Human Molecular Genetics, 2014, Vol. 23, No. 3 **820–830**
doi:10.1093/hmg/ddt464
Advance Access published on September 23, 2013

Meta-analysis of genome-wide association data identifies novel susceptibility loci for obesity

Yu-Fang Pei[1,2], Lei Zhang[1,2], Yongjun Liu[2], Jian Li[2], Hui Shen[2], Yao-Zhong Liu[2], Qing Tian[2], Hao He[2], Shuyan Wu[1], Shu Ran[1], Yingying Han[1], Rong Hai[1], Yong Lin[1], Jingying Zhu[1], Xue-Zhen Zhu[1], Christopher J. Papasian[3] and Hong-Wen Deng[1,2,*]

[1]Center of System Biomedical Sciences, University of Shanghai for Science and Technology, Shanghai 200093, P R China [2]Center for Bioinformatics and Genomics, Department of Biostatistics and Bioinformatics, Tulane University, New Orleans, LA 70112, USA [3]Department of Basic Medical Science, University of Missouri-Kansas City, Kansas City, MO 64108, USA

Received March 13, 2013; Revised September 2, 2013; Accepted September 17, 2013

Obesity is a major public health problem with strong genetic determination. Multiple genetic variants have been implicated for obesity by conducting genome-wide association (GWA) studies, primarily focused on body mass index (BMI). Fat body mass (FBM) is phenotypically more homogeneous than BMI and is more appropriate for obesity research; however, relatively few studies have been conducted on FBM. Aiming to identify variants associated with obesity, we carried out meta-analyses of seven GWA studies for BMI-related traits including FBM, and followed these analyses by *de novo* replication. The discovery cohorts consisted of 21 969 individuals from diverse ethnic populations and a total of over 4 million genotyped or imputed SNPs. The *de novo* replication cohorts consisted of 6663 subjects from two independent samples. To complement individual SNP-based association analyses, we also carried out gene-based GWA analyses in which all variations within a gene were considered jointly. Individual SNP-based association analyses identified a novel locus 1q21 [*rs2230061*, *CTSS* (Cathepsin S)] that was associated with FBM after the adjustment of lean body mass (LBM) ($P = 3.57 \times 10^{-8}$) at the genome-wide significance level. Gene-based association analyses identified a novel gene *NLK* (nemo-like kinase) in 17q11 that was significantly associated with FBM adjusted by LBM. In addition, we confirmed three previously reported obesity susceptibility loci: 16q12 [*rs62033400*, $P = 1.97 \times 10^{-14}$, *FTO* (fat mass and obesity associated)], 18q22 [*rs6567160*, $P = 8.09 \times 10^{-19}$, *MC4R* (melanocortin 4 receptor)] and 2p25 [*rs939583*, $P = 1.07 \times 10^{-7}$, *TMEM18* (transmembrane protein 18)]. We also found that *rs6567160* may exert pleiotropic effects to both FBM and LBM. Our results provide additional insights into the molecular genetic basis of obesity and may provide future targets for effective prevention and therapeutic intervention.

INTRODUCTION

Obesity is a serious public health problem associated with an increased risk of type 2 diabetes, hypertension, cardiovascular disease and certain forms of cancer (1). In the USA, over two-thirds of the population are overweight and about one-third are obese (2). The annual economic cost attributable to obesity in the USA is as high as $100 billion (3).

Obesity is under strong genetic control (4–8). Body mass index (BMI) and fat body mass (FBM) are two important indices of obesity and both have strong genetic determination with heritability of 0.4–0.7 (4–8). BMI has been used widely for obesity research because it can be measured conveniently at low cost for large samples. However, BMI alone may not be accurate enough for assessing the extent of excessive fat accumulation and thus the health hazard of obesity because (1) body weight is mainly composed of FBM and lean body mass (LBM), which are under different development mechanisms; and (2) BMI may not always represent body fat appropriately since the relationship between BMI and FBM

*To whom correspondence should be addressed at: Center for Bioinformatics and Genomics, Department of Biostatistics and Bioinformatics, School of Public Health and Tropical Medicine, Tulane University, 1440 Canal St., Suite 2001, New Orleans, LA 70112, USA. Tel: +1 5049881310; Email: hdeng2@tulane.edu

is not always linear (9). FBM is considered phenotypically more homogeneous than BMI and may be advantageous for obesity research (4).

More than 10 genome-wide association (GWA) studies have been published for identifying genetic variants underlying obesity (10–21), primarily focused on BMI. These GWA studies have identified a number of obesity genes/loci, such as *FTO* (fat mass and obesity associated), *MC4R* (melanocortin 4 receptor), *NPC1*, *PTER*, *MAF*, *NRXN3*, *KCNMA1*, *PAX5*, *MRPS22, C12orf51* and *CTNNBL1*. Meta-analytic combination of multiple GWA studies is an efficient way to enhance the power of gene identification by greatly increasing the sample sizes. Several GWA meta-analyses identified additional susceptibility genes/loci underlying obesity (22–26), such as *TMEM18* (transmembrane protein 18), *KCTD15, GNPDA2, SH2B1, MTCH2, NEGR1*, *SDCCAG8*, *TNKS/MSRA*, RBJ/AD CY3/POMC, QPCTL/GIPR, SLC39A8, TMEM160, FANCL, CADM2, LRP1B, PTBP2, MTIF3/GTF3A, ZNF608, RPL2 7A/TUB, NUDT3/HMGA1,*OLFM4* and *HOXB5*. However, the genetic variants identified so far only explain a small fraction of the heritability attributable to obesity (24).

In the present study, aiming to identify additional genetic variants associated with human obesity, we conducted a GWA meta-analysis for obesity-related phenotypes in populations of diverse ethnicities. The discovery samples included seven GWA studies ($n = 21\,969$) and millions of genotyped/imputed SNPs from the 1000 Genomes Project (1000G) (27) reference panels. Significant SNPs were further *de novo* genotyped in two independent samples ($n = 6663$). Gene-based association analysis is a useful complement to individual SNP-based GWA analysis (28). In the presence of allelic heterogeneity, it is more powerful than traditional individual SNP-based analytical approaches (29). In addition to the meta-analytic association testing per SNP, we also evaluated the significance of all genes across the genome using the gene-based analyses.

RESULTS

The discovery meta-analysis included 21 969 individuals from the following seven GWA studies (see Supplementary Material, Methods and Materials): OOS ($n = 998$) (30), KCOS ($n = 2283$), COS ($n = 1624$), FHS ($n = 4760$) (31), IFS ($n = 1478$) (32), WHI-AA ($n = 7478$) and WHI-HIS ($n = 3348$). Basic characteristics of the samples are listed in Table 1. The majority of individuals were women ($n = 17\,955$). No population outlier was observed with principal components analysis (33). Imputation with the 1000G reference panels generated over 4 million SNPs that were qualified for further analyses (Supplementary Material, Table S1).

Meta-analysis of individual SNPs

Prior to meta-analysis, association statistics in each individual sample were corrected using genomic control (GC) approach (lambda = 0.98–1.08). A logarithmic quantile–quantile plot of the meta-analysis test statistics showed a marked deviation in the tail of the distribution, implying existence of true associations in these samples (Fig. 1).

