

Common Variation in *SIM1* Is Reproducibly Associated With BMI in Pima Indians

Michael Traurig,¹ Janel Mack,¹ Robert L. Hanson,¹ Maya Ghoussaini,² David Meyre,² William C. Knowler,¹ Sayuko Kobes,¹ Philippe Froguel,² Clifton Bogardus,¹ and Leslie J. Baier¹

OBJECTIVE—Haploinsufficiency of *SIM1* is a cause of rare monogenic obesity. To assess the role of *SIM1* in polygenic obesity, this gene was analyzed in the Pima Indian population, which has a high prevalence of obesity.

RESEARCH DESIGN AND METHODS—*SIM1* was sequenced in 96 individuals. Variants ($n = 46$) were genotyped in a population-based sample of 3,250 full-heritage Pima Indians and in a separate replication sample of 2,944 predominantly non-full-heritage subjects from the same community.

RESULTS—Variants spanning the upstream region of *SIM1* through intron 8 were associated with BMI in the full-heritage Pima Indians, where the strongest associations ($P \sim 10^{-4}$ to 10^{-6}) were with common variants (risk allele frequency 0.61–0.67). The difference in mean BMI between individuals homozygous for the major allele compared with homozygotes for the minor allele was ~ 2.2 kg/m² ($P = 2 \times 10^{-5}$ for rs3213541). These associations replicated in the separate sample of subjects from the same community ($P = 5 \times 10^{-3}$ for rs3213541). The strongest associations ($P = 4 \times 10^{-7}$, controlled for age, sex, birth year, and heritage) were seen in the combined sample ($n = 6,194$). The risk allele for obesity was more common in full-heritage Pimas than in the mixed-heritage subjects. Two variants (rs3734353 and rs3213541) were also genotyped in 1,275 severely obese and 1,395 lean control subjects of French European ancestry. The Pima risk alleles were the minor alleles in the European samples, and these variants did not display any significant association ($P > 0.05$).

CONCLUSIONS—Common variation in *SIM1* is associated with BMI on a population level in Pima Indians where the risk allele is the major allele. *Diabetes* 58:1682–1689, 2009

Obesity is a major cause of health disparities among minority populations. The few major reports of genes contributing to polygenic obesity have predominately focused on populations of European descent. The Pima Indians of Arizona have a high rate of obesity and, as a consequence, have an extraordinarily high prevalence of type 2 diabetes (1). To identify genetic variation that contributes to obesity in Pimas, genes that have a potential role in body weight

regulation are being analyzed, and ~ 50 biologic candidate genes have been analyzed to date with no conclusive association with obesity reported up to now in this specific population. The most significant findings are with the transcription factor *SIM1*, which is the human homolog of the *Drosophila* single-minded (*sim*) developmental gene (2–4). *SIM1* is essential for the development of the hypothalamic paraventricular nucleus (5), which plays a central role in the regulation of appetite and body weight (6). Several chromosomal deletions and a balanced translocation involving chromosome 6q16–6q21, which encompasses the *SIM1* locus, have been identified in children with monogenic severe obesity (7–14). Mice haploinsufficient for *SIM1* are hyperphagic and are sensitive to diet-induced obesity (15–17), whereas mice overexpressing *SIM1* have decreased food intake (18,19). Therefore, based on the biological role of *SIM1* and the human and mouse genetic studies, *SIM1* was analyzed as a candidate gene for obesity in the Pima Indians of Arizona.

RESEARCH DESIGN AND METHODS

All subjects are part of an ongoing population-based longitudinal study of the etiology of type 2 diabetes in the Gila River Indian Community in Central Arizona, where most of the residents are Pima Indians (20). For the initial studies of *SIM1*, all individuals ($n = 3,250$) were selected who had available DNA and measures of diabetes and BMI after the age of 15 years and whose heritage was reported as full Pima and/or Tohono O'odham (a closely related tribe). For subsequent replication studies, all participants ($n = 2,944$) were selected who had available DNA and measures of BMI regardless of heritage (mixed-heritage replication set). Thus, most of the individuals in the replication set were not of full Pima heritage (on average their reported heritage was 1/2 Pima and 3/4 American Indian, which may include other tribes), and there were 72 residents who reported no Native American heritage.

However, 140 full-heritage Pima individuals whose DNA was not available when the initial full-heritage samples were collected were also included in the replication study. BMI was computed for all exams at age ≥ 15 years and the maximum BMI was used for analysis. Because BMI can be influenced by diabetes progression and disease treatment, the maximum BMI measured at a nondiabetic exam was additionally analyzed. Subjects who did not have a BMI recorded from a nondiabetic exam at age ≥ 15 years (i.e., developed diabetes before age 15 years or were first seen at our clinic when they were older and had already developed diabetes) were excluded from these analyses; therefore, the nondiabetic BMI analyses were restricted to 2,789 subjects for the full-heritage sample and 2,647 subjects for the replication sample. Thus, the combined sample including the initial full-heritage individuals and those included in the mixed-heritage replication study included 6,194 individuals in total and 5,436 who had been examined when nondiabetic. All subjects provided written informed consent before participation. This study was approved by the Institutional Review Board of the National Institute of Diabetes and Digestive and Kidney Diseases.

French obese children (BMI greater than the 97th percentile for sex and age) were recruited at the CNRS-UMR8090 Unit in Lille ($n = 420$), at the Children's Hospital, Toulouse ($n = 92$), and at the Trousseau Hospital ($n = 90$). The obese adult subgroup consisted of 673 morbidly obese (BMI ≥ 40 kg/m²) adults collected at the Department of Nutrition of the Hôtel Dieu Hospital in Paris or at the CNRS-UMR8090 Unit in Lille. The control group consisted of 1,395 nonobese (BMI < 27 kg/m²) normoglycemic (fasting glycemia < 5.56 mmol/l) French Caucasian adults pooled from four separate studies: 394 subjects were recruited at the CNRS-UMR8090, 265 were re-

From the ¹Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Disease, National Institutes of Health, Phoenix, Arizona; and ²CNRS 8090–Institute of Biology, Pasteur Institute, Lille, France.

Corresponding author: Leslie Baier, lbaier@phx.niddk.nih.gov.

Received 7 January 2009 and accepted 16 April 2009.

Published ahead of print at <http://diabetes.diabetesjournals.org> on 28 April 2009. DOI: 10.2337/db09-0028.

