Trans-ethnic analysis of metabochip data identifies two new loci associated with BMI

J Gong\textsuperscript{1}, KK Nishimura\textsuperscript{1}, L Fernandez-Rhodes\textsuperscript{2}, J Haessler\textsuperscript{1}, S Bien\textsuperscript{1}, M Graff\textsuperscript{2}, U Lim\textsuperscript{3}, Y. Lu\textsuperscript{4,5,6}, M Gross\textsuperscript{7}, M Fornage\textsuperscript{6}, S Yoneyama\textsuperscript{2}, CR Isasi\textsuperscript{8}, P Buzkova\textsuperscript{9}, M Daviglus\textsuperscript{11}, D-Y Lin\textsuperscript{12}, R Tao\textsuperscript{12}, R Goodloe\textsuperscript{13}, WS Bush\textsuperscript{14}, E Farber-Eger\textsuperscript{13}, J Boston\textsuperscript{13}, HH Dilks\textsuperscript{15}, G Ehret\textsuperscript{16,17}, CC Gu\textsuperscript{18}, CE Lewis\textsuperscript{19}, K-DH Nguyen\textsuperscript{16}, R Cooper\textsuperscript{20}, M Leppert\textsuperscript{21}, MR Irvin\textsuperscript{22}, EP Bottiglieri\textsuperscript{16}, LR Wilkens\textsuperscript{23}, CA Haiman\textsuperscript{23}, L Park\textsuperscript{23}, KR Monroe\textsuperscript{23}, I Cheng\textsuperscript{24}, DO Strom\textsuperscript{23}, CS Carlson\textsuperscript{25}, R Jackson\textsuperscript{25}, L Kuller\textsuperscript{26}, D Houston\textsuperscript{27}, C Kooperberg\textsuperscript{1}, S Buyske\textsuperscript{28,29}, LA Hindorff\textsuperscript{10}, DC Crawford\textsuperscript{30}, RJF Loos\textsuperscript{4,5,31}, L Le Marchand\textsuperscript{32}, TC Matise\textsuperscript{32}, KE North\textsuperscript{30} and U Peters\textsuperscript{1}

OBJECTIVE: Body mass index (BMI) is commonly used to assess obesity, which is associated with numerous diseases and negative health outcomes. BMI has been shown to be a heritable, polygenic trait, with close to 100 loci previously identified and replicated in multiple populations. We aim to replicate known BMI loci and identify novel association in a trans-ethnic study population.

SUBJECTS: Using eligible participants from the Population Architecture using Genomics and Epidemiology consortium, we conducted a trans-ethnic meta-analysis of 102 514 African Americans, Hispanics, Asian/Native Hawaiian, Native Americans and European Americans. Participants were genotyped on over 200 000 SNPs on the Illumina Metabochip custom array, or imputed into the 1000 Genomes Project (Phase I). Linear regression of the natural log of BMI, adjusting for age, sex, study site (if applicable), and ancestry principal components, was conducted for each race/ethnicity within each study cohort. Race/ethnicity-specific, and combined meta-analyses used fixed-effects models.

RESULTS: We replicated 15 of 21 BMI loci included on the Metabochip, and identified two novel BMI loci at 1q41 (rs2820436) and 2q31.1 (rs10930502) at the Metabochip-wide significance threshold (P < 2.5 x 10^-7). Bioinformatic functional investigation of SNPs at these loci suggests a possible impact on pathways that regulate metabolism and adipose tissue.

CONCLUSION: Conducting studies in genetically diverse populations continues to be a valuable strategy for replicating known loci and uncovering novel BMI associations.

International Journal of Obesity (2018) 42, 384–390; doi:10.1038/ijo.2017.304

INTRODUCTION

Obesity is a heritable risk factor for a large number of serious health conditions.\textsuperscript{1–4} It already imposes an enormous burden on the public health system and will continue to impact the cost of medical care through the predicted rise in diseases linked to chronic obesity.\textsuperscript{5} In US ethnicities, obesity rates vary in African Americans (36.2%), Hispanics/Latinos (31.5%), Native Americans (41.2%), European Americans (27.9%) and Asians (9.9%).\textsuperscript{6} Body mass index (BMI) heritability studies estimate that up to 70% of BMI variability may be attributed to genetic factors.\textsuperscript{7–11} Although this might suggest that genetic traits contribute to racial/ethnic differences in rates of obesity, the relative importance of genetics compared with diet, behavior and socioeconomic factors is under continued investigation.\textsuperscript{12} However, it is indisputable that many minority groups have been disproportionately affected by the obesity epidemic and obesity research in minorities must remain a public health priority.