The discovery GWA meta-analyses identified two genomic regions in 18q22 and 16q12 at the genome-wide significance (GWS) level (5×10^{-8}). In 18q22, the most prominent association was observed at *rs6567160* for BMI ($P = 7.23 \times 10^{-14}$, Table 2). Association at the same locus was also found for FBM ($P = 1.91 \times 10^{-7}$), of which the lead SNP was *rs17782313* ($P = 8.57 \times 10^{-9}$). These two SNPs are 22.0 kb apart and in strong linkage disequilibrium (LD) with each other ($r^2 = 0.93$), therefore probably representing a causal effect from the same source. The association between *rs6567160* and FBM(adj) ($P = 0.02$) and percentage of fat mass (PFM, $P = 1.59 \times 10^{-4}$) was weak. In 16q12, the GWS signal was found only for BMI, with the lead SNP being

Table 1. Basic characteristics of study samples

Gender	Stage	Sample	Ancestry	Sample size	Age (years)	Height (m)	Weight (kg)	BMI (kg/m^2)	FBM (kg)	PFM (%)
Female	Discovery	OOS	Caucasian	498 (487)	50.1 (17.7)	1.64 (0.06)	71.22 (15.86)	26.58 (5.96)	26.61 (10.43)	36.95 (7.08)
		KCOS	Caucasian	1727 (1727)	51.6 (12.9)	1.63 (0.06)	71.45 (16.04)	26.82 (5.98)	25.27 (10.78)	33.85 (7.73)
		COS	Chinese	823 (822)	37.5 (13.8)	1.58 (0.05)	54.63 (8.09)	21.78 (3.05)	16.13 (4.87)	29.30 (5.36)
		FHS	Caucasian	2606 (1787)	66.1 (13.2)	1.59 (0.07)	69.18 (15.67)	27.19 (5.74)	27.04 (8.52)	41.51 (7.29)
		IFS	Caucasian	1478 (1433)	32.7 (7.2)	1.65 (0.06)	71.65 (16.89)	26.21 (5.97)	25.33 (12.05)	36.41 (9.20)
		WHI-AA	African American	7478 (709)	60.9 (6.9)	1.62 (0.06)	83.15 (17.72)	31.00 (6.33)	37.53 (12.65)	45.26 (6.79)
		WHI-HIS	Hispanic	3348 (408)	60.7 (7.2)	1.57 (0.06)	73.87 (15.62)	28.80 (5.45)	32.58 (10.66)	44.73 (6.90)
	Replication	OOSR	Caucasian	2261 (2261)	49.1 (14.7)	1.63 (0.06)	72.05 (17.76)	26.96 (6.50)	26.64 (11.44)	35.41 (7.34)
		COSR	Chinese	1580 (1580)	53.4 (7.7)	1.58 (0.05)	53.36 (7.34)	21.25 (2.93)	15.35 (4.76)	28.60 (5.32)
		OOS	Caucasian	500 (488)	50.4 (18.9)	1.78 (0.07)	89.04 (14.93)	28.15 (4.47)	23.40 (8.88)	26.13 (6.59)
		KCOS	Caucasian	556 (556)	50.7 (16.0)	1.76 (0.07)	87.12 (16.74)	28.12 (4.85)	20.75 (9.35)	22.89 (6.88)
	Discovery	COS	Chinese	801 (800)	31.4 (11.9)	1.70 (0.06)	65.74 (9.64)	22.65 (2.93)	11.86 (5.14)	17.78 (5.75)
		FHS	Caucasian	2154 (999)	64.8 (12.4)	1.73 (0.08)	85.04 (15.44)	28.13 (4.48)	22.35 (6.20)	28.95 (5.93)
Male	Replication	IFS	Caucasian	0	–	–	–	–	–	–
		WHI-AA	African American	0	–	–	–	–	–	–
		WHI-HIS	Hispanic	0	–	–	–	–	–	–
		OOSR	Caucasian	1662 (1662)	43.5 (13.5)	1.76 (0.07)	84.09 (17.12)	27.05 (5.14)	19.78 (9.65)	22.63 (7.06)
		COSR	Chinese	1160 (1160)	63.8 (9.4)	1.70 (0.05)	63.80 (9.46)	22.07 (2.93)	11.00 (4.76)	16.93 (5.78)

Notes: Sample size is presented as number of subjects with available BMI (FBM).
Means (standard deviation) for age, height and weight were calculated based on subjects with BMI information.

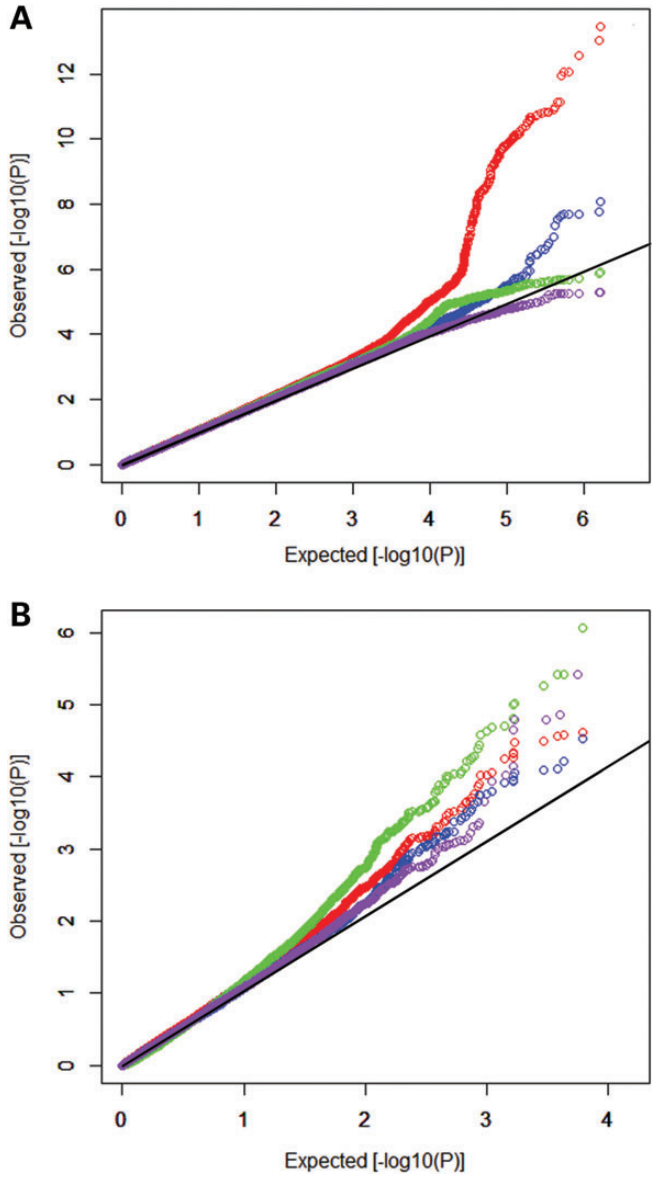

Figure 1. Logarithmic quantile–quantile (QQ) plot of the meta-analysis test statistics. (**A**) The QQ plot for SNP-based analyses for BMI (red dot), FBM (blue dot), FBM(adj) (green dot) and PFM (purple dot); and (**B**) the QQ plot for the gene-based analyses for BMI (red dot), FBM (blue dot), FBM(adj) (green dot) and PFM (purple dot).

rs62033400 ($P = 7.50 \times 10^{-12}$, Table 2). The association of *rs62033400* with FBM ($P = 8.49 \times 10^{-5}$), FBM(adj) ($P = 0.03$) and PFM ($P = 0.04$) was weak. Besides these two loci, three additional loci were identified at the level of suggestive significance (5×10^{-6}): 2p25, 1q41 and 1q21. The lead SNPs were *rs7569210* for BMI ($P = 4.07 \times 10^{-7}$) in 2p25, *rs320466* for FBM ($P = 3.50 \times 10^{-7}$) in 1q41 and *rs6693120* for FBM(adj) ($P = 1.26 \times 10^{-6}$) in 1q21. In summary, two genomic loci were identified at the GWS level and three additional loci were identified at the suggestive level, for multiple phenotypes.