© 2009 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. See <http://creativecommons.org/licenses/by-nc-nd/3.0/> for details.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

cruited by the Fleurbaix-Laventie Ville Santé study (21), 365 from the HAGUENEAU study (22), and 371 from the SUVIMAX study (23).

Sequencing and genotyping. To identify novel variants, the coding regions, 5'- and 3'-untranslated regions (UTRs), 2.1 kb upstream of the promoter, and a 2-kb conserved region in intron 8 of *SIM1* were sequenced in 96 obese Pima subjects (BMI 50–80 kg/m²). Sequencing was done using Big Dye terminator (Applied Biosystems) on an automated DNA capillary sequencer (model 3730; Applied Biosystems). In addition to the variants identified by sequencing ($n = 16$), 30 tag SNPs with a minor allele frequency ≥ 0.1 and an $r^2 \geq 0.8$ that cover the unsequenced intronic regions were selected from the HapMap Chinese (CHB) population. The 46 variants were genotyped in the full-heritage Pima Indians, and most of the variants ($n = 43$) were genotyped in the mixed-heritage replication sample using the SNplex Genotyping System 48-plex (Applied Biosystems) on an automated DNA capillary sequencer (model 3730; Applied Biosystems). All Pima Indian genotypic data passed our quality control criteria, which require a successful genotypic call rate on $>85\%$ of the samples, a deviation from Hardy-Weinberg equilibrium of $P > 0.001$, and a blind duplicate genotyping of 330 samples with a discrepancy rate of $<2.5\%$. The chromosomal locations and flanking sequence for the five novel variants (SIM1-1, Arg665His, Thr361Ile, SIM1-2, and SIM1-3) that were genotyped are shown in an online-only appendix (Table A1, available at <http://diabetes.diabetesjournals.org/cgi/content/full/db09-0028/DC1>). Variants rs3734353 and rs3213541 were genotyped in the French cohort using Taqman Technology (Applied Biosystems). Genotyping error rate calculated from duplicate genotypes of 250 individuals was 0% for both variants. In addition, the two SNPs were in Hardy-Weinberg equilibrium.

Statistical analysis. Statistical analyses were performed using the SAS of the SAS Institute (Cary, NC). For continuous variables, linear regression models were used to assess the association between BMI and genotype (assuming an additive model) with adjustment for covariates including age, sex, and birth year; the logarithmic transformation of BMI was taken in these analyses to reduce skewness. In the mixed-heritage replication group, the individual estimate of European admixture was also used as a covariate; these estimates were derived by the method of Hanis et al. (24) from 32 markers with large differences in allele frequency between populations (25). The generalized estimating equation procedure was used to account for family membership, since some subjects were siblings. Although the generalized estimating equation procedure accounts for the familial nature of the data, this test is not robust to population stratification. To provide a test that is robust to stratification, a modification of the method of Abecasis et al. (26) was used in which the association is partitioned into between- and within-family components.

A test of the equivalence of the between- and within-family effects provides a test of the null hypothesis of no population stratification effect for the marker in question. A combined test of association for the full-heritage and replication groups was conducted by the inverse variance method (27). Linkage disequilibrium (D' and r^2) was analyzed using the Haploview program (Haploview, <http://www.broad.mit.edu/mpg/haploview>). Haplotype "blocks" were defined using the four-gamete method, with a gametic frequency >0.01 taken as indicative of significant recombination (28), and analyses of the association of haplotypes within each block with BMI were conducted. In these analyses, one to four variants were determined to define the common haplotypes (frequency >0.01) within each block. The probability that an individual carried one or two copies of each of the common haplotypes within a block was calculated by modification of the zero-recombinant haplotype method as previously described (29). The MLINK program (30) was used to assign a probability of carriage of a given haplotype, and these probabilities were analyzed in a fashion analogous to that for individual variants. An "exhaustive" analysis was conducted in which all common haplotypes for all possible combinations of one to four variants within each block were analyzed. For the French Caucasian cohort, tests for deviation from the Hardy-Weinberg equilibrium and for association were performed with the De Finetti program (<http://ihg.gsf.de/cgi-bin/hw/hwa1.pl>).

RESULTS

Sequencing and genotyping. To identify novel variants in *SIM1*, all 11 exons, the 5'- and 3'-UTRs, 2.1 kb upstream of the 5'-UTR, and a 2-kb highly conserved region in intron 8 immediately adjacent to exon 8 were sequenced in 96 obese (BMI ≥ 50 kg/m²) Pima Indians. Sequencing of these regions identified several variants, including two previously identified missense substitutions in exon 9, Pro352Thr (rs3734354), and Ala371Val (rs3734355), and five novel polymorphisms (sequences are shown in online

Table A1), which included two rare nonsynonymous amino acid changes in exons 9 and 11 (Thr361Ile and Arg665His, respectively). Variants ($n = 16$) identified by sequencing along with an additional 30 tag SNPs chosen from the dbSNP public database spanning 35 kb upstream of *SIM1* to 25 kb downstream of *SIM1* were genotyped in a population-based sample of 3,250 full-heritage Pima Indians (online Table A2). The 46 variants fell into seven haplotype blocks (A–G, defined by the four-gamete method; Fig. 1A). The two largest blocks were block B, which spans much of the 3' region of *SIM1*, and block F, which spans half of intron 8 through the 5' region flanking *SIM1* (Fig. 1A). Several variants in blocks E and F were significantly associated with BMI in the full-heritage Pima Indians (P values ranging from 5×10^{-3} – 7×10^{-6} ; adjusted for age, sex, and birth year; Fig. 1B). These variants were associated with the maximum recorded BMI from any exam after the age of 15 years ($n = 3,250$) as well as the maximum BMI recorded at a nondiabetic exam after the age of 15 years ($n = 2,789$) (online Table A2; representative variants for blocks E and F are shown in Table 1). The differences in mean BMIs for the individuals homozygous for the major allele (M/M) versus individuals homozygous for the minor allele (m/m) for either analysis were ~ 2.2 kg/m² (Table 1 and online Table A2). The four missense mutations, Pro352Thr (rs3734354), Thr361Ile, Ala371Val (rs3734355), and Arg665His, were not associated with BMI (online Table A2).