Genome-wide association studies (GWAS) in European ancestry populations have successfully identified numerous genetic variants associated with BMI, firmly establishing the importance of genetic factors on obesity.\textsuperscript{13–15} However, examining genetic associations in minority groups may reveal previously unidentified BMI loci and help to pinpoint causal variants. Conducting analyses...
in underrepresented minority populations has been shown to improve the statistical power to detect novel loci by increasing allele frequency and the variance of allele counts for some genetic variants.\textsuperscript{16–18} A recent fine-mapping study in African Americans benefited from the lower linkage disequilibrium (LD) patterns when identifying independent signals in known BMI loci, and also found two novel loci, presumably aided by the gain in power owing to the higher minor allele frequencies of these variants in those with African genetic ancestry.\textsuperscript{16} GWAS restricted to minority populations have had similar successes, uncovering additional BMI loci previously unidentified in studies of exclusively European ancestry.\textsuperscript{2–23} To date, the largest and most comprehensive BMI GWAS included individuals of both European and non-European descent, confirmed 41 known loci, and found 56 novel BMI-associated loci.\textsuperscript{24} The results from these studies highlights the feasibility and benefits of using diverse human populations as a strategy to broaden our knowledge of BMI genetics.

To identify additional BMI loci, we leveraged the multietnic design of the Population Architecture using Genomics and Epidemiology (PAGE) consortium to conduct a discovery meta-analysis in up to 102,514 individuals. Using this approach, we identified two novel BMI-associated loci, rs2820436 (1q41) and rs10930502 (2q31.1).

**MATERIALS AND METHODS**

**Study population**

The PAGE consortium is funded by the National Human Genome Research Institute to investigate the epidemiologic architecture of well-replicated genetic variants associated with human diseases or traits.\textsuperscript{25} The PAGE-I study, initiated in 2008, consists of a coordinating center and four consortia, each with access to large, diverse population-based studies. The four consortia are: Epidemiologic Architecture for Genes Linked to Environment (EAGLE), which is based on data from Vanderbilt University Medical Center’s biorepository linked to de-identified electronic health records (EAGLE-BioVU); the Multietnic Cohort Study (MEC); the Women’s Health Initiative (WHI); and Causal Variants Across the Life Course (CALICO), 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 (SOL) and the Strong Heart Study.\textsuperscript{25} The PAGE-II study, initiated in 2013, added the Charles Bronfman Institute for Personalized Medicine at Mount Sinai Medical Center, BioMe BioBank (MSSM). For specific analyses in this paper, PAGE reached out to additional studies, including GenNet and the Hypertension Genetic Epidemiology Network (HyperGen) to increase the African American sample size. The Supporting Information includes detailed descriptions of each study.

American African, Hispanic, Asian/Native Hawaiian, Native American and European participants from the ARIC, EAGLE-BioVU, CHS, CARDIA, MEC, MSSM, SOL, WHI, GenNet and HyperGen were eligible for inclusion in this study (Supplementary Table S1). Race/ethnicity was self-reported in most studies except for EAGLE-BioVU, where race/ethnicity was administratively-reported and recorded in the electronic health record.\textsuperscript{26,27} All studies were approved by Institutional Review Boards at their respective sites, and all study participants agree to a self-reported informed consent. The Vanderbilt University Internal Review Board has determined that data contained in EAGLE-BioVU are considered limited data sets as defined by the Health Insurance Portability and Accountability Act (HIPAA) and are in accordance with provisions of Title 45, Code of Federal Regulations, part 46 (45 CFR 46) that define criteria for ‘non-human subjects’ research.

The final sample of minorities from PAGE included 35,606 African American, 26,048 Hispanic/Latino, 22,466 Asian/Native Hawaiian, 17,859 European American and 535 Native American participants (Supplementary Table S1).

**Anthropometric measurements**

BMI was calculated by taking the ratio of the weight (kg) and height squared (m\(^2\)). For ARIC, CHS, CARDIA, HyperGEN, GenNet and WHI, BMI was calculated from height and weight measured at the time of study enrollment. In EAGLE-BioVU and MSSM, the median height and weight was calculated across all complete medical histories. MEC used self-reported height and weight. A validation study within MEC was conducted to assess the validity of these measures and showed that self-reported BMI was sometimes underestimated, but the difference was small (< 1 BMI unit) compared to the findings from national surveys.\textsuperscript{28} To reduce the influence of outliers on the analysis, individuals who were underweight BMI (< 18.5 kg m\(^{-2}\)) and extremely overweight (BMI > 70 kg m\(^{-2}\)) were excluded, and BMI values were natural log transformed to correct for the right-skewed distribution of BMI.

