^\text{10}\), which systematically integrates the 1,000 Genomes SNP data (1KGP, April 2012) with chromatin features of interest. In order to capture regulatory elements, we used 73 different chromatin features generated by next-generation sequencing technologies in brain and adipose tissues from the NIH Epigenomics Roadmap\(^\text{11}\) as well as known DNaseI hypersensitive locations, FAIRE-seq peaks, and CTCF binding sites from more than 100 different cell types, which were collected from the ENCODE data\(^\text{12}\).

All SNPs with an \(r^2\) value >0.5 with each index SNP in the 1KGP AFR populations in a 1Mb window around each index variant were catalogued. We used the UCSC Genome Browser (http://genome.ucsc.edu/) to illustrate the correlated SNPs which overlap chromatin features from these tissues as well as chromatin features from seven cell lines utilized in the ENCODE Project (Supplementary Figures 5a–c). All of the results from these analyses are provided in Supplementary Tables 16–19.

eQTL Analyses

Liver, subcutaneous, and omental fat tissue—The determination of eQTLs in liver, subcutaneous and omental fat tissue have been described in detail previously\(^\text{13}\). In brief, liver, subcutaneous, and omental fat tissue were obtained from patients of European ancestry who underwent bariatric surgery. Expression of a total of 39,280 oligonucleotide probes targeting transcripts representing 34,266 known and predicted genes was assessed. All patients were genotyped on a genome-wide SNP array and association between SNPs and gene expression data was adjusted for age, race, gender, and surgery year using linear regression. Results are presented in Supplementary Table 14 and Supplementary Figure 4.

Brain cortical tissue—We examined the cis-associations (defined as genes within 1 Mb) between each of the BMI SNPs and expression of nearby genes in brain (cortical tissue)\(^\text{14}\). The eQTL analyses have been described in detail previously (GEO database: GSE8919)\(^\text{14}\). In brief, DNA and RNA of neuropathologically normal cortical brain samples of 193 individuals (average age [range]: 81 [65–100] yrs) of European ancestry were isolated and genotyped for a genome-wide SNP array and HapMap genotypes were imputed. RNA expression was assessed for 24,357 transcripts of which 14,078 transcripts met the QC criteria. Association analyses between SNPs and expression data assumed an additive model and were adjusted for sex and age at death. Results are presented in Supplementary Table 14 and Supplementary Figure 4.

Association Testing of Previously Established BMI Loci

To characterize alleles that might better represent the biologically functional variant at the 36 previously-discovered BMI loci, we searched for LD proxies among individuals of African ancestry. Using HapMap data (CEU or JPT/CHB) to estimate LD, we identified all SNPs that were correlated (\(r^2 \geq 0.4\)) with the index SNP (within 250 kb, or larger to include a nearby gene). Next, we tested these SNPs for association with BMI in the Stage 1 African ancestry sample. We applied a locus-specific significance criterion \(\alpha\), which accounts for multiple testing [the number of tag SNPs in the HapMap YRI population that capture (\(r^2 \geq 0.8\)) all common SNPs (MAF \(\geq 0.05\)) correlated with the index signal in the HapMap CEU or JPT/CHB populations]. This alpha level does not account for the number of regions evaluated and reflects a balance between the need to correct for multiple comparisons and the prior knowledge that each region harbors a risk variant for BMI. We also looked for novel independent associations, focusing on the genotyped and imputed SNPs that were uncorrelated with the index signal in the initial GWAS populations (\(r^2 < 0.2\)). Here, we applied a Bonferroni correction for defining novel associations as significant in each region,
as \( 0.05 \) the total number of tags needed to capture \( (r^2 \geq 0.8) \) all common risk alleles across all risk regions in the YRI population (\( \alpha = 6.7 \times 10^{-6} \)).