To assess whether the association with BMI could be replicated in a separate group of subjects, the variants were further genotyped in a population-based sample of individuals from the same longitudinal study, most of whom were of mixed heritage ($n = 2,944$, Pima heritage ranging from 0/8th to 8/8th). The variants in blocks E and F were reproducibly associated with maximum BMI from any exam, as well as maximum BMI from a nondiabetic exam in this group (mixed-heritage and mixed-heritage nondiabetic replication sets; online Table A2; representative variants are shown in Table 1). Combining the initial full-heritage Pima Indians set with the mixed-heritage replication set ($n = 6,194$) provided the strongest associations with BMI (e.g., rs3213541 $P = 4 \times 10^{-7}$; Table 1 and online Table A2). When the mean maximum BMIs based on genotypes for rs3213541 were stratified by age, an increase in BMI among the risk (G) allele carriers for rs3213541 was observed at nearly all ages (Fig. 2A), and this increase appears to be consistent for both men and women (Fig. 2B and C). There was no significant interaction with age, suggesting that the fluctuation observed in men (Fig. 2B) after the age of 45 years is likely due to a smaller sample size. The magnitude of the BMI difference is similar among individuals who are predominately Pima heritage (more than half Pima heritage) compared with individuals who are predominately of different heritage (less than half Pima heritage) (Fig. 3A and B, respectively), except at older ages (>45 years), where the number of subjects becomes small, making the values of mean BMI somewhat less reliable. However, the alleles associated with high BMI (risk alleles) are less common among the mixed-heritage individuals compared with the full-heritage Pima Indians (Table 1 and online Table A2). For example, the risk (G) allele for increased BMI for rs3213541 has a frequency of 0.62 among the full-heritage Pima subjects and 0.54 among mixed-heritage subjects. Among the 72 individuals for whom there was no reported American Indian heritage, the frequency of the rs3213541 G allele

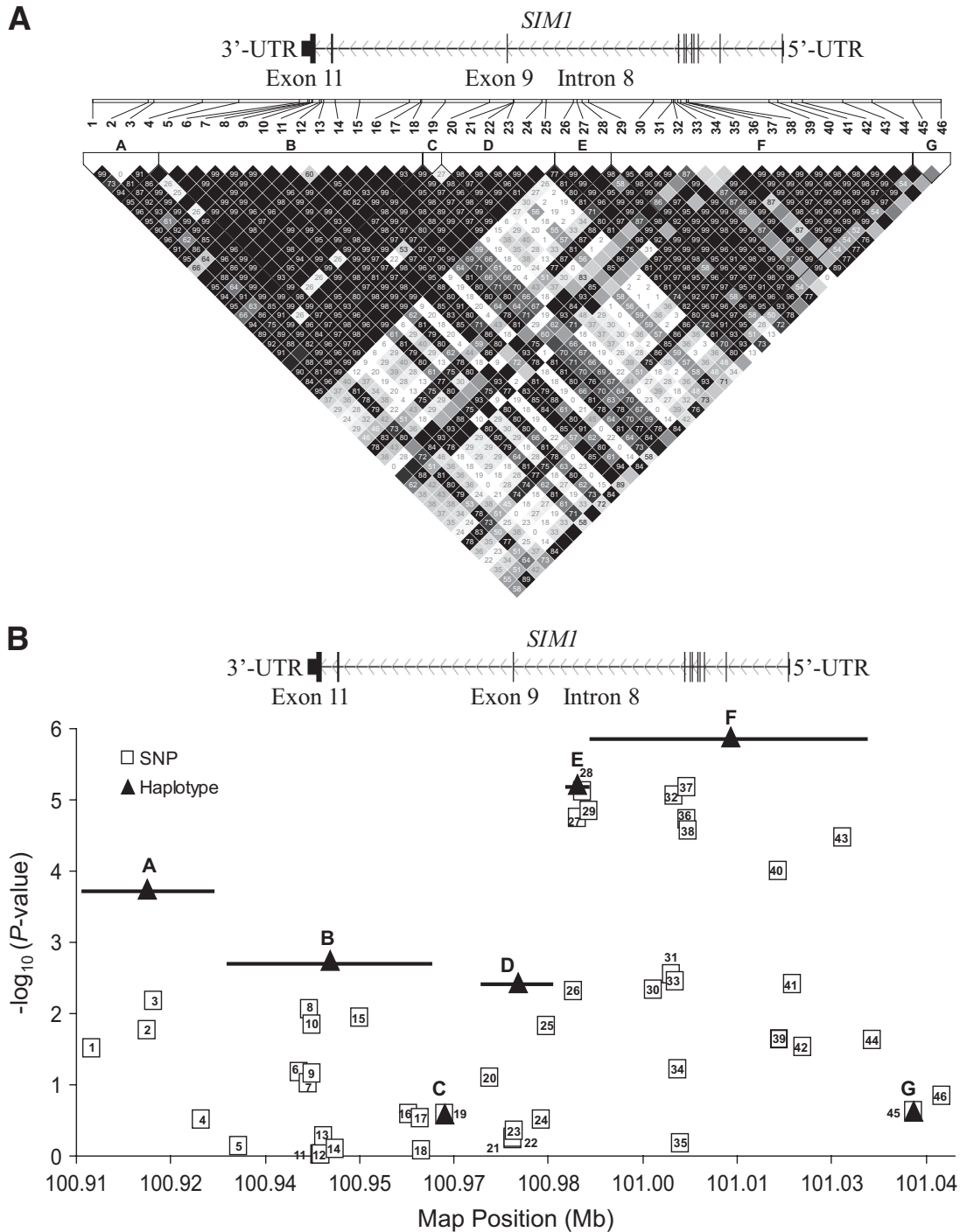


FIG. 1. Linkage disequilibrium plot (A) and association analyses (B) for the 46 variants genotyped in *SIM1* and adjacent 3' and 5' regions. **A:** Linkage disequilibrium is shown as D' . The 46 variants separate into seven haplotype blocks (A–G). Haplotypes were determined using the four-gamete method. The structure of the human *SIM1* gene (minus strand) is shown above the linkage disequilibrium plot. The variant numbers (1–46) correspond to the variant numbers in online Table A2. **B:** Plot of additive P values for the associations between the 46 individual variants and maximum BMI among a population-based sample of 3,250 full-heritage Pima Indians. Open boxes are the 46 individual variants that were genotyped. The numbers in the open boxes correspond to the numbers shown in Fig. 1A. The best P value obtained within each of the seven haplotypes (A–G) across this region is also shown. P values are plotted at the haplotype's midpoint (▲). The haplotype boundaries are delineated by the black vertical lines. Details of the haplotypes are shown in Table 2. All P values were adjusted for age, sex, nuclear family membership, and birth year.