Results from MANTRA were in concordance with the above findings. Specifically, MANTRA identified the same two loci 18q22 and 16q12 at the GWS level ($\log_{10}(BF) > 6$), with the

Table 2. Association results of the five *de novo* genotyped SNPs

Trait	SNP	Gene	Alleles	Discovery Effect (se)	*P*-value	I^2	Log_{10} (BF)	Replication Effect (SE)	*P*-value	I^2	Log_{10} (BF)	Combined Effect (SE)	*P*-value	I^2	Log_{10} (BF)
BMI	*rs6567160*	*MC4R*	T/C	0.09 (0.01)	$\mathbf{7.23 \times 10^{-14}}$	0	***12.11***	0.10 (0.02)	4.42×10^{-6}	0	***24.30***	0.09 (0.01)	$\mathbf{8.09 \times 10^{-19}}$	0	***34.69***
BMI	*rs62033400*	*FTO*	A/G	0.08 (0.01)	$\mathbf{7.50 \times 10^{-12}}$	0	***9.71***	0.07 (0.02)	6.00×10^{-4}	0	1.98	0.08 (0.01)	$\mathbf{1.97 \times 10^{-14}}$	0	***12.08***
BMI	*rs939583*	*TMEM18*	T/C	−0.07 (0.01)	1.03×10^{-6}	21	4.77	−0.06 (0.03)	0.03	0	–	−0.07 (0.01)	1.07×10^{-7}	8	5.64
FBM	*rs6567160*	*MC4R*	T/C	0.09 (0.02)	1.91×10^{-7}	11	5.56	0.09 (0.02)	3.88×10^{-5}	0	3.13	0.09 (0.01)	$\mathbf{3.24 \times 10^{-11}}$	0	***8.80***
FBM	*rs320466*	*RPS6KC1*	T/C	0.11 (0.02)	3.50×10^{-7}	0	5.02	0.01 (0.02)	0.78	72	0.18	0.05 (0.01)	4.92×10^{-4}	66	2.76
FBM(adj)	*rs2230061*	*CTSS*	G/A	−0.07 (0.02)	2.98×10^{-6}	0	4.34	−0.06 (0.02)	2.82×10^{-3}	0	1.51	−0.06 (0.01)	$\mathbf{3.57 \times 10^{-8}}$	0	***6.29***

Notes: FBM(adj) represents FBM adjusted for LBM. I^2 is the measure of heterogeneity (%).
P-values reached GWS level or $\log_{10}(BF) > 6$ are in bold italics.
The second allele is the effect allele.

same two lead SNPs (*rs6567160* $\log_{10}(BF) = 12.1$ and *rs62033400* $\log_{10}(BF) = 9.7$) for BMI. In 18q22, *rs17782313* was again the lead SNP ($\log_{10}(BF) = 6.7$) for FBM. No association was found for the other phenotypes. For the three suggestively associated loci, the MANTRA signals ranged from 4.34 to 5.56 (Table 2).

One SNP from each of these five loci was selected for *de novo* genotyping in two replication samples. SNPs were selected based on the strongest association signals, with two exceptions in 2p25 and 1q21. *rs7569210* ($P = 4.07 \times 10^{-7}$) and *rs6693120* ($P = 1.26 \times 10^{-6}$) were the most significant SNPs in these two regions. However, the genotyping probes for these two SNPs failed to be amplified in one of our two replication samples with *de novo* genotyping. They were replaced by *rs939583* ($P = 1.03 \times 10^{-6}$) and *rs2230061* ($P = 2.98 \times 10^{-6}$), which were successfully genotyped. In the joint analysis of the two replication samples, four SNPs were successfully replicated at the nominal level 0.05, while the last one *rs320466* was not significant ($P = 0.78$) (Table 2). Joint analyses of both discovery and replication samples retained *rs6567160* and *rs62033400* at the GWS level. In addition, the signal of *rs2230061* ($P = 3.57 \times 10^{-8}$) reached the GWS level. Association of *rs939583* got stronger as well ($P = 1.07 \times 10^{-7}$), although it did not reach the GWS level. At last, *rs320466* was filtered out because of a much weaker association signal ($P = 4.92 \times 10^{-4}$). Analyses of the replication samples with MANTRA gave similar results (Table 2). For all four SNPs that were successfully replicated, the results from the random-effects model were the same as those from the fixed-effects model (data not shown), and their MANTRA $\log_{10}(BF)$ values were larger than one. At *rs320466*, the MANTRA $\log_{10}(BF)$ was only 0.18.

Of the four loci described here, three (18q22, 16q12 and 2p25) had been previously reported to be associated with BMI-related phenotypes at the GWS level, while the last one (1q21) was novel. *rs6567160* in 18q22 was significant at the GWS level for both BMI ($P = 8.09 \times 10^{-19}$) and FBM ($P = 3.24 \times 10^{-11}$) in the combined analyses. Allele C at this imputed SNP increased both BMI and FBM (Fig. 1). It is located in an intergenic region between *MC4R* (209.4 kb) and *PMAIP1* (257.6 kb), of which *MC4R* was of primary interest. Genetic variation within *MC4R* was previously reported to be associated with both obesity and height (25,34,35). The association of *rs6567160* with FBM in the present study was likely to be a direct one rather than induced by height, because FBM was already adjusted by both height and height^2 in association analyses. *rs62033400* in 16q12 was strongly associated with BMI ($P = 1.97 \times 10^{-14}$). Allele G at this imputed SNP increased BMI (Fig. 1). It is located in an intron region of the *FTO* gene, which is a well-known gene associated with fat mass and obesity (36). Although the combined *P*-value 1.07×10^{-7} for the third SNP *rs939583* in 2p25 did not reach the GWS level, the successful replication ($P = 0.03$) indicated that this locus might be associated with obesity. Allele C at this imputed SNP decreased BMI. It is located in an intergenic region between *FAM150B* (334.2 kb) and *TMEM18* (45.4 kb). *TMEM18* was previously reported to be associated with BMI (26), thus strengthening the confidence on the association identified here. Finally, *rs2230061* in the only novel locus 1q21 was associated with FBM after the adjustment by LBM ($P = 3.57 \times 10^{-8}$). Allele A of this imputed SNP decreased residue of FBM (Fig. 1). *rs2230061* is located in an intron region of the gene *CTSS* (Cathepsin S), which was not reported previously. Associations between the four identified SNPs and all the studied phenotypes are listed in Supplementary Material, Table S2.

Despite being strongly associated with FBM ($P = 3.24 \times 10^{-11}$), *rs6567160* was weakly associated with FBM(adj) ($P = 2.32 \times 10^{-3}$). To explore possible explanation, we tested its association with LBM and with LBM after the adjustment of FBM, namely LBM(adj). The association with LBM was significant at the GWS level ($P = 9.35 \times 10^{-13}$). Interestingly, the association with LBM(adj) still existed, although the signal was weak ($P = 1.43 \times 10^{-4}$), implying that *rs6567160* may exert pleiotropic effects to both FBM and LBM (Table 3). Allele C at *rs6567160* increased both FBM and LBM (Fig. 2).

Gender-stratified *P*-values at the identified SNPs seemed to be different between the two gender groups. Specifically, *P*-values were always smaller in females than in males, especially for *rs939583* (Table 4). However, gender-specific effects were not significant with the SNP-by-sex interaction analyses. One potential reason for the smaller *P*-values in the females was the larger number of female subjects.

We estimated the cumulative effects of the identified SNPs on variation of the studied phenotypes in KCOS and COS samples. Individual SNPs explained 0.05–0.64% of phenotypic variations, and the variation explained by all four SNPs ranged between 0.46 and 1.61%, depending on samples and phenotypes (Table 5).

Replication with previously reported results

We first checked the replicability of our findings in the GIANT summary results of BMI (24). Of the four identified SNPs, *rs6567160* and *rs939583* existed in the GIANT data sets, and both were extremely significant ($P = 1.82 \times 10^{-22}$ and 5.34×10^{-22}).

Table 3. Pleiotropic effects of *rs6567160* on FBM and LBM

Trait	Discovery Beta (SE)	*P*-values	Replication Beta (SE)	*P*-values	Combined Beta (SE)	*P*-values
BMI	0.09 (0.01)	***7.23 × 10^{-14}***	0.10 (0.02)	4.42×10^{-6}	0.09 (0.01)	***8.09 × 10^{-19}***
FBM	0.09 (0.02)	1.91×10^{-7}	0.09 (0.02)	3.88×10^{-5}	0.09 (0.01)	***3.24 × 10^{-11}***
LBM	0.10 (0.02)	5.42×10^{-8}	0.10 (0.02)	3.67×10^{-6}	0.10 (0.01)	***9.35 × 10^{-13}***
FBM(adj)	0.04 (0.02)	0.02	0.04 (0.02)	0.06	0.04 (0.01)	2.32×10^{-3}
LBM(adj)	0.05 (0.02)	4.31×10^{-3}	0.05 (0.02)	0.01	0.05 (0.01)	1.43×10^{-4}

Notes: FBM(adj) represents FBM adjusted for LBM and LBM(adj) represents LBM adjusted for FBM.
P-values reached GWS level are in bold italics.