**Genotyping and imputation**

Genotyping was performed using the Metabochip, whose design has been described elsewhere.\textsuperscript{29} In brief, the Metabochip is a custom Illumina Select genotyping array of ~200,000 SNP markers and was designed to cost-effectively analyze putative association signals identified through GWAS meta-analyses of many obesity-related metabolic and cardiovascular traits. Imputation of Metabochip SNPs was conducted in MEC African Americans and Hispanics, MSSM African Americans and Hispanics, and WHI African Americans (SHARE) and Europeans. Study-specific reference samples\textsuperscript{30} or reference samples from 1000 Genomes Phase I\textsuperscript{31} were used. The programs MaCh and minimac were used for phasing and imputation, respectively.\textsuperscript{32–34} A summary of genotyping and imputation performance for each participating study has been published previously\textsuperscript{35} and reproduced in Supplementary Table S2.

Within each race/ethnicity, related participants were identified within and between studies using PLink.\textsuperscript{36} Identity by descent was estimated and when apparent first-degree relative pairs were identified, the member with the lower call rate was excluded from further analyses, with the exception of GenNet, SOL and HyperGen. These studies accounted for family structure using linear mixed models (GenNet, HyperGen) or with general-ized estimated equations which incorporate clusters of first-degree relative pairs/household members (SOL).\textsuperscript{37} In the remaining studies, participants with an inbreeding coefficient F > 0.15 were excluded. Ancestry principal components were generated using the Eigensoft software\textsuperscript{38,39} using either an unrelated subset, or in the 1000 Genomes reference populations, which were then projected into the study sample. Ancestral outliers were excluded from further analyses, as described previously.\textsuperscript{40} Additional information is included in the Supporting Information.

A total of 88,505 individuals were genotyped with the Metabochip, and an additional 14,009 with GWAS data were imputed into the 1000 Genomes Project\textsuperscript{41} or study-specific reference samples.\textsuperscript{30} For individuals with imputed data, only the Metabochip genetic variants were examined. Genotype data were cleaned by standard quality control procedures as described in the Supporting Information.

**Analysis**

As has been done in previous publications,\textsuperscript{6,41} BMI values were natural log transformed to account for the right-skewed distribution. Extreme BMI values (< 18.5 kg m\(^{-2}\) or > 70 kg m\(^{-2}\) were excluded from the analysis, with the assumption that these outliers could be attributable to data coding errors or an underlying rare condition outside the scope of this investigation. Given that CARDIA participants were generally younger, and young adults may have naturally low BMI measurements, the 18.5 kg m\(^{-2}\) exclusion criteria was waived for this cohort. These analyses were restricted to adults 20 years or older. The population was stratified by study and self-identified race/ethnicity, with each subgroup analyzed separately. Multivariable linear regressions for each study-specific minority group were adjusted for age, sex, study site (if applicable) and ancestry principal components (Supplementary Table S2). A sex × age interaction term was included in all models (except WHI, which only includes women) to account for possible effect modification by sex. The sex × age interaction term was intended to account for potential sex-specific effects on BMI that vary by sex and age, given that obesity risk and body composition are known to vary by age, and our study population includes both elderly participants and young adults older than 20 years of age. The results from each ethnicity, and for all ethnicities combined, were meta-analyzed using an inverse-variance weighted fixed-effects model in METAL.\textsuperscript{42} No inflation was observed in this meta-analysis (inflation factor \(\lambda = 0.97\)). SNPs with the smallest \(P\) values within a locus was considered the lead SNP. BMI associations were considered statistically significant if the \(P\)-value surpassed the Bonferroni corrected threshold of significance (\(P < 2.5 \times 10^{-7}\)), correcting for ~200,000 SNPs included on the Metabochip array. The locus considered novel if the lead SNP was not in LD
(\(r^2 < 0.1\) in any 1000 Genomes population) with a previously published known BMI loci. The list of known BMI loci was obtained by extracting records from the GWAS Catalog of the National Human Genome Research Institute (http://www.ebi.ac.uk/gwas/, accessed 26 April 2016) and through a literature search (April 2017) identifying publications based on high-throughput genotyping arrays that are not genome-wide (and thus, excluded from the GWAS Catalog).16 BMI studies examining GxE associations,43,44 and internal publications from collaborators that we expect to be published within the next year (Turcott V, in progress). Bioinformatic functional follow-up was performed for the most significant index SNP and all SNPs in high LD with the index SNP (\(r^2 > 0.8\) in African 1000 Genomes Population). HaploReg v4.2 and the UCSC Genome Browser from the Roadmap epigenomics project were used to assess whether variants in each of these loci were positioned in a putative enhancer or promoter specific to adipose tissue. GTex expression data were also used to assess whether any of the loci overlapped eQTL results.