was 0.40. HapMap data for rs3213541 show that the G allele is the minor allele for all four populations (G allele frequencies: Caucasians, 0.36; Chinese, 0.34; Japanese, 0.44; and Africans, 0.13 [International HapMap Project]); therefore, the risk allele for obesity in Pima Indians is the major allele but it is the minor allele in non-Native

American populations. Given the allele frequency differences of rs3213541 (and other variants in linkage disequilibrium) between American Indians and other populations, it is possible that these associations with BMI could be influenced by admixture even though our analyses controlled for heritage. Therefore, within-family association

TABLE 1
 Representative variants from linkage disequilibrium blocks E and F and their associations with BMI in the full-heritage Pima Indian, mixed-heritage replication, and combined (full-heritage + mixed-heritage) population sets

Variant*	M/m	mAF	Full-heritage Pima population [mean BMI† (n)]					Mixed-heritage replication [mean BMI† (n)]					Combined	
			MM	M/m	m/m	P‡	mAF	MM	M/m	m/m	P‡	P‡		
26. rs3798498	G/A	0.14	37.4 (2,271)	36.2 (711)	37.0 (53)	0.005	0.20	35.0 (1,784)	34.3 (906)	32.1 (114)	0.25	0.005		
			36.6 (1,959)	35.1 (614)	36.3 (45)	0.0007		34.2 (1,603)	33.8 (829)	31.1 (97)	0.50			
28. rs241816	C/T	0.33	37.8 (1,386)	36.8 (1,318)	35.5 (361)	0.000007	0.39	35.2 (1,072)	34.4 (1,347)	33.8 (438)	0.06	0.000006		
			37.0 (1,180)	35.8 (1,148)	34.9 (313)	0.000009		34.3 (950)	33.8 (1,229)	32.9 (396)	0.24			
30. rs241812	G/A	0.18	37.6 (1,986)	36.4 (876)	36.8 (79)	0.004	0.27	35.2 (1,425)	34.1 (1,053)	32.9 (171)	0.08	0.001		
			36.7 (1,718)	35.3 (749)	35.4 (69)	0.0002		34.4 (1,283)	33.7 (961)	32.1 (148)	0.32			
33. SIM1-2	G/A	0.004	37.1 (3,027)	41.9 (24)	—	0.003	0.002	34.6 (2,932)	39.3 (15)	—	0.10	0.001		
			36.2 (2,606)	40.5 (21)	—	0.008		33.9 (2,645)	36.3 (13)	—	0.24			
34. SIM1-3	G/A	0.05	37.3 (2,741)	35.8 (267)	38.6 (7)	0.06	0.04	34.7 (2,522)	33.2 (214)	30.7 (8)	0.01	0.003		
			36.4 (2,368)	34.7 (224)	38.3 (5)	0.02		34.1 (2,272)	32.7 (193)	29.7 (7)	0.05			
36. rs3213541	G/A	0.39	38.1 (1,154)	36.8 (1,387)	35.9 (456)	0.00002	0.46	35.6 (867)	34.6 (1,406)	33.4 (613)	0.005	0.000004		
			37.3 (976)	35.9 (1,215)	35.0 (393)	0.00004		34.7 (762)	34.0 (1,284)	32.6 (553)	0.06			
37. rs3830139	G/T	0.39	38.1 (1,166)	36.7 (1,433)	35.9 (452)	0.000007	0.45	35.6 (874)	34.6 (1,425)	33.4 (612)	0.005	0.000002		
			37.3 (890)	35.9 (1,063)	35.1 (386)	0.0002		34.7 (769)	34.0 (1,305)	32.7 (553)	0.06			
38. rs3734353	C/A	0.38	38.1 (1,138)	36.7 (1,330)	36.0 (443)	0.00003	0.45	35.6 (906)	34.6 (1,411)	33.4 (599)	0.005	0.000007		
			37.2 (971)	35.8 (1,150)	35.1 (387)	0.00007		34.6 (799)	34.1 (1,290)	32.7 (540)	0.06			
42. rs4840133	C/T	0.18	37.4 (2,059)	36.6 (909)	36.0 (89)	0.03	0.24	35.2 (1,675)	34.1 (1,076)	33.0 (171)	0.05	0.004		
			36.5 (1,777)	35.6 (781)	34.8 (78)	0.001		34.3 (1,508)	33.6 (980)	31.9 (147)	0.19			

*Numbers preceding the variant identifier correspond to the variant numbers in Fig. 1A and B and in online Table A2. M, major allele; m, minor allele as determined in full-heritage Pima Indians. mAF, minor allele frequency. †For each variant, the mean maximum BMI is given in the upper row and the mean maximum BMI restricted to only exams when the subjects were nondiabetic are shown in the lower row. ‡P values are given for an additive model and adjusted for age, sex, birth date, family membership, and Pima heritage. SIM1-2 and SIM1-3 are novel variants located at chromosomes 6:101001126 bps and 101001582 bps, respectively (Human Genome, Build 36.1). See online Table A1 for a more detailed description of the novel variants.