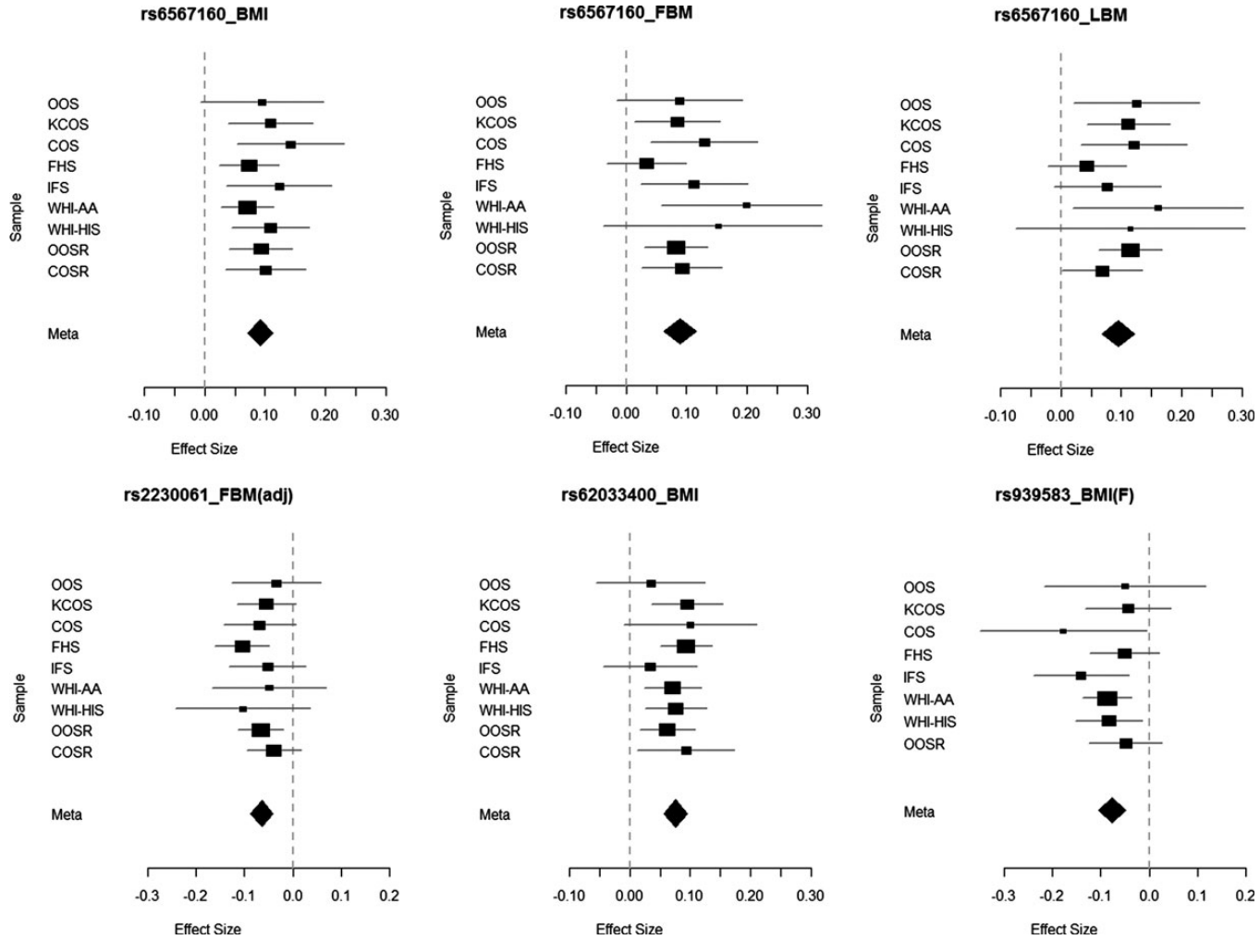

Figure 2. Forest plot for the identified SNPs associated with obesity-related traits. X-axis represents the effect size. The title of each sub-figure represents a specific combination of SNP and trait. BMI(F) refers to results for BMI in females and FBM(adj) refers to results for FBM adjusted for LBM.

Table 4. Sex-stratified results for the four identified SNPs

Trait	SNP	Gene	Alleles	Females Beta (SE)	*P*-value	Males Beta (SE)	*P*-value	*P*-value of SNP-by-sex interaction test
BMI	*rs6567160*	*MC4R*	T/C	0.09 (0.01)	$\boldsymbol{2.22 \times 10^{-12}}$	0.11 (0.02)	7.23×10^{-8}	0.42
BMI	*rs62033400*	*FTO*	A/G	0.07 (0.01)	$\boldsymbol{9.62 \times 10^{-10}}$	0.09 (0.02)	2.17×10^{-6}	0.87
BMI	*rs939583*	*TMEM18*	T/C	−0.08 (0.01)	$\boldsymbol{4.83 \times 10^{-8}}$	−0.03 (0.03)	0.26	0.51
FBM	*rs6567160*	*MC4R*	T/C	0.09 (0.02)	6.28×10^{-8}	0.09 (0.02)	7.79×10^{-5}	0.70
FBM(adj)	*rs2230061*	*CTSS*	G/A	−0.07 (0.01)	6.06×10^{-7}	−0.05 (0.02)	0.01	0.31

Notes: FBM(adj) represents FBM adjusted for LBM.
The second allele is the effect allele.
P-values reached GWS level are in bold italics.

The other two SNPs *rs62033400* and *rs2230061* did not exist in the GIANT data set. Scrutiny of their nearby SNPs implied that they were well replicated. For example, the SNP *rs3751812* near *rs62033400* (LD, $r^2 = 0.99$) was extremely significant ($P = 1.14 \times 10^{-59}$) in the GIANT results. Another SNP *rs4509581* near *rs2230061* ($r^2 = 0.87$) was significant as well at the nominal level 0.05 ($P = 0.03$).

We then checked the replicability of previously reported loci in our summary results. By searching in the GWAS catalog (37), we identified 32 distinct genomic regions and 39 SNPs that were reported previously at the GWS level. The signals of these SNPs in our results are listed in Supplementary Material, Table S3. Of the 39 SNPs, 3 were not available in our results, and 22 were significant at the nominal level 0.05.

Gene-based meta-analysis

A total of 17 483 genes were included in the gene-based association test. The Bonferroni correction was used to declare the significance level ($0.05/17\,483 = 2.86 \times 10^{-6}$). In this analysis, we identified one gene *NLK* (nemo-like kinase; 17q11) that was significant for FBM(adj) ($P = 8.77 \times 10^{-7}$, Table 6). This gene had not been reported previously. Exploration of allelic structures showed that common variants (MAF > 5%) accounted for the majority of genetic variation in *NLK*. Variants were common in all but the WHI-AA samples. In the WHI-AA sample, 25 of the total 77 SNPs were less common with MAF between 1 and 5%, but no rare variants (MAF < 1%) were involved. All of the 77 SNPs were well imputed ($r^2 > 0.3$) and the average r^2 was as high as 0.78. The gene remained significant ($P = 6.82 \times 10^{-5}$) after excluding these less common variants from the combined analyses, although the *P*-value was higher than the significance level. No gene was identified at the GWS level for the other phenotypes. Although the four candidate genes (*CTSS*, *MC4R*, *FTO* and *TMEM18*) identified in the single variants analyses did not reach the significance level in the gene-based analyses, their signals were all nominally significant ($P < 0.05$, Supplementary Material, Table S4).

DISCUSSION

In the present study, we carried out a GWA meta-analysis of seven GWA studies of diverse populations for obesity, followed by *de novo* replication. Using both individual SNP- and gene-based analyses, we identified two novel loci 1q21 and 17q11 associated with FBM after the adjustment of LBM. We also confirmed three previously reported loci (16q12, 18q22 and 2p25) at the GWS level.