RESULTS

The Metabochip array contains high density genetic variants at 21 previously published GWAS-identified BMI loci. We first assessed these known loci to evaluate the reproducibility of these loci in a multiethnic study population. Our study confirmed 15 of the 21 previously known BMI loci, significant at \(P < 5.8 \times 10^{-5}\), an approximate Bonferroni multiple testing correction for the average 866 SNPs at each BMI locus (Supplementary Table S3). Among the Metabochip previously known BMI loci that failed to replicate, the meta-analysis p-values approached significance, with most in the \(10^{-4}\) range.

When we examined the remaining Metabochip content, we found an additional 14 loci associated with BMI, which achieved a Metabochip-wide significance level of \(P < 2.5 \times 10^{-7}\), correcting for \(\approx 200,000\) SNPs on the Metabochip array (Supplementary Table S4). Eleven of these loci (or SNPs in high LD, \(r^2 > 0.8\), with these loci) were in LD (\(r^2 > 0.1\) in any 1000 Genomes population) with BMI loci previously identified since the development of the Metabochip.14,21–24,46–49 A twelfth SNP (rs11927381) no longer achieved Metabochip-wide significance after conditioning on a nearby SNP (rs1516725) that had previously been associated with BMI.44 Thus, we discovered two novel BMI-associated loci: 1q41 (rs2820436) and 2q31.1 (rs10930502) (Table 1). No evidence of heterogeneity was observed across studies at these two loci, with Cochran’s Q heterogeneity p-values of 0.65 and 0.94, for rs2820436 and rs10930502, respectively (Table 1, Figure 1).

The minor allele frequencies of these SNPs differed across the different ethnic groups (Table 1). rs2820436 was most frequent among PAGE African Americans (CAF = 0.48), and least frequent in Asians (CAF = 0.20), with the strongest association seen in the African Americans (\(P = 8.34E-04\) and Hispanic/Latinos (\(P = 1.61E-04\)). Although rs10930502 was also most frequent among African Americans (CAF = 0.70) and European Americans (CAF = 0.70), and least frequent among Asians (CAF = 0.33), the association was strongest among the Asians (\(P = 1.45E-03\)) and European Americans (\(P = 8.89E-03\)). Generally, the observed allele frequencies in our own study population were similar to those from the same ethnic groups in the 1000 Genomes populations. Both of these SNPs were analyzed in the most recent and largest BMI GWAS study to date (\(P_{(rs2820436)} = 1.02E-02\); \(P_{(rs10930502)} = 2.91E-04\),24 and were directionally consistent with our own results, providing additional support for these variants.

The variant rs10930502 was included on the Metabochip to follow-up on significant and suggestive signals from the largest available GWAS meta-analysis on BMI, whereas rs2820436 was included on the array for fine-mapping regions associated with waist-to-hip ratio (WHR). Given that rs2820436 was included on the Metabochip due to its previously published association with a non-BMI trait, we evaluated whether the associations with BMI were independent using individuals where WHR data were available (\(n = 53,481\)). When the association between rs2820436 and BMI was adjusted by WHR, the overall association did not noticeably change. Conversely, when the association between rs2820436 and WHR was examined, adjusting for BMI, this \(P\)-value also achieved Metabochip-wide Bonferroni significance (\(P = 3.09E-10\)). These findings suggest that this loci may influence multiple phenotypes related to body composition.

Functional investigation of the SNPs supports their likely involvement in lipid metabolism. We found that rs2820436 strongly tagged (\(r^2 = 0.94\) in 1000 Genomes Phase I Africans) a putative enhancer variant, rs2605096, positioned in an eQTL for the gene lysophospholipase-like 1 (LYPLAL1) previously associated with adiponectin,50 adiposity,51 cholesterol, T2D and WHR.52 Although rs10930502 was positioned in an eQTL for a lincRNA in adipose tissue, it did not strongly tag a putative regulatory variant. However, it was in moderate LD (\(r^2 = 0.48, D' = 0.85\) in 1000 Genomes Phase I Africans) with variant rs34636594 at 2q31.1, which was positioned in a transcription factor-binding enhancer in adipose tissue. lincRNAs are highly tissue specific and typically co-expressed with neighboring genes and thus we hypothesize that the 2q31.1 association may exert its effects on the candidate gene SLC25A12, through regulation of lincRNA.