**Detection of recent positive selection in Africans and Europeans at a novel BMI locus**

We evaluated the evidence for recent positive selection at our novel loci using several statistical techniques, the BioVU African American GWAS data, and data from the International HapMap Project and the Human Genome Diversity Project (HGDP). We compared adjusted allele frequencies among BioVU, and HapMap phase 3 participants from West African Yoruban (YRI) and East African Luhyia (LWK) using Treeselect\(^{15}\). The LWK sample is differentiated from the YRI and samples of African Americans\(^{16}\). Allele frequencies in the African American sample were adjusted by subtracting the expected contribution of European alleles, where \( p_{AA} \) is the allele frequency in African Americans obtained from experimental data, \( p_{EA} \) is the allele frequency in Europeans obtained from HapMap, \( p_{AF} \) is the estimated allele frequency in African founders, and \( \alpha \) is the average proportion of ancestry from Europeans, or 0.2. The adjustment is then performed by solving the following expression for \( p_{AF} \):

\[
p_{AF} = \frac{p_{AA} - \alpha p_{EA}}{(1 - \alpha)}
\]

We also evaluated the HapMap Phase II and HGDP data with the integrated haplotype score (iHS)\(^{17}\) and Haplogetter and the cross-population extended haplotype homozygosity (XP-EHH) statistic using the HGD selection browser\(^{18,19}\). We also evaluated BioVU using 5,000 random autosomal SNPs with STRUCTURE v2.3.3, and on average the participants were 20.7% European and 79.3% African ancestry\(^{20,21}\).

We observed evidence for recent selection near the \( KLHL32 \) gene within the YRI HapMap data using iHS (Supplementary Figure 4) and in the HGPD African participants (Supplementary Figures 5a–d). Nominal evidence of selection was observed within YRI and African American populations using the Treeselect statistic, with the transcription factor binding site SNP rs1206131 (\( p = 0.003 \) in the African Americans, and \( p = 0.005 \) in YRI and at the SNP rs9387284 (\( p = 0.004 \) in the YRI and \( p = 0.026 \) in the African Americans) (Supplementary Figure 6a, b). The Treeselect method also demonstrated a significant allele frequency differentiation between African and African-ancestry populations (Fst~0.01) at the transcription factor binding site SNP rs1206131. In panel (b), rs1206131 is the most significant SNP for this test in the region +/- 400kb. The test from the African American branch of the tree in (a) was slightly less significant at rs1206131 and the most significant SNP was downstream, which is also under the iHS and XP-EHH peaks from Africans in the HGDP and HapMap data. The graph of HGDP allele frequencies at this SNP shows that the ancestral T allele has increased frequencies throughout Africa relative to other major global populations (Supplementary Figure 7). Average (standard deviation, maximum) \( F_{ST} \) values in this region between YRI and African American were 0.001(0.001, 0.015), between YRI and CEU were 0.040 (0.045, 0.304), and between African American and CEU were 0.011(0.013, 0.082).

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.
Acknowledgments

A full listing of acknowledgments is detailed in the Supplementary Note.

References

Figure 1.
Manhattan plot displaying results of the BMI association meta-analysis in the Stage 1 studies. Colored genomic loci indicate novel associations (red) and those detected previously (blue).
Figure 2.
Regional plots of three novel genome-wide significant loci identified in men and women of African ancestry. (a) rs7708584 (GALNT10 region), (b) rs974417 (KLHL32 region), and (c) rs10261878 (MIR148A/NFE2L3 region). For 2a and b, Stage 1 p-value represented by purple circle and Stage 1+2+3 p-value represented by purple square; for 2c, Stage 1 p-value represented by purple circle, African ancestry Stage 1+2+3 p-value represented by purple diamond, and African ancestry + GIANT p-value represented by purple square. SNPs are plotted by their position 500kb on either side of the index SNP on the chromosome against their association ($-\log_{10} P$) with BMI using the Stage 1 data. SNPs surrounding the top SNPs are colored to indicate the local LD structure using pairwise $r^2$ data from the May 2012 AFR panel of the 1000 genomes.
Figure 3.
Effect estimates (95% CI) per BMI-increasing allele for the 3 novel loci discovered in individuals of African ancestry (1st section, in descending order of African effect size), the 32 loci discovered in individuals of European ancestry (2nd section, in descending order of European effect size), and the 4 loci discovered in individuals of Asian ancestry (3rd section, in descending order of Asian effect size). Results for individuals of African ancestry depicted by red dots (Stage 1+2+3 for novel loci, Stage 1 for previously-discovered loci); results for individuals of European ancestry depicted by black squares from Speliotes EK et al, 20107; and results for individuals of Asian ancestry depicted by green triangles from Okada Y et al, 20129 and Wen W et al, 201210.
Table 1