The two candidate genes implied by the two identified loci were *CTSS* and *NLK*. The *CTSS* gene belongs to a family of cystein protease that includes other proteases (e.g. *CTSK* and *CTSL*) involved in the development of obesity (38). It plays a central role in extracellular matrix remodeling by stimulating adipocyte differentiation through degrading fibronectin, one of the main components of the extra cellular matrix (39). *In vitro* studies showed that *CTSS* expression and Cathepsin S secretion in adipose tissue were induced by *LPS*, *TNF-α* and *IL-1β*, proinflammatory factors that are secreted by cells such as macrophages or smooth muscle cells (40). The involvement of *CTSS* in the development of obesity was also supported by extensive gene expression studies and clinical studies (40–42). The *NLK* gene is a negative regulator of the canonical Wnt/beta-catenin signaling pathway (43), which was demonstrated to play an important role in maintaining pre-adipocytes in an undifferentiated state by inhibiting adipogenic gene expression (44). Our finding, therefore, provides additional evidence on the novel hypothetical mechanism by which *NLK* contributes to the development of obesity.

The three candidate genes that we confirmed in the present study were *FTO*, *TMEM18* and *MC4R*. They were all well established to be associated with BMI by previous studies (25,34). We found that *rs6567160,* near *MC4R,* had a pleiotropic effect to both FBM and LBM, implying that *MC4R* may regulate the development of both fat and lean masses. This gene plays an important role in the regulation of food intake, energy expenditure and catabolism (45,46). Humans with rare functional mutations in their *MC4R* sequence are known to develop severe early-onset obesity (47) and analogous phenotypes are seen in murine models of *MC4R* disruption (34). In mice, stimulation of *MC4R* has been demonstrated to decrease food-seeking behavior, to increase basal metabolic rate and to decrease LBM (45). Blockade of central melanocortin signaling increased both LBM and FBM in a rat model (48), supporting the bidirectional role played by the *MC4R* in obesity development.

PFM is one of the commonly used traits for body fat mass (31). However, no significant findings were obtained with PFM in the present study at either the GWS or suggestive significance level. Instead, both novel findings were identified with original fat/lean

Table 5. Phenotypic variation explained by the identified SNPs

Trait	SNP	Variance explained (%)	
		COS	KCOS
BMI	*rs6567160*	0.64	0.43
BMI	*rs62033400*	0.21	0.45
BMI(F)	*rs939583*	0.52	0.05
FBM	*rs6567160*	0.54	0.26
FBM(adj)	*rs2230061*	0.22	0.15
PFM	*rs6567160*	0.33	0.13
BMI	Total	1.11	0.90
BMI(F)	Total	1.61	0.73
FBM	Total	1.34	0.68
FBM(adj)	Total	1.10	0.49
PFM	Total	0.97	0.46

Notes: FBM(adj) represents FBM adjusted for LBM and BMI(F) refers to results for BMI in females.

Table 6. Genes identified by gene-based association test ($P < 1 \times 10^{-5}$)

Gene information				*P*-values in different samples				
Name	Chr	Start	Stop	Caucasian	COS	WHI-AA	WHI-HIS	Combined
NLK	17	23 393 814	23 547 531	0.04	9.92×10^{-3}	1.70×10^{-5}	0.07	***8.77 × 10⁻⁷***
RNF213	17	75 928 320	75 984 673	0.09	0.06	2.01×10^{-4}	2.43×10^{-3}	3.75×10^{-6}
DHX40	17	54 997 667	55 040 484	0.12	0.65	3.60×10^{-5}	9.55×10^{-4}	3.87×10^{-6}
TBX4	17	56 888 588	56 916 446	0.18	0.77	6.10×10^{-5}	4.66×10^{-4}	5.55×10^{-6}
PMP22	17	15 073 820	15 109 369	0.92	0.05	7.10×10^{-5}	2.08×10^{-3}	9.50×10^{-6}
C7orf57	7	48 041 641	48 067 419	0.45	2.46×10^{-3}	8.33×10^{-4}	8.28×10^{-3}	9.87×10^{-6}

Notes: The results for Caucasian were calculated based on *P*-values from GWA studies of individuals with European ancestry (fixed-effect meta-analysis of OOS, KCOS, FHS and IFS GWA studies). *P*-values that reached genome-wide gene-based significance levels are in bold italics. All these genes are identified for FBM(adj).

mass measurements. This result implies that modeling on original measured parameters may be advantageous compared with modeling on a synthesized trait (e.g. PFM). Although PFM is the surrogate of fat mass, it has a linear relationship with percentage of lean mass (PLM, PFM+PLM = 1) so that PFM and PLM are equivalent from an analytical point of view. In other words, results of PFM have the same interpretations to both fat and lean mass. Because fat and lean have different biological and developmental mechanisms and also because PLM and PFM have a linear relationship, interpretation of the results on PFM may be somehow difficult.

The only SNP that was not significant in the *de novo* replication results was *rs320466* for FBM. In addition to a false-positive signal, the reasons for the failure of replication may be the following: first, genetic heterogeneity caused by different evolution histories may result in differential genotype–phenotype associations (49). Second, interacting variants in the background may modulate individual marker effects, diminishing or increasing marginal associations of the studied polymorphisms (50). Third, significant associations are usually declared at genetic markers that are in LD with the causal site, rather than the causal site itself. Therefore, the failure to replicate could be a result of different patterns of LD between populations.

Gene-based analysis has the capacity to identify genes that are missed by SNP-based analyses. This is because gene-based analysis is more powerful in certain conditions by jointly modeling onto multiple SNPs within a gene. In our study, none of the SNPs located within *NLK* was significant. However, when modeling all SNPs together, the gene was significant. One of the advantages of the gene-based analysis is the ability to take into account allelic heterogeneity (i.e. presence of more than one susceptibility allele in a gene), which is low-powered by individual SNP-based analysis. We did not validate the multiple imputed SNPs within the *NLK* through experiments; however, we believe our results were reliable because of the following two reasons: (i) all included variants were common (MAF $>$ 5%) or less common (MAF $>$ 1%); and (ii) all SNPs were well imputed ($r^2 > 0.3$), in practice the average r^2 was 0.78.

The samples that we analyzed were from populations of diverse ethnicities. Although using different samples in one study might increase the false-negative rates due to sample heterogeneity, the loci identified here are not likely to be due solely to heterogeneity. First, neither *Cochran*'s Q statistic nor I^2 measures demonstrated evidence of heterogeneity at any of the identified SNPs. Second, replication in distinct samples makes the false-positive findings less likely. Third, the random-effects model, which is more robust against heterogeneity, generated similar results to the fixed-effects model. Fourth, the results from MANTRA, which combines GWAS from different ethnic groups by taking advantage of the expected similarity in allelic effects between the most closely related populations, gave highly concordant results. Finally, multiple previously reported loci (e.g. *FTO, TMEM18* and *MC4R*) were replicated using the same analytical approach. Therefore, the identified loci may represent true genetic susceptible loci for obesity that are shared across populations.

In summary, by meta-analyzing multiple GWA studies and following by *de novo* replication, we have identified two novel loci for obesity, as well as confirmed three previously reported loci. Further studies are warranted to explore the biological functions of the genetic variants identified in the present study.

MATERIALS AND METHODS

Study populations

This study included multiple samples from different research and/or clinical centers. All samples were approved for use by respective institutional ethics review boards, and all participants provided written informed consent.

In the discovery stage, seven GWA samples were included, of which three were from in-house studies and four were from the public database of genotype and phenotype (dbGAP) (51). The three in-house samples included those from the Omaha Osteoporosis Study (OOS, with 998 unrelated Caucasian subjects), Kansas-City Osteoporosis Study (KCOS, with 2283 unrelated Caucasian subjects) and China Osteoporosis Study (COS, with 1624 unrelated Chinese Han subjects). The fourth sample was selected from the Framingham Heart Study (FHS), a longitudinal and prospective cohort comprising over 16 000 Caucasian subjects spanning three generations (52,53). Based on the first two generations of the FHS families, we identified 4760 subjects with BMI information and 2786 subjects with FBM information for use in this study. The fifth sample was selected from the Indiana Fragility Study (IFS), a quantitative and cross-sectional cohort comprising 1493 premenopausal Caucasian sister pairs (54). Among them, 1478 subjects with BMI information and 1433 subjects with FBM information were included in the analyses. The sixth and seventh samples were selected from the Women's Health Initiative (WHI) observational study, a partial factorial randomized and longitudinal cohort with over 12 000 genotyped women of African-American or Hispanic ancestry (55). The sixth sample included 7478 individuals of African-American ancestry (WHI-AA) with BMI information and 709 individuals with FBM information. The seventh sample included 3348 individuals of Hispanic ancestry (WHI-HIS) with BMI information and 408 individuals with FBM information.