DISCUSSION

This trans-ethnic meta-analysis replicated 15 of 21 previously known BMI loci included on the Metabochip. Of the six loci that did not reach statistical significance in our own study, two of these had lead SNPs that were very rare, with CAF < 0.01 in 1000 Genomes populations and PAGE racial/ethnic subgroups. As most

Table 1. Novel BMI-associated loci

| Metabochip Loci | Chr:BP | Gene | A1/A2 | Population | N | CAF | Beta | s.e. | P-value | HetP |
|----------------|--------|------|-------|------------|---|-----|------|------|---------|------|
| rs2820436 (1q41) | 1:219640680 | LYPLAL1/ZC3H11B | A/C | Combined | 94255 | 0.3876 | 0.0049 | 0.0009 | 3.79E-08 | 0.65 |
| AA | 35606 | 0.4782 | 0.0051 | 0.0015 | 8.34E-04 | NA |
| HA | 26046 | 0.4395 | 0.0062 | 0.0017 | 1.61E-04 | NA |
| AS | 14210 | 0.1952 | 0.0021 | 0.0021 | 3.20E-01 | NA |
| EA | 17859 | 0.3446 | 0.0024 | 0.0132 | 8.56E-01 | NA |
| NA | 534 | 0.3446 | 0.0024 | 0.0132 | 8.56E-01 | NA |
| rs10930502 (2q31.1) | 2:172890588 | METAP1D | A/G | Combined | 94256 | 0.6794 | 0.0004 | 0.0009 | 1.35E-07 | 0.94 |
| AA | 35999 | 0.7000 | 0.0004 | 0.0017 | 1.70E-02 | NA |
| HA | 26043 | 0.6555 | 0.0043 | 0.0018 | 1.40E-02 | NA |
| AS | 14220 | 0.3327 | 0.0056 | 0.0018 | 1.45E-03 | NA |
| EA | 17859 | 0.6971 | 0.0056 | 0.0021 | 8.89E-03 | NA |
| NA | 535 | 0.6794 | 0.0102 | 0.0137 | 4.56E-01 | NA |

Abbreviations: A1, coded allele; A2, non-coded allele; AA, African American; AS, Asian; BP, base pair position hg19/GRCh37; CAF, coded allele frequency; Chr, chromosome; EA, European American; HA, Hispanic American; HetP, heterogeneity P-value; NA, Native American; StdErr, Standard Error.
of these loci were originally discovered in GWAS studies with much larger sample sizes, 13–15,24,33–35 our smaller study was likely insufficiently powered to replicate the rarer variants (Supporting Information). Other loci that we failed to replicate had lead SNPs that were more frequent in Europeans than in non-Europeans. Given that only 17% of our study sample consisted of those with European ancestry, insufficient power may also have contributed to our inability to replicate some of these loci, especially if these were European-specific associations.

Interestingly, both of the novel loci we identified, rs2820436 (1q41) and rs10930502 (2q31.1), are common in those with European ancestry, with a frequency of 0.68 and 0.31 in 1000 Genomes Europeans, respectively. Previous large, European-based BMI GWAS studies may have failed to detect these associations due to population-specific GxG interactions, or GxE interactions linked to cultural, socioeconomic, or behavior risk factors, resulting in a more pronounced effect on BMI in minority groups compared with Europeans. For both novel SNPs reported here, the largest betas in our study occurred in a non-European subgroup, suggesting that the genotypes might have a greater effect on BMI among non-Europeans (Supplementary Table S5). Should non-European population-specific effects exist, our large sample of minority subjects may have yielded more power to detect those associations compared to previous GWAS studies that may have been underpowered to detect population-specific effects related to a certain minority group.

Another possible explanation for why these associations were not detected in previous GWAS efforts is that these SNPs may be a poor proxy for the underlying causal SNP in European populations, but are a better proxy for the causal SNP in non-European populations. LD patterns are known to differ by genetic ancestry. It is possible that these SNPs are in poor LD with the causal SNP in those with European ancestry, but in high LD with the causal SNPs in those with non-European ancestry. This would cause the association to be weaker or non-significant in Europeans owing to exposure misclassification, where the tag SNP is an inaccurate indicator for the presence of the causal SNP. Given that the Metabochip was designed to facilitate fine-mapping in non-Europeans, it is not surprising that some of the Metabochip tag SNPs may perform better at estimating causal genotype-phenotype associations in a predominantly non-European study population.