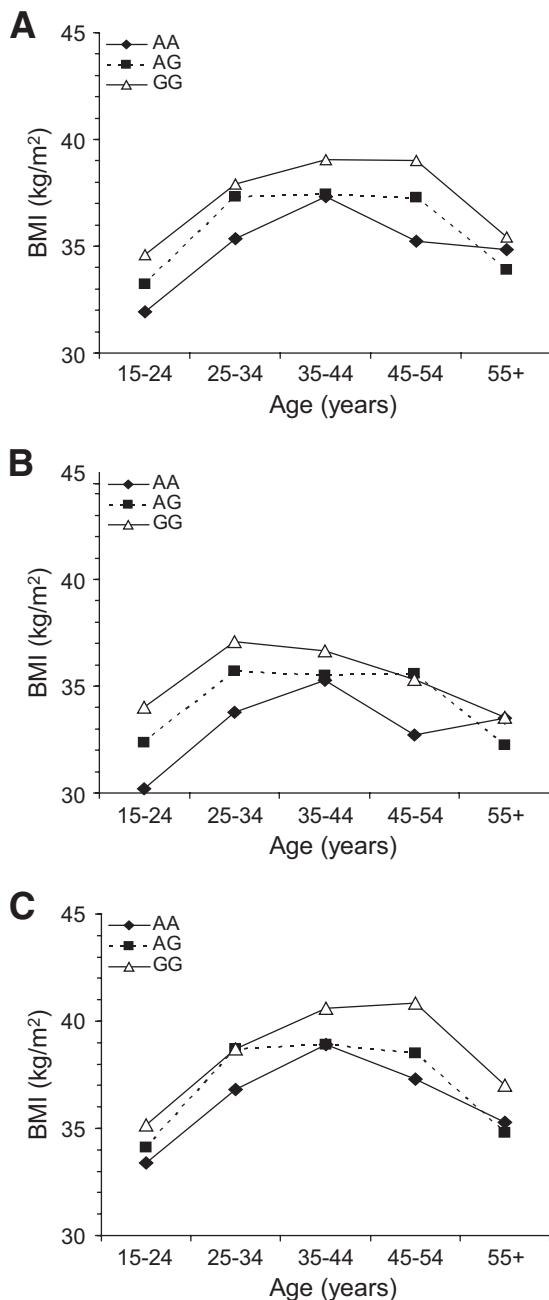


FIG. 2. Mean maximum BMIs based on genotypes for rs3213541 in the combined (full-heritage + mixed-heritage) population set stratified by age (A) or age and sex (B and C). The BMI is the maximum BMI recorded within the specified age range. The number of subjects in each age-group by genotype (GG, GA, and AA) is as follows: A: Both men and women combined: 15–24 ($n = 686, 1,021, 402$), 25–34 ($n = 512, 748, 235$), 35–44 ($n = 460, 593, 227$), 45–54 ($n = 217, 252, 120$), ≥ 55 ($n = 125, 154, 65$). B: Men: 15–24 ($n = 310, 478, 187$), 25–34 ($n = 221, 319, 102$), 35–44 ($n = 173, 247, 87$), 45–54 ($n = 69, 103, 60$), ≥ 55 ($n = 51, 57, 23$). C: Women: 15–24 ($n = 376, 543, 215$), 25–34 ($n = 291, 429, 133$), 35–44 ($n = 287, 346, 140$), 45–54 ($n = 148, 149, 69$), ≥ 55 ($n = 74, 97, 43$).

tests that are robust to population stratification were also done. The within-family analyses were less significant but consistent with the overall general associations ($P = 0.04 - 0.1$), while tests for population stratification were not statistically significant. Given the strong overall association, these results suggest that these associations are not solely the result of admixture.

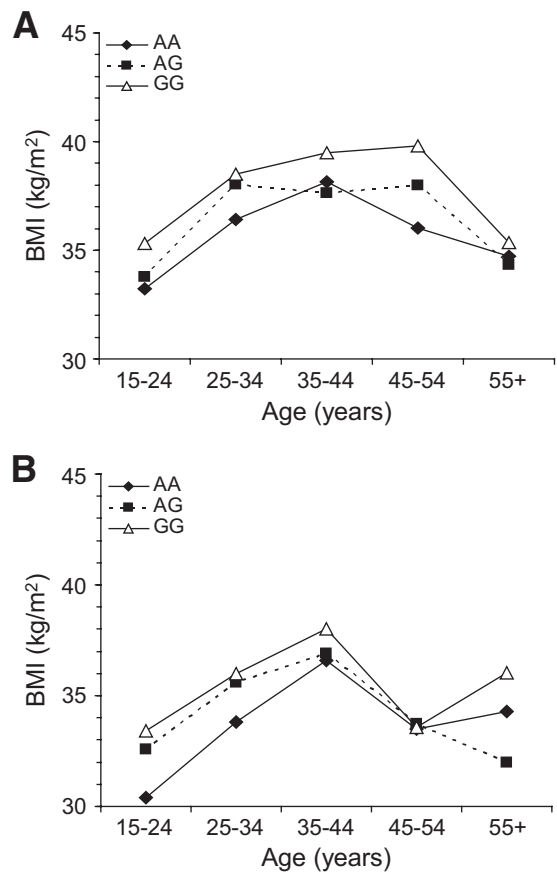


FIG. 3. Mean maximum BMIs based on genotypes for rs3213541 among subjects who are predominately Pima heritage (more than half Pima heritage) (A) and who are predominately of different heritage (less than half Pima heritage) (B) stratified by age. The BMI is the maximum BMI recorded within the specified age range. The number of subjects in each age-group by genotype (GG, GA, AA) is as follows: A: More than half Pima heritage: 15–24 ($n = 461, 628, 227$), 25–34 ($n = 408, 533, 150$), 35–44 ($n = 372, 434, 147$), 45–54 ($n = 191, 208, 88$), ≥ 55 ($n = 108, 123, 49$). B: Less than half Pima heritage: 15–24 ($n = 225, 393, 175$), 25–34 ($n = 104, 215, 85$), 35–44 ($n = 88, 159, 80$), 45–54 ($n = 26, 44, 41$), ≥ 55 ($n = 17, 31, 16$).

There was little, if any, association of these variants with type 2 diabetes, with only a few variants achieving nominal statistical significance ($P < 0.05$; see online Table A2). Most of the nominal associations were no longer significant after controlling for BMI. For example, the diabetes odds ratio for rs3213541 is 1.09 per copy of the G allele (95% CI 0.99–1.19, $P = 0.08$); after control for BMI, the odds ratio is attenuated to 1.05 per copy of the G allele (95% CI 0.96–1.16, $P = 0.31$).

To examine whether any of the variants had an effect in addition to the most strongly associated variants, associations were further analyzed conditional on that observed for rs3213541 for each of the 45 other variants. For variants that were not highly concordant with rs3213541 ($r^2 < 0.78$), the effect of rs3213541 remained significant ($P < 0.001$), whereas in most cases, the other variant was not significant ($P > 0.05$). The exceptions were the rare novel SIM1–2 variant located in intron 8 and the rare variant rs7766596, both of which remained significant despite controlling for rs3213541 (SIM1–2, $P = 0.01$, and rs7768342, $P = 0.03$); however, these associations are modest and may reflect chance findings. These results are consistent with the hypothesis that the primary associa-

TABLE 2

Tag SNP combinations providing the strongest association within each of seven haplotypes with BMI