For replication, the selected SNPs were *de novo* genotyped in two independent samples, one with 3923 unrelated Caucasian individuals selected from OOS and KCOS (OOSR), and the other with 2740 unrelated Chinese Han individuals selected from COS (COSR).

Phenotype measurements and modeling

BMI was calculated as body mass (in kilograms) divided by the square of height (in meters). FBM and LBM were measured/calculated with dual-energy X-ray absorptiometry scanners (Lunar Corp., Madison, WI, USA; or Hologic Inc., Bedford, MA, USA) following the manufacturer protocols. PFM was calculated as FBM divided by the sum of FBM and LBM. In addition to BMI, FBM and PFM, we analyzed FBM after the adjustment of LBM, denoted by FBM(adj). The focus of this latter phenotype was on the fat part of body weight only. For all phenotypes, covariates were screened among gender, age, age^2, height (except for BMI), height^2 (except for BMI) and scanner ID (in WHI, except for BMI) with the step-wise linear regression model. In the admixed WHI-AA and WHI-HIS samples, the first five principle components derived from genome-wide genotype data were

included as covariates to correct for potential population stratification/admixture (33). Phenotypes were adjusted by significant covariates. Adjusted phenotypes were then normalized by inverse quantile of the standard normal distribution and the normalized phenotypes were used in subsequent association analyses.

Genotyping and quality control

Individual GWA samples were genotyped by high-throughput SNP genotyping arrays (Affymetrix Inc., Santa Clara, CA, USA; or Illumina Inc., San Diego, CA, USA) following the manufacturer protocols. Quality controls of genotype data were implemented using PLINK (56) with the following criteria applied: individual missingness 5%, SNP call rate 95% and Hardy–Weinberg equilibrium (HWE) P-value 1.0×10^{-5}. For familial samples (FHS and IFS), all genotypes with the Mendel error were set to missing. Population outliers were monitored by principal components derived from genome-wide genotypes (33).

Genotype imputation

In the discovery stage, imputation was performed for each GWA study using 1000G sequence variants as a reference panel (as of August 2010). Reference samples included 283 individuals of European ancestry, 193 individuals of Asian ancestry and 174 individuals of African ancestry. Imputation was performed by comparing the respective panel with the closest ancestry.

Prior to imputation, a consistency test of allele frequency between GWA samples and reference samples was examined with the Chi-square test. To correct for potential misstrandedness, SNPs that failed consistency test ($P < 1.0 \times 10^{-6}$) were transformed into reverse strand. SNPs that failed consistency again were removed from the GWA studies.

To distribute imputation computation to multiple parallel CPUs, chromosomes were split into non-overlapping fragments each of 10 Mega base-pair length. In each fragment, haplotypes of the individual GWA study were phased by a MArkov Chain Haplotyping algorithm (MACH) (57,58). For familiar samples (FHS and IFS), 200 unrelated founder individuals were randomly selected to estimate model parameters, which were then used to impute all family members. Based on phased haplotypes, untyped genotypes were then imputed by a computationally efficient imputing algorithm Minimac (59). SNPs with r^2 score less than 0.3 as estimated by Minimac were considered of low imputation accuracy. SNPs of low accuracy or of MAF < 0.01 in any one of the samples were excluded from subsequent SNP-based association analyses.

Association testing

Associations between phenotypes and SNPs (both genotyped and imputed) were tested under an additive mode of inheritance. For samples of unrelated subjects, association was examined by the linear regression model with MACH2QTL (57,58), in which allele dosage was taken as the predictor for the phenotype. Empirical assessments show that MACH2QTL provides an effective means of evaluating evidence for association at untyped markers (60). For familial samples (FHS and IFS), a mixed linear model was used in which the effect of genetic relatedness within each pedigree was also taken into account (61,62). To correct for potential population stratification, GC inflation factor (63) was estimated for each GWA study and study-specific summary statistics were corrected prior to the meta-analysis.

Meta-analysis

Summary association statistics from each GWA study were combined to perform weighted fixed-effects meta-analyses with METAL (64), in which weights were proportional to standard error of regression coefficient. *Cochran*'s Q statistic and I^2 were estimated with METAL to quantify between-study heterogeneity effect. Significant heterogeneity effect was defined as $I^2 > 50\%$ or Q-statistic P-value < 0.1. We only focused on SNPs with the association directions consistent in all the seven GWA studies. The random-effects model of meta-analysis was applied to validate significant SNPs, with the R package '*rmeta*' (65).

During the preparation of the present study, Morris developed a new meta-analysis method MANTRA that is specific to samples of diverse ethnicities (66). One advantage of this method is the ability to model the expected similarity in allelic effects between closely related ethnic groups. As an important complement and validation, we also analyzed data with MANTRA. Genome-wide significance level for MANTRA was set at the base 10 logarithm of Bayes factor $\log_{10}(\mathrm{BF}) > 6$.

De novo replication analysis

SNPs with P-values $< 5.0 \times 10^{-6}$ in the discovery samples were considered as significant. One SNP from each distinct genomic region was subjected to *de novo* replication in two independent samples, namely OOSR and COSR. The OOSR sample was genotyped using a fluorescent resonance energy transfer-based KASP genotyping platform with technical support from Kbiosciences UK Company. The COSR sample was genotyped using the ligation detection reaction method with technical support from the Shanghai Biowing Applied Biotechnology Company. Quality control included individual missingness 5%, SNP call rate 90% and HWE P-value 1.0×10^{-5}. Association testes in the *de novo* samples were performed with PLINK (56). Joint analyses of both discovery and replication samples were performed with METAL (64) and MANTRA (66).

SNP-by-sex interaction analysis

To examine gender-specific effects, we performed gene-by-sex interaction analyses for the identified SNPs. In each individual study, interaction was tested by a linear regression model with PLINK (56), in which gender was encoded into a binary variable. P-values from individual studies were then meta-analyzed for a combined signal with Fisher's method (67).

Replication analysis with previously reported results

We checked the replicability of our findings in another larger publicly available meta-analysis of BMI conducted by the GIANT consortium (24), in which 249 796 subjects were involved. We also checked the replicability of previously reported loci in our discovery samples.