Our findings demonstrate the value of conducting GWAS in non-European populations, both when replicating findings previously discovered in large, often European-centric GWAS, and for discovering novel associations which may be population-specific, or have stronger effects in those with non-European ancestry. Finally, the functional findings provide additional evidence for the biological relevance of these new loci in the BMI phenotype, which warrant further investigation. Although these results are intriguing, additional replication is needed, especially using study populations that include underrepresented individuals. Both of these SNPs are most frequent in those with African ancestry, and our association in rs10930502 appears to be the strong in those with Asian ancestry.

Many genetic studies of BMI with larger sample sizes have been published and comparatively, we were underpowered to detect and replicate weaker associations, especially in less-frequent variants. It is possible that additional novel, or population-specific loci may be found in a larger, trans-ethnic study population. However, we assembled one of the largest and most diverse non-European study populations and were still able to confirm 15 of the 21 known BMI loci included on the Metabochip. Although the Metabochip was designed to replicate and fine-map
loci known to be associated with 23 disease-related traits, its content is not genome-wide and non-MetaboChip loci were not evaluated in this study. Yet, the inclusion of strong and well-established metabolically-related loci allowed us to identify a potential pleiotropic association with WHR. Studies that replicate our findings are advised to isolate the association that contributes specifically to BMI, given that our associations with BMI remain significant after adjusting for WHR. Through accompanying research efforts, we will benefit from the MetaboChip’s increased marker density to fine-map these associations and further describe the relationship between these loci, BMI, and related phenotypes.56–58

Certainly, there are challenges associated with multithetic genetic studies, but there are also legitimate benefits, which may help explain more of the BMI heritability. The dearth of studies that include underrepresented populations only sustains disparities in genetic research, inhibits our ability to identify population-specific genetic risk factors, and hinders the development and application of genetic findings in real-world clinical settings.56–58

Our findings are promising and perhaps more importantly, demonstrate the need to conduct additional genetic studies of complex traits in non-European individuals.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