Tag SNPs*	Individual SNPs		Haplotype†					Frequency	P‡	
	M/m	P	Allele 1	Allele 2	Allele 3	Allele 4	Allele 5			
Haplotype A										
2. rs10872573	T/C	0.02	T						0.78	2 × 10 ⁻⁴
3. rs17792983	C/T	0.007		C						
4. rs1506083	T/A	0.30			—					
Haplotype B										
9. rs17060480	A/T	0.07	A						0.16	2 × 10 ⁻³
12. SIM-1	C/T	0.95		—						
13. Arg665His	G/A	0.54			—					
15. rs3798514	A/G	0.01				G				
17. rs3798508	G/A	0.29					G			
Haplotype C										
19. rs7747189	G/A	0.25	G						0.58	0.25
Haplotype D										
20. rs9390366	T/G	0.29	T						0.16	4 × 10 ⁻³
22. Thr361Ile	C/T	0.56		—						
24. rs9386144	T/C	0.31			T					
25. rs3822980	T/C	0.02				T				
Haplotype E										
26. rs3798498	G/A	0.005	G						0.67	6 × 10 ⁻⁶
27. rs3798497	C/T	0.00002		—						
28. rs241816	C/T	0.000007			C					
Haplotype F										
34. SIM-3	G/A	0.06	G						0.63	1 × 10 ⁻⁶
41. rs3756915	G/A	0.004		G						
43. rs7773043	G/A	0.00003			G					
44. rs6910391	T/C	0.02				—				
Haplotype G										
45. rs6901854	C/T	0.24	C						0.75	0.24

*Tag SNPs (numbers correspond to Fig. 1A and B and online Table A2) are shown for each of the seven haplotype blocks (A–G). †All possible combinations of tag SNPs were analyzed for each block, and the haplotype providing the strongest association with BMI is given. The midpoints for each of the haplotypes are plotted in Fig. 1B. ‡P values are adjusted for age, sex, and birth year.

tion reflects the effect rs3213541 or a strongly concordant variant.

To assess whether a specific haplotype provided a stronger association than a single variant, tag SNPs that captured the common variation ($r^2 > 0.8$) within each of the seven haplotype blocks (Fig. 1A–G) were determined for the 46 variants spanning *SIM1*. The lowest P values obtained from tag SNP combinations within each block for the initial full-heritage set are listed in Table 2 and are also shown in Fig. 1B (plotted at the haplotype block's midpoint; black triangles). Consistent with the single variant analysis, the haplotypes providing the strongest associations with BMI were in haplotype blocks E and F (spanning intron 8 through the 5' region of *SIM1*) and were highly concordant with rs3213451 (Table 2).

To determine whether common variation in *SIM1* had a significant effect on obesity in a non-Native American population, two common variants (rs3734353 and rs3213541) were also genotyped in French Caucasian case/control samples consisting of 602 unrelated severely obese children, 673 unrelated morbidly obese adults, and 1,395 unrelated normoglycemic nonobese control subjects. The French Caucasians were selected for replication because in a prior genome-wide linkage study in this population, the most significant linkage for obesity was on chromosome 6q22.31–6q23.2, which contains the *SIM1* locus (31). The two variants were in high linkage disequilibrium

among the Caucasians ($r^2 = 0.87$ in HAPMAP), and neither were associated with obesity in the case/control samples (e.g., rs3734353: $P = 0.70$ for obese children/control subjects, $P = 0.38$ for obese adults/control subjects, $P = 0.43$ for obese children + obese adults/control subjects). The allele frequencies for these two variants differed between the two populations; for example, the risk allele (C) for rs3734353 is the major allele (frequency = 0.62) in Pima Indians but the minor allele (frequency = 0.30) in French Caucasians. However, the overall pattern of linkage disequilibrium across the *SIM1* locus is quite similar between these two ethnic groups (online Fig. A1). Because of the different study designs, the results in the French Caucasian subjects are not directly comparable with those in the Pima Indians. To obtain a comparable OR estimate, the predominately Native American samples (full-heritage Pima Indian and mixed-heritage samples) were classified into case (BMI >40 kg/m²; $n = 1,694$) and control (nondiabetic and BMI <30 kg/m²; $n = 1,272$) subjects. With this classification, the OR for severe obesity is 1.26 per copy of the G allele for rs3213541 compared with the estimate of 0.95 for French Caucasian subjects, and Cochran's Q test for homogeneity indicates significant heterogeneity between Pima Indian and French subjects ($P = 0.001$). However, there was no statistically significant interaction with self-identified American Indian heritage among the Pima Indian families ($P = 0.80$).

DISCUSSION

In this study, noncoding variants spanning intron 8 through to the 5' region of *SIM1* showed the strongest association with BMI in full-heritage Pima Indians, and these associations also replicated in a group of mixed-heritage individuals from the same community. Meyre et al. (31) had previously reported linkage to obesity in French Caucasians in the region of chromosome 6 that includes the *SIM1* locus, but the variants most strongly associated with BMI in Pima Indians did not replicate in the French population. Linkage to BMI on chromosome 6q was not identified in our prior genome-wide linkage scan in Pima Indians (32). Hung et al. (33) previously reported that the coding Ala371Val polymorphism (rs3734355) was modestly associated with BMI in Caucasian males; however, none of the coding variants—Pro352Thr (rs3734354), Ala371Val (rs3734355), Thr361Ile, and Arg665His—were associated with BMI in our study of Pima Indians. Ahituv et al. (34) also found no association between either the Ala371Val (rs3734355) or the Pro352Thr (rs3734354) in their case/control study for obesity in Caucasians.

Few genetic variants with reproducible associations with obesity have been identified, and most of those that have been identified have been studied primarily in European populations. The present study identifies several variants in *SIM1* that are reproducibly associated with BMI in individuals from a longitudinal study, most of whom have Pima Indian heritage. The issue of statistical significance in genetic association studies is complicated. Given the potential for artifactual associations and the fact that most variants are unlikely a priori to be associated with a given trait, most statisticians recommend a stringent threshold for declaring an association "significant." The exact *P* value threshold is a matter of controversy, but for genome-wide studies, a value in the vicinity of $P < 5 \times 10^{-8}$ (35,36) is generally considered significant. Whether the same level of stringency is required for a candidate gene, such as *SIM1*, is debatable, but the poor record of reproducibility for such variants suggests similar criteria may be appropriate. Such stringency, however, comes at the cost of potentially missing true associations, particularly in small populations in whom the potential for additional replication studies is limited. The present study achieves *P* values approaching levels of genome-wide significance for the association of *SIM1* variants with BMI, with replication observed in two separate samples of Pima Indians, comprising >6,000 individuals in toto. The effect of these common variants on BMI is notable, where the difference in mean BMI by genotype is ~2.2 kg/m² among the initial sample of full-heritage Pimas and ~2.4 kg/m² in the combined sample of full-heritage and mixed-heritage subjects, an effect corresponding to ~1% of the variance in BMI in the population. In comparison, the difference in mean BMI by genotype for *FTO*, the most highly replicated gene for polygenic obesity identified to date, is ~1 kg/m² among Caucasians (37) and 1.6 kg/m² among Pimas (38).