Gene-based meta-analysis

Gene-based association meta-analyses were performed to identify genes associated with BMI-related phenotypes. As gene-based analyses require information of LD structure, which may diverge between different ethnic groups, we tested gene-based association within each ethnic group, and then combined signals from all groups into a summary statistic. Quality control of SNPs for inclusion was the same as that for SNP-based analyses: imputation accuracy $r^2 > 0.3$ and MAF > 0.01. For ethnic groups containing only one individual GWA study (COS, WHI-AA, WHI-HIS), *P*-values from that study were analyzed individually with VEGAS (29). For the Caucasian group containing multiple GWA studies (OOS, KCOS, FHS and IFS), those multiple studies were meta-analyzed first, and then the combined *P*-values were analyzed with VEGAS (29). In using VEGAS, we adopted the following steps: (i) SNPs were assigned to genes according to their physical positions at the UCSC Genome Browser hg18 assembly. Gene boundaries were defined as 50 kb from the 5′ and 3′ UTRs; (ii) for a gene containing *n* SNPs, individual SNP *P*-values were converted to upper tail Chi-squared statistics with 1 degree of freedom. Then the gene-based statistic was defined as the sum of these Chi-squared statistics; (iii) using the Monte Carlo approach, a large number of multivariate normal vectors *z* with mean 0 and variance matrix of pairwise LD *r* values were simulated. The *r* values were estimated from the respective HapMap phase 2 samples (30) with the closest ancestry. The simulated gene-based test statistics was the sum of squares of the elements of *z*; and (iv) the empirical gene-based *P*-value was calculated as the proportion of simulated test statistics that exceeded the observed gene-based statistic. Gene-based *P*-values from different ethnic populations were combined by the Fisher's method (67).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank Dr Andrew P. Morris (Wellcome Trust Centre for Human Genetics) for providing us with the MANTRA software. Thanks also to the reviewer and editor for their constructive comments and suggestions.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by grants from the National Institutes of Health (P50AR055081, R01AG026564, R01AR050496, RC2DE020756, R01AR057049 and R03TW008221); grant from the National Natural Science Foundation of China (31100902); Shanghai Leading Academic Discipline Project (S30501); and young teacher startup funds from Shanghai University of Science and Technology (slg11018 and slg11019). The Framingham Heart Study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University (Contract No. N01-HC-25195). This manuscript was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University or NHLBI. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. Funding support for the Framingham Whole Body DXA dataset was provided by NIH grants R01 AR/AG 41398. The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap through dbGaP accession phs000007.v14.p6. Funding support for the Genetic Determinants of Bone Fragility was provided through the NIA Division of Geriatrics and Clinical Gerontology. Genetic Determinants of Bone Fragility is a genome-wide association study funded as part of the NIA Division of Geriatrics and Clinical Gerontology. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the NIA Division of Geriatrics and Clinical Gerontology and the NIA Division of Aging Biology. Support for the collection of data sets and samples was provided by the parent grant, Genetic Determinants of Bone Fragility (P01-AG018397). Funding support for the genotyping which was performed at the Johns Hopkins University Center for Inherited Diseases Research was provided by the NIH NIA. The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap through dbGaP accession phs000138.v2.p1. The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts N01WH22110, 24152, 32100-2, 32105-6, 32108-9, 32111-13, 32115, 32118-32119, 32122, 42107-26, 42129-32 and 44221. This manuscript was not prepared in collaboration with investigators of the WHI, has not been reviewed and/or approved by the Women's Health Initiative (WHI) and does not necessarily reflect the opinions of the WHI investigators or the NHLBI.WHI PAGE is funded through the NHGRI Population Architecture Using Genomics and Epidemiology (PAGE) network (Grant Number U01 HG004790). Assistance with phenotype harmonization, SNP selection, data cleaning, meta-analyses, data management and dissemination, and general study coordination, was provided by the PAGE Coordinating Center (U01HG004801-01). The data sets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap through dbGaP accession phs000200.v6.p2.

REFERENCES

1. Kopelman, P.G. (2000) Obesity as a medical problem. *Nature*, **404**, 635–643.
2. Ogden, C.L., Carroll, M.D., Curtin, L.R., McDowell, M.A., Tabak, C.J. and Flegal, K.M. (2006) Prevalence of overweight and obesity in the United States, 1999–2004. *JAMA*, **295**, 1549–1555.
3. Wolf, A.M. and Colditz, G.A. (1998) Current estimates of the economic cost of obesity in the United States. *Obes. Res.*, **6**, 97–106.
4. Albanese, C.V., Diessel, E. and Genant, H.K. (2003) Clinical applications of body composition measurements using DXA. *J. Clin. Densitom*, **6**, 75–85.
5. Bell, C.G., Walley, A.J. and Froguel, P. (2005) The genetics of human obesity. *Nat. Rev. Genet.*, **6**, 221–234.
6. Deng, H.W., Lai, D.B., Conway, T., Li, J., Xu, F.H., Davies, K.M. and Recker, R.R. (2001) Characterization of genetic and lifestyle factors for

determining variation in body mass index, fat mass, percentage of fat mass, and lean mass. *J. Clin. Densitom*, **4**, 353–361.
7. 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.
8. Stunkard, A.J., Foch, T.T. and Hrubec, Z. (1986) A twin study of human obesity. *JAMA*, **256**, 51–54.
9. Saunders, C.L., Chiodini, B.D., Sham, P., Lewis, C.M., Abkevich, V., Adeyemo, A.A., de Andrade, M., Arya, R., Berenson, G.S., Blangero, J. *et al.* (2007) Meta-analysis of genome-wide linkage studies in BMI and obesity. *Obesity (Silver Spring)*, **15**, 2263–2275.
10. Melka, M.G., Bernard, M., Mahboubi, A., Abrahamowicz, M., Paterson, A.D., Syme, C., Lourdusamy, A., Schumann, G., Leonard, G.T., Perron, M. *et al.* (2012) Genome-wide scan for loci of adolescent obesity and their relationship with blood pressure. *J. Clin. Endocrinol. Metab.*, **97**, E145–E150.
11. Jiao, H., Arner, P., Hoffstedt, J., Brodin, D., Dubern, B., Czernichow, S., van't Hooft, F., Axelsson, T., Pedersen, O., Hansen, T. *et al.* (2011) Genome wide association study identifies KCNMA1 contributing to human obesity. *BMC Med. Genomics*, **4**, 51.
12. Wang, K., Li, W.D., Zhang, C.K., Wang, Z., Glessner, J.T., Grant, S.F., Zhao, H., Hakonarson, H. and Price, R.A. (2011) A genome-wide association study on obesity and obesity-related traits. *PLoS ONE*, **6**, e18939.
13. Cotsapas, C., Speliotes, E.K., Hatoum, I.J., Greenawalt, D.M., Dobrin, R., Lum, P.Y., Suver, C., Chudin, E., Kemp, D., Reitman, M. *et al.* (2009) Common body mass index-associated variants confer risk of extreme obesity. *Hum. Mol. Genet.*, **18**, 3502–3507.
14. Norris, J.M., Langefeld, C.D., Talbert, M.E., Wing, M.R., Haritunians, T., Fingerlin, T.E., Hanley, A.J., Ziegler, J.T., Taylor, K.D., Haffner, S.M. *et al.* (2009) Genome-wide association study and follow-up analysis of adiposity traits in Hispanic Americans: the IRAS Family Study. *Obesity (Silver Spring)*, **17**, 1932–1941.
15. Meyre, D., Delplanque, J., Chevre, J.C., Lecoeur, C., Lobbens, S., Gallina, S., Durand, E., Vatin, V., Degraeve, F., Proenca, C. *et al.* (2009) Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat. Genet.*, **41**, 157–159.
16. Liu, Y.J., Liu, X.G., Wang, L., Dina, C., Yan, H., Liu, J.F., Levy, S., Papasian, C.J., Drees, B.M., Hamilton, J.J. *et al.* (2008) Genome-wide association scans identified CTNNBL1 as a novel gene for obesity. *Hum. Mol. Genet.*, **17**, 1803–1813.
17. Scuteri, A., Sanna, S., Chen, W.M., Uda, M., Albai, G., Strait, J., Najjar, S., Nagaraja, R., Orru, M., Usala, G. *et al.* (2007) Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet.*, **3**, e115.
18. Hinney, A., Nguyen, T.T., Scherag, A., Friedel, S., Bronner, G., Muller, T.D., Grallert, H., Illig, T., Wichmann, H.E., Rief, W. *et al.* (2007) Genome wide association (GWA) study for early onset extreme obesity supports the role of fat mass and obesity associated gene (FTO) variants. *PLoS ONE*, **2**, e1361.
19. Herbert, A., Gerry, N.P., McQueen, M.B., Heid, I.M., Pfeufer, A., Illig, T., Wichmann, H.E., Meitinger, T., Hunter, D., Hu, F.B. *et al.* (2006) A common genetic variant is associated with adult and childhood obesity. *Science*, **312**, 279–283.
20. Paternoster, L., Evans, D.M., Nohr, E.A., Holst, C., Gaborieau, V., Brennan, P., Gjesing, A.P., Grarup, N., Witte, D.R., Jorgensen, T. *et al.*. (2011) Genome-wide population-based association study of extremely overweight young adults—the GOYA study. *PLoS ONE*, **6**, e24303.
21. Cho, Y.S., Go, M.J., Kim, Y.J., Heo, J.Y., Oh, J.H., Ban, H.J., Yoon, D., Lee, M.H., Kim, D.J., Park, M. *et al.* (2009) A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nat. Genet.*, **41**, 527–534.
22. Bradfield, J.P., Taal, H.R., Timpson, N.J., Scherag, A., Lecoeur, C., Warrington, N.M., Hypponen, E., Holst, C., Valcarcel, B., Thiering, E. *et al.* (2012) A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat. Genet.*, **44**, 526–531.
23. Scherag, A., Dina, C., Hinney, A., Vatin, V., Scherag, S., Vogel, C.I., Muller, T.D., Grallert, H., Wichmann, H.E., Balkau, B. *et al.*. (2010) Two new loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and German study groups. *PLoS Genet.*, **6**, e1000916.
24. 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., Magi, R. *et al.* (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.*, **42**, 937–948.
25. 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.
26. 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.* (2009) Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.*, **41**, 25–34.
27. 1000 Genomes Project ConsortiumAbecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E. and McVean, G.A. (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061–1073.
28. Neale, B.M. and Sham, P.C. (2004) The future of association studies: gene-based analysis and replication. *Am. J. Hum. Genet.*, **75**, 353–362.
29. Liu, J.Z., McRae, A.F., Nyholt, D.R., Medland, S.E., Wray, N.R., Brown, K.M., Hayward, N.K., Montgomery, G.W., Visscher, P.M., Martin, N.G. *et al.* (2010) A versatile gene-based test for genome-wide association studies. *Am. J. Hum. Genet.*, **87**, 139–145.
30. Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M. *et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851–861.
31. Kilpelainen, T.O., Zillikens, M.C., Stancakova, A., Finucane, F.M., Ried, J.S., Langenberg, C., Zhang, W., Beckmann, J.S., Luan, J., Vandenput, L. *et al.* (2011) Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nat. Genet.*, **43**, 753–760.
32. 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., Magi, R. *et al.* (0000) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.*, **42**, 937–948.
33. 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.
34. Loos, R.J., Lindgren, C.M., Li, S., Wheeler, E., Zhao, J.H., Prokopenko, I., Inouye, M., Freathy, R.M., Attwood, A.P., Beckmann, J.S. *et al.* (2008) Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat. Genet.*, **40**, 768–775.
35. Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S. *et al.* (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, **467**, 832–838.
36. Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R., Elliott, K.S., Lango, H., Rayner, N.W. *et al.* (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, **316**, 889–894.
37. Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA*, **106**, 9362–9367.
38. Hooton, H., Angquist, L., Holst, C., Hager, J., Rousseau, F., Hansen, R.D., Tjonneland, A., Roswall, N., van der A, D.L., Overvad, K. *et al.*. (2012) Dietary factors impact on the association between CTSS variants and obesity related traits. *PLoS ONE*, **7**, e40394.
39. Taleb, S., Cancello, R., Clement, K. and Lacasa, D. (2006) Cathepsin s promotes human preadipocyte differentiation: possible involvement of fibronectin degradation. *Endocrinology*, **147**, 4950–4959.
40. Taleb, S., Lacasa, D., Bastard, J.P., Poitou, C., Cancello, R., Pelloux, V., Viguerie, N., Benis, A., Zucker, J.D., Bouillot, J.L. *et al.* (2005) Cathepsin S, a novel biomarker of adiposity: relevance to atherogenesis. *FASEB J.*, **19**, 1540–1542.
41. Taleb, S. and Clement, K. (2007) Emerging role of cathepsin S in obesity and its associated diseases. *Clin. Chem. Lab Med.*, **45**, 328–332.
42. Naour, N., Rouault, C., Fellahi, S., Lavoie, M.E., Poitou, C., Keophiphath, M., Eberle, D., Shoelson, S., Rizkalla, S., Bastard, J.P. *et al.* (2010) Cathepsins in human obesity: changes in energy balance predominantly affect cathepsin s in adipose tissue and in circulation. *J. Clin. Endocrinol. Metab.*, **95**, 1861–1868.
43. Ishitani, T., Ninomiya-Tsuji, J., Nagai, S., Nishita, M., Meneghini, M., Barker, N., Waterman, M., Bowerman, B., Clevers, H., Shibuya, H. *et al.*