KKH was supported by a National Cancer Institute training grant: Cancer Prevention Training in Nutrition, Exercise and Genetics (R25CA094880). LFR was supported by the Cardiovascular Disease Epidemiology Training Grant from the National Heart, Lung and Blood Institute (T32HL007555) and the American Heart Association (AHA) predoctoral grant (13PRE16100015). The Population Architecture Using Genomics and Epidemiology (PAGE-II) program is funded by the National Human Genome Research Institute (NHGRI), supported by U01HG004803 (CALiCo), U01HG004798 (EAGLE), U01HG004802 (MEC), U01HG004790 (WHI) and U01HG004801 (Coordinating Center), and their respective NHGRI ARRA supplements. The Population Architecture Using Genomics and Epidemiology (PAGE-II) program is funded by the National Human Genome Research Institute (NHGRI), with co-funding from the National Institute on Minority Health and Health Disparities (NHWD), supported by U01HG007416 (CALiCo), U01HG007417 (ISMMS), U01HG007395 (NHWD) and U01HG007419 (Coordinating Center). The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The complete list of PAGE members can be found at PAGE website (http://www.pagestudy.org). The data and materials included in this report result from a collaboration between the following studies: The ‘Epidemiologic Architecture for Genes Linked to Environment’ (EAGLE) is funded through the NHGRI PAGE program (U01HG007498-01) and its NHGRI ARRA supplement. The data set(s) used for the analyses described were obtained from Vanderbilt University Medical Center’s BioVU which is supported by institutional funding and by the Vanderbilt CTSA grant UL TR000445 from NCATS/NIH. The Vanderbilt University Center for Human Genetics Research, Computational Genomics Core provided computational and/or analytical support for this work. The Multicentric Cohort study (MEC) characterization of epidemiological architecture is funded through NHGRI (HGP00402, and HGP003797) and the NHGRI PAGE program (U01HG007357, U01HG004802 and its NHGRI ARRA supplement). The MEC study is funded through the National Cancer Institute (CA164973, R37CA54281, R01 CA 063464, 01CA133619, 01CA136792 and 01CA198758). The data sets used for the analyses described in this manuscript were obtained from dbGAP under accession phs000220. Funding support for the ‘Epidemiology of putative genetic variants: The Women’s Health Initiative’ study is provided through the NHGRI PAGE program (U01HG004790 and its NHGRI ARRA supplement). Funding support for the ‘Exonic variants and their relation to complex traits in minorities of the WHI’ study is provided through the NHGRI PAGE program (U01HG007376, U01HG004790). The WHI study is funded by the National Heart, Lung, and Blood Institute; NIH; and U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 3211-13, 32115, 32118-32119, 32122, 42107-26, 42129-32 and 42221. WHI PAGE-II is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HS057820100005C, HS0578201100006C, HS0578201100007C, HS0578201100008C, HS0578201100009C, HS0578201100010C, HS0578201100011C and HS0578201100012C. R01HL78671, R01HL59367 and R01HL068964; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN262200625226C. The authors thank the staff and participants of this important study for their infrastructural support and were partially supported by Grant Number UL1RR025051, a component of the National Institutes of Health and NIH Roadmap for Medical Research. The Coronary Artery Risk Development in Young Adults (CARDIA) study is supported by the following National Institutes of Health, National Heart, National Lung and Blood Institute contracts: N01-HC-95095; N01-HC-98047; N01-HC-98048; N01-HC-98049; N01-HC-45134; N01-HC-05187; and N01-HC-45205. CARDIA is conducted and supported by the National Heart, Lung, and Blood Institute in collaboration with the University of Alabama at Birmingham (HHSN268201300025C & HHSN268201300026C), Northwestern University (HHSN268201300027C), University of Minnesota (HHSN268201300028C), Kaiser Foundation Research Institute (HHSN268201300029C) and Johns Hopkins University School of Medicine (HHSN268200900041C). CARDIA is also partially supported by the Intramural Research Programs of the National Institute on Aging, the National Human Genome Research Institute and the Centers for Disease Control and Prevention, through cooperative agreements, with the WHI Biobank study was provided through the National Human Genome Research Institute (NHGRI) (U01HG007417). The data sets used for the analyses described in this manuscript were obtained from dbGAP under accession phs000223 (ARIC), phs000236, (CARDIA), phs000301 (CHS), phs000555 (HCHS/SOL). The opinions expressed in this paper are those of the author(s) and do not necessarily reflect the views of the Indian Health Service. GenNet is one of four networks in the Family Blood Pressure Program, established in 1995 and supported by a series of agreements with the NIH National Heart, Lung and Blood Institute. Samples and data from The Charles Bronfman Institute for Personalized Medicine at the Icahn School of Medicine at Mount Sinai (New York) were provided by The Charles Bronfman Institute for Personalized Medicine (IPM) BioMe Biobank used in this study were provided by The Charles Bronfman Institute for Personalized Medicine at the Icahn School of Medicine at Mount Sinai (New York). Phenotype data collection was supported by The Andrea and Charles Bronfman Philanthropies. Funding support for the Population Architecture Using Genomics and Epidemiology (PAGE) (IPM) BioMe Biobank data collection was provided by the National Human Genome Research Institute (U01HG007417). The data sets used for the analyses described in this manuscript were obtained from dbGAP under accession phs000925. The Hypertension Genetics Epidemiology Network (HyperGEN) study was supported by National Heart, Lung and Blood Institute contracts HL068694 and HL055673. Assistance with phenotype harmonization, SNP selection and annotation, data cleaning, data management, integration and dissemination and general study coordination was provided by the PAGE Coordinating Center (U01HG07419, U01HG04801-01 and its NHGRI ARRA supplement). The National Institutes of Mental Health also contributes to the support for the Coordinating Center. We gratefully acknowledge Dr Ben Voight for sharing the MetaboChip SNP LD and minor allele frequency statistics estimated in the Malmö
REFERENCES

1 Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nat Rev Cancer* 2004; 4: 579–591.

2 Kopelman PG. Obesity as a medical problem. *Nature* 2000; 404: 635–643.

3 Miller WM, Norian-Jaroe KS, Ulltstone M, Yanez J, McCullough PA. Obesity and lipids. *Curr Cardioil Rep* 2005; 7: 465–470.

4 Vuucenik I, Stains JP. Obesity and cancer risk: evidence, mechanisms, and recommendations. *Ann N Y Acad Sci* 2012; 1271: 37–43.

5 Cavley J, Meyerhoeffer C. The medical care costs of obesity: an instrumental approach. *J Health Econ* 2012; 31: 219–230.

6 Danane G, Ding EL, Mozaffarian D, Taylor B, Rehm J, Murray CJ et al. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PloS Med* 2009; 6: e1000058.

7 Olshansky SJ, Passaro DJ, Hershov RC, Layden J, James BA, Brody J et al. A potential decline in life expectancy in the United States in the 21st century. *N Engl J Med* 2005; 352: 1138–1145.

8 Blackwell DL, Lucas JW, Clarke TC. Summary health statistics for U.S. adults: national health interview survey 2012. *Vital Health Stat 10* 2014; 1:161.

9 Hjelmborg J, Fagnani C, Silventoinen K, McGue M, Korkelä M, Christensen K et al. Genetic influences on growth traits of BMI: a longitudinal study of adult twins. *Obesity (Silver Spring)* 2008; 16: 847–852.