Our results also reflect a genetic heterogeneity between the French Caucasians and the Pima Indians. Indeed, two common variants (rs3734353 and rs3213541) associated with obesity in the Pima Indians are not associated in French European subjects. Moreover, different allele frequencies are observed for these two variants among European and Pima Indian backgrounds (the major allele in Pima is the minor allele in Europeans), the obesity risk

allele being more prevalent in the obesity-prone Pima Indian population.

Kublaoui et al. (17) examined the role of *Sim1* in hyperphagic obesity in mice and reported that *Sim1* and *Mc4r* are both expressed in the paraventricular nucleus, overexpression of *Sim1* in agouti yellow (*A^y*) mice inhibited hyperphagia and reduced fat mass in the *A^y/Sim1* transgenic mice (19). *Sim1*^{+/-} mice had higher levels of food intake than wild-type mice after treatment with melanotan II (an α -melanocyte-stimulating hormone analog and melanocortin 4 receptor agonist) (17), and oxytocin mRNA and protein levels were notably reduced in *Sim1*^{+/-} mice (39). Based on these findings, they proposed a simplified model for hyperphagic obesity that involves a signaling pathway from *Mc4r* to *Sim1* to oxytocin in the paraventricular nucleus neurons (17,19,39).

In conclusion, *SIM1* is emerging as a potential critical component of the MC4R signaling pathway known to regulate appetite in humans. The present study shows that common variation in this gene provides strong replicated associations with BMI in Pima Indians, where the common allele is the risk allele for obesity.

ACKNOWLEDGMENTS

This study was supported by the intramural research program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health. M.T. was supported by a grant from the American Diabetes Association.

No potential conflicts of interest relevant to this article were reported.

We gratefully acknowledge the volunteers from the Gila River Indian Community, whose cooperation made these studies possible.

REFERENCES

1. Knowler WC, Pettitt DJ, Saad MF, Charles MA, Nelson RG, Howard BV, Bogardus C, Bennett PH. Obesity in the Pima Indians: its magnitude and relationship with diabetes. *Am J Clin Nutr* 1991;53:S1543-S1551
2. Thomas JB, Crews ST, Goodman CS. Molecular genetics of the single-minded locus: a gene involved in the development of the *Drosophila* nervous system. *Cell* 1988;52:133-141
3. Nambu JR, Franks RG, Hu S, Crews ST. The single-minded gene of *Drosophila* is required for the expression of genes important for the development of CNS midline cells. *Cell* 1990;63:63-75
4. Crews S, Franks R, Hu S, Matthews B, Nambu J. *Drosophila* single-minded gene and the molecular genetics of CNS midline development. *J Exp Zool* 1992;261:234-244
5. Michaud JL, Rosenquist T, May NR, Fan CM. Development of neuroendocrine lineages requires the bHLH-PAS transcription factor *SIM1*. *Genes Dev* 1998;12:3264-3275
6. Kalra SP, Dube MG, Pu S, Xu B, Horvath TL, Kalra PS. Interacting appetite-regulating pathways in the hypothalamic regulation of body weight. *Endocr Rev* 1999;20:68-100
7. Turleau C, Demay G, Cabanis MO, Lenoir G, de Grouchy J. 6q1 monosomy: a distinctive syndrome. *Clin Genet* 1988;34:38-42
8. Villa A, Urioste M, Bofarull JM, Martínez-Frías ML. De novo interstitial deletion q16.2q21 on chromosome 6. *Am J Med Genet* 1995;55:379-383
9. Holder JL Jr, Butte NF, Zinn AR. Profound obesity associated with a balanced translocation that disrupts the *SIM1* gene. *Hum Mol Genet* 2000;9:101-108
10. Gilhuis HJ, van Ravenswaaij CM, Hamel BJ, Gabreëls FJ. Interstitial 6q deletion with a Prader-Willi-like phenotype: a new case and review of the literature. *Eur J Paediatr Neurol* 2000;4:39-43
11. Faivre L, Cormier-Daire V, Lapiere JM, Colleaux L, Jacquemont S, Geneviève D, Saunier P, Munnich A, Turleau C, Romana S, Prieur M, De Blois MC, Vekemans M. Deletion of the *SIM1* gene (6q16.2) in a patient with a Prader-Willi-like phenotype. *J Med Genet* 2002;39:594-596
12. Varela MC, Simões-Sato AY, Kim CA, Bertola DR, De Castro CI, Koifmann CP. A new case of interstitial 6q16.2 deletion in a patient with Prader-Willi-