(1999) The TAK1-NLK-MAPK-related pathway antagonizes signalling between beta-catenin and transcription factor TCF. *Nature*, **399**, 798–802.
44. Ross, S.E., Hemati, N., Longo, K.A., Bennett, C.N., Lucas, P.C., Erickson, R.L. and MacDougald, O.A. (2000) Inhibition of adipogenesis by Wnt signaling. *Science*, **289**, 950–953.
45. Marks, D.L., Ling, N. and Cone, R.D. (2001) Role of the central melanocortin system in cachexia. *Cancer Res.*, **61**, 1432–1438.
46. Rolland, Y., Onder, G., Morley, J.E., Gillette-Guyonet, S., Abellan van Kan, G. and Vellas, B. (2011) Current and future pharmacologic treatment of sarcopenia. *Clin. Geriatr. Med.*, **27**, 423–447.
47. Farooqi, I.S., Keogh, J.M., Yeo, G.S., Lank, E.J., Cheetham, T. and O'Rahilly, S. (2003) Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N. Engl. J. Med.*, **348**, 1085–1095.
48. Scarlett, J.M., Bowe, D.D., Zhu, X., Batra, A.K., Grant, W.F. and Marks, D.L. (2010) Genetic and pharmacologic blockade of central melanocortin signaling attenuates cardiac cachexia in rodent models of heart failure. *J. Endocrinol.*, **206**, 121–130.
49. Economou, M., Trikalinos, T.A., Loizou, K.T., Tsianos, E.V. and Ioannidis, J.P. (2004) Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis. *Am. J. Gastroenterol.*, **99**, 2393–2404.
50. Greene, C.S., Penrod, N.M., Williams, S.M. and Moore, J.H. (2009) Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS ONE*, **4**, e5639.
51. Mailman, M.D., Feolo, M., Jin, Y., Kimura, M., Tryka, K., Bagoutdinov, R., Hao, L., Kiang, A., Paschall, J., Phan, L. *et al.* (2007) The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.*, **39**, 1181–1186.
52. Dawber, T.R., Meadors, G.F. and Moore, F.E. Jr. (1951) Epidemiological approaches to heart disease: the Framingham Study. *Am. J. Public Health Nations Health*, **41**, 279–281.
53. Feinleib, M., Kannel, W.B., Garrison, R.J., McNamara, P.M. and Castelli, W.P. (1975) The Framingham Offspring Study. Design and preliminary data. *Prev. Med.*, **4**, 518–525.
54. Koller, D.L., Ichikawa, S., Lai, D., Padgett, L.R., Doheny, K.F., Pugh, E., Paschall, J., Hui, S.L., Edenberg, H.J., Xuei, X. *et al.* (2010) Genome-wide association study of bone mineral density in premenopausal European-American women and replication in African-American women. *J. Clin. Endocrinol. Metab.*, **95**, 1802–1809.
55. 1998) Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. *Control Clin. Trials*, **19**, 61–109.
56. 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. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
57. Li, Y., Willer, C.J., Ding, J., Scheet, P. and Abecasis, G.R. (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.*, **34**, 816–834.
58. Li, Y., Willer, C., Sanna, S. and Abecasis, G. (2009) Genotype imputation. *Annu. Rev. Genomics Hum. Genet.*, **10**, 387–406.
59. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. and Abecasis, G.R. (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.*, **44**, 955–959.
60. Pei, Y.F., Zhang, L., Li, J. and Deng, H.W. (2010) Analyses and comparison of imputation-based association methods. *PLoS ONE*, **5**, e10827.
61. Zhang, L., Li, J., Pei, Y.F., Liu, Y. and Deng, H.W. (2009) Tests of association for quantitative traits in nuclear families using principal components to correct for population stratification. *Ann. Hum. Genet.*, **73**, 601–613.
62. Chen, W.M. and Abecasis, G.R. (2007) Family-based association tests for genomewide association scans. *Am. J. Hum. Genet.*, **81**, 913–926.
63. Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics*, **55**, 997–1004.
64. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.
65. Lumley, T. (2009) rmeta: Meta-analysis. R package, version 2.16.
66. Morris, A.P. (2011) Transethnic meta-analysis of genomewide association studies. *Genet. Epidemiol.*, **35**, 809–822.
67. Fisher, R.A. (1925) *Statistical Methods for Research Workers*. Oliver and Boyd (Edinburgh).