10 Maes HH, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet* 1997; 27: 325–351.

11 Stunkard AJ, Focht TT, Hrubec Z. A twin study of human obesity. *JAMA* 1986; 256: 51–54.

12 McAllister EJ, Dhurandhar NV, Keith SW, Aronne LJ, Barger J, Baskin M et al. Ten putative contributors to the obesity epidemic. *Crit Rev Food Sci Nutr* 2009; 49: 868–913.

13 Berndt SI, Gustafsson S, Magri R, Ganna A, Wheeler E, Feitosa MF et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet* 2013; 45: 501–512.

14 Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010; 42: 937–948.

15 Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, Heid IM et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 2009; 41: 25–34.

16 Gong J, Schumacher F, Lim U, Hindorff LA, Hindorff J, Haessler J, Buyske S et al. An integrated map of genetic variation from 1,092 human genomes. *Nat Genet* 2010; 42: 352–358.

1744 Graff M, Scott RA, Justice AE, Young KL, Feitosa MF, Barata L et al. Genetic association analysis under complex survey sampling: the Hispanic Community Health Study/Study of Latinos. *Am J Hum Genet* 2014; 95: 675–688.

18 Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PloS Genet* 2006; 2: e190.

19 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; 38: 904–909.

20 Buyse S, Wu Y, Cant CL, Cheng I, Assimes TL, Dumitrescu L et al. Evaluation of the metabochip genotyping array in African Americans and implications for fine mapping of GWAS-identified loci: the PAGE study. *PloS One* 2012; 7: e35651.

21 Fesmimayer MD, North KE, Ritchie MD, Lim U, Franceschini N, Wilkins LR et al. 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)* 2013; 21: 835–846.

22 Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* 2010; 26: 2190–2191.

23 Justice AE, Winkler TW, Feitosa MF, Grall M, Fisher VA, Young K et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat Commun* 2017; 8: 14977.

24 Grall M, Scott RA, Justice AE, Young KL, Feitosa MF, Barata L et al. Genome-wide physical activity interactions in adiposity - A meta-analysis of 200,452 adults. *PloS Genet* 2017; 13: e1006528.

25 Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2010; 40: D930–D934.

26 Melen E, Granell R, O)objavainen M, Strachan D, Gonzalez JR, Wjst M et al. Genome-wide association study of body mass index in 23,000 individuals with and without asthma. *Clin Exp Allergy* 2013; 43: 463–474.

27 Stergioukoulis E, Gaillard R, Tavare JM, Balthasar N, Loos RJ, Taal HR et al. Genome-wide association study of height-adjusted BMI in childhood identifies functional variant in ADCYS3. *Obesity (Silver Spring)* 2014; 22: 2252–2259.

28 Warington NM, Howe LD, Paternoster L, Kaakinen M, Hernala S, Hukari V et al. A genome-wide association study of body mass index across early life and childhood. *Int J Epidemiol* 2015; 44: 700–712.

29 Winkler TW, Justice AE, Grall M, Barata L, Feitosa MF, Chu S et al. The influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PloS Genet* 2015; 11: e1003378.

30 Dastani Z, Hivert MF, Timpson N, Perry JR, Yuan X, Grall M et al. Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multiethnic meta-analysis of 45,891 individuals. *PloS Genet* 2012; 8: e1002607.

31 Lindgren CM, Heid IM, Randall JC, Lammia C, Steinhorsdottir V, Qi L et al. Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PloS Genet* 2009; 5: e1000508.
52 Heid IM, Jackson AU, Randall JC, Winkler TW, Qi L, Steinthorsdottir V et al. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat Genet 2010; 42: 949–960.

53 Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet 2009; 41: 18–24.

54 Graff M, Ngwa JS, Workalemahu T, Homuth G, Schipf S, Teumer A et al. Genome-wide analysis of BMI in adolescents and young adults reveals additional insight into the effects of genetic loci over the life course. Hum Mol Genet 2013; 22: 3597–3607.

55 Elks CE, Perry JR, Sulem P, Chasman DI, Franceschini N, He C et al. Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. Nat Genet 2010; 42: 1077–1085.

56 Bustamante CD, Burchard EG, De la Vega FM. Genomics for the world. Nature 2011; 475: 163–165.

57 Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. Nat Rev Genet 2010; 11: 356–366.

58 Haga SB. Impact of limited population diversity of genome-wide association studies. Genet Med 2010; 12: 81–84.

Supplementary Information accompanies this paper on International Journal of Obesity website (http://www.nature.com/ijo)