- like phenotype and investigation of SIM1 gene deletion in 87 patients with syndromic obesity. *Eur J Med Genet* 2006;49:298–305
13. Bonaglia MC, Ciccone R, Gimelli G, Gimelli S, Marelli S, Verheij J, Giorda R, Grasso R, Borgatti R, Pagone F, Rodriguez L, Martinez-Frias ML, van Ravenswaaij C, Zuffardi O. Detailed phenotype-genotype study in five patients with chromosome 6q16 deletion: narrowing the critical region for Prader-Willi-like phenotype. *Eur J Hum Genet* 2008;16:1443–1449
 14. Wang J, Turner L, Lomax B, Eydoux P. A 5-Mb microdeletion at 6q16.1-q16.3 with *SIM1* gene deletion and obesity. *Am J Med Genet A* 2008;146A:2975–2978
 15. Michaud JL, Boucher F, Melnyk A, Gauthier F, Goshu E, Lévy E, Mitchell GA, Himms-Hagen J, Fan CM. *Sim1* haploinsufficiency causes hyperphagia, obesity and reduction of the paraventricular nucleus of the hypothalamus. *Hum Mol Genet* 2001;10:1465–1473
 16. Holder JL Jr, Zhang L, Kublaoui BM, DiLeone RJ, Oz OK, Bair CH, Lee YH, Zinn AR. *Sim1* gene dosage modulates the homeostatic feeding response to increased dietary fat in mice. *Am J Physiol Endocrinol Metab* 2004;287:E105–E113
 17. Kublaoui BM, Holder JL Jr, Gemelli T, Zinn AR. *Sim1* haploinsufficiency impairs melanocortin-mediated anorexia and activation of paraventricular nucleus neurons. *Mol Endocrinol* 2006;20:2483–2492
 18. Yang C, Gagnon D, Vachon P, Tremblay A, Levy E, Massie B, Michaud JL. Adenoviral-mediated modulation of *Sim1* expression in the paraventricular nucleus affects food intake. *J Neurosci* 2006;26:7116–7120
 19. Kublaoui BM, Holder JL Jr, Tolson KP, Gemelli T, Zinn AR. *SIM1* overexpression partially rescues agouti yellow and diet-induced obesity by normalizing food intake. *Endocrinology* 2006;147:4542–4549
 20. Knowler WC, Bennett PH, Hamman RF, Miller M. Diabetes incidence and prevalence in Pima Indians: a 19-fold greater incidence than in Rochester, Minnesota. *Am J Epidemiol* 1978;108:497–505
 21. Maillard G, Charles MA, Lafay L, Thibault N, Vray M, Borys JM, Basdevant A, Eschwège E, Romon M. Macronutrient energy intake and adiposity in non obese prepubertal children aged 5–11 y (the Fleurbaix Laventie Ville Sante Study). *Int J Obes Relat Metab Disord* 2000;24:1608–1617
 22. Vu-Hong TA, Durand E, Deghmoun S, Boutin P, Meyre D, Chevenne D, Czernichow P, Froguel P, Levy-Marchal C. The *INS VNTR* locus does not associate with smallness for gestational age (SGA) but interacts with SGA to increase insulin resistance in young adults. *J Clin Endocrinol Metab* 2006;91:2437–2440
 23. Hercberg S, Preziosi P, Briançon S, Galan P, Triol I, Malvy D, Roussel AM, Favier A. A primary prevention trial using nutritional doses of antioxidant vitamins and minerals in cardiovascular diseases and cancers in a general population: the SU.VI.MAX study: design, methods, and participant characteristics: SUPPLEMENTATION EN VITAMINES ET MINÉRAUX ANTIOXYDANTS. *Control Clin Trials* 1998;19:336–351
 24. Hanis CL, Chakraborty R, Ferrell RE, Schull WJ. Individual admixture estimates: disease associations and individual risk of diabetes and gall-bladder disease in Mexican-Americans in Starr County, Texas. *Am J Phys Anthropol* 1986;70:433–441
 25. Tian C, Hinds DA, Shigeta R, Adler SG, Lee A, Pahl MV, Silva G, Belmont JW, Hanson RL, Knowler WC, Gregersen PK, Ballinger DG, Seldin MF. A genome-wide single nucleotide polymorphism panel for Mexican American admixture mapping. *Am J Hum Genet* 2007;80:1014–1023
 26. Abecasis GR, Cardon LR, Cookson WOC. A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 2000;66:279–292
 27. Petitti DB. Statistical methods in meta-analysis. In *Meta-Analysis, Decision Analysis and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine*. Petitti DB, Ed. Oxford, Oxford University Press, 2000, p. 94–118
 28. Hudson RR, Kaplan N. Statistical properties of the number of recombination events in the history of a sample of sequences. *Genetics* 1985;111:147–164
 29. Vozarova de Courten B, Hanson RL, Funahashi T, Lindsay RS, Matsuzawa Y, Tanaka S, Thameem F, Gruber JD, Froguel P, Wolford JK. Polymorphisms in the adiponectin gene *ACDC* are not associated with type 2 diabetes in Pima Indians. *Diabetes* 2005;54:284–289
 30. Lathrop GM, Lalouel JM. Easy calculations of lod scores and genetic risks on small computers. *Am J Hum Genet* 1980;36:460–465
 31. Meyre D, Lecoecur C, Delplanque J, Francke S, Vatin V, Durand E, Weill J, Dina C, Froguel P. A genome-wide scan for childhood obesity-associated traits in French families shows significant linkage on chromosome 6q22.31-q23.2. *Diabetes* 2004;53:803–811
 32. Hanson RL, Ehm MG, Pettitt DJ, Prochazka M, Thompson DB, Timberlake D, Foroud T, Kobes S, Baier L, Burns DK, Almasy L, Blangero J, Garvey WT, Bennett PH, Knowler WC. An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. *Am J Hum Genet* 1998;63:1130–1138
 33. Hung CC, Luan J, Sims M, Keogh JM, Hall C, Wareham NJ, O'Rahilly S, Farooqi IS. Studies of the *SIM1* gene in relation to human obesity and obesity-related traits. *Int J Obes (Lond)* 2007;31:429–434
 34. Ahituv N, Kavaslar N, Schackwitz W, Ustaszewska A, Martin J, Hebert S, Doelle H, Ersoy B, Kryukov G, Schmidt S, Yosef N, Ruppin E, Sharan R, Vaisse C, Sunyaev S, Dent R, Cohen J, McPherson R, Pennacchio LA. Medical sequencing at the extremes of human body mass. *Am J Hum Genet* 2007;80:779–791
 35. International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;437:1299–1320
 36. Dudbridge F, Gonsanto A. Estimation of significance thresholds in genomewide association studies. *Genet Epidemiol* 2008;32:227–234
 37. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, Knight B, Patch AM, Ness AR, Ebrahim S, Lawlor DA, Ring SM, Ben-Shlomo Y, Jarvelin MR, Sovio U, Bennett AJ, Melzer D, Ferrucci L, Loos RJ, Barroso I, Wareham NJ, Karpe F, Owen KR, Cardon LR, Walker M, Hitman GA, Palmer CN, Doney AS, Morris AD, Smith GD, Hattersley AT, McCarthy MI. A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 2007;316:889–894
 38. Rong R, Hanson RL, Ortiz D, Wiedrich C, Kobes S, Knowler WC, Bogardus C, Baier LJ. Association analysis of variation in/near *FTO*, *CDKAL1*, *SLC30A8*, *HHEX*, *EXT2*, *IGF2BP2*, *LOC387761* and *CDKN2B* with type 2 diabetes and related quantitative traits in Pima Indians. *Diabetes* 2009;58:478–488
 39. Kublaoui BM, Gemelli T, Tolson KP, Wang Y, Zinn AR. Oxytocin deficiency mediates hyperphagic obesity of *Sim1* haploinsufficient mice. *Mol Endocrinol* 2008;22:1723–1734