Summary of the eight independent SNPs that were associated with BMI at genome-wide significant ($p < 5.0 \times 10^{-8}$) levels in men and women of African ancestry

<table>
<thead>
<tr>
<th>Nearest gene</th>
<th>Previously identified BMI loci</th>
<th>Newly identified BMI loci</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rs543874</td>
<td>rs7586879</td>
</tr>
<tr>
<td>Chr</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Position (Build 37)</td>
<td>177889480</td>
<td>25116977</td>
</tr>
<tr>
<td>Alleles</td>
<td>G/A</td>
<td>T/C</td>
</tr>
<tr>
<td>EAF</td>
<td>0.25</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Stage 1

| N | 38899 | 38948 | 39097 | 39080 | 39103 | 38219 | 39120 | 39101 |
| β (SE) | 0.057 (0.009) | 0.042 (0.010) | 0.048 (0.009) | 0.074 (0.012) | 0.062 (0.010) | 0.050 (0.009) | 0.040 (0.008) | 0.030 (0.008) |
| p-value | 1.80E-10 | 1.05E-05 | 2.70E-08 | 2.27E-09 | 2.41E-10 | 8.02E-09 | 1.49E-06 | 1.66E-04 |

Stage 2

| N | 6805 | 6817 | 6817 | 6769 | 6817 | 6817 | 6816 | 6817 |
| β (SE) | 0.074 (0.020) | 0.073 (0.020) | 0.067 (0.021) | 0.068 (0.027) | 0.045 (0.021) | 0.047 (0.018) | 0.053 (0.018) | 0.017 (0.017) |
| p-value | 1.49E-04 | 3.12E-04 | 1.19E-03 | 0.012 | 0.032 | 9.35E-03 | 3.47E-03 | 0.330 |

Stage 3

| N | N/A | N/A | N/A | N/A | N/A | 25337 | 25451 | 25308 |
| β (SE) | 0.026 (0.010) | 0.015 (0.009) | 0.029 (0.009) | 0.078 (0.009) |
| p-value | 7.08E-03 | 0.091 | 1.01E-03 |

Combined

| N | 45704 | 45765 | 45914 | 45849 | 45920 | 70373 | 71387 | 194931 |
| β (SE) | 0.060 (0.008) | 0.047 (0.009) | 0.051 (0.008) | 0.073 (0.011) | 0.059 (0.009) | 0.040 (0.006) | 0.031 (0.006) | 0.032 (0.005) |
| p-value | 2.00E-13 | 3.60E-08 | 1.60E-10 | 1.05E-10 | 2.96E-11 | 3.37E-11 | 6.88E-08 | 1.23E-10 |

Explained variance (%) | 0.21% | 0.19% | 0.20% | 0.10% | 0.07% | 0.04% | 0.02% | 0.03% |

a Effect allele listed first;
Frequencies from Stage 1 sample;

Using results from Stage 2 for previously-identified BMI loci and results from Stage 2+Stage 3 for newly-identified BMI loci, the total fraction of variance explained was calculated using the formula \[2f(1−f)*a^2]\times 100, where \(f\) is the frequency of the variant and \(a\) is the additive effect of the variant (see Thorleifsson G et al, 2009);

Combined=African ancestry stages 1+2+3+GIANT [GIANT results are N=123706, \(\beta\) (SE)=0.045 (0.011), p-value=2.21E-05]. SNP, single nucleotide polymorphism; Chr, chromosome; EAF, effect allele frequency; \(\beta\) (beta estimate) reported in inverse-normally transformed units; SE, standard error. P-values for between-study heterogeneity all >0.1.