

Appetite regulation genes are associated with body mass index in black South African adolescents: a genetic association study

Zané Lombard,^{1,2} Nigel J Crowther,³ Lize van der Merwe,^{4,5} Punita Pitamber,¹ Shane A Norris,⁶ Michèle Ramsay¹

To cite: Lombard Z, Crowther NJ, van der Merwe L, *et al.* Appetite regulation genes are associated with body mass index in black South African adolescents: a genetic association study. *BMJ Open* 2012;**2**:e000873. doi:10.1136/bmjopen-2012-000873

► Prepublication history and additional materials for this paper are available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2012-000873>).

Received 13 January 2012
Accepted 20 April 2012

This final article is available for use under the terms of the Creative Commons Attribution Non-Commercial 2.0 Licence; see <http://bmjopen.bmj.com>

For numbered affiliations see end of article.

Correspondence to

Dr Zané Lombard;
zane.lombard@wits.ac.za

ABSTRACT

Background: Obesity is a complex trait with both environmental and genetic contributors. Genome-wide association studies have identified several variants that are robustly associated with obesity and body mass index (BMI), many of which are found within genes involved in appetite regulation. Currently, genetic association data for obesity are lacking in Africans—a single genome-wide association study and a few replication studies have been published in West Africa, but none have been performed in a South African population.

Objective: To assess the association of candidate loci with BMI in black South Africans. The authors focused on single nucleotide polymorphisms (SNPs) in the *FTO*, *LEP*, *LEPR*, *MC4R*, *NPY2R* and *POMC* genes.

Design: A genetic association study.

Participants: 990 randomly selected individuals from the larger Birth to Twenty cohort (a longitudinal birth cohort study of health and development in Africans).

Measures: The authors genotyped 44 SNPs within the six candidate genes that included known BMI-associated SNPs and tagSNPs based on linkage disequilibrium in an African population for *FTO*, *LEP* and *NPY2R*. To assess population substructure, the authors included 18 ancestry informative markers. Weight, height, sex, sex-specific pubertal stage and exact age collected during adolescence (13 years) were used to identify loci that predispose to obesity early in life.

Results: Sex, sex-specific pubertal stage and exact age together explain 14.3% of the variation in log(BMI) at age 13. After adjustment for these factors, four SNPs were individually significantly associated with BMI: *FTO* rs17817449 ($p=0.022$), *LEP* rs10954174 ($p=0.0004$), *LEP* rs6966536 ($p=0.012$) and *MC4R* rs17782313 ($p=0.045$). Together the four SNPs account for 2.1% of the variation in log(BMI). Each risk allele was associated with an estimated average increase of 2.5% in BMI.

Conclusions: The study highlighted SNPs in *FTO* and *MC4R* as potential genetic markers of obesity risk in South Africans. The association with two SNPs in the 3' untranslated region of the *LEP* gene is novel.

ARTICLE SUMMARY

Article focus

- This is a replication study aiming to reproduce BMI association findings from European cohorts in a South African population.
- This study focused on genes linked to appetite control that were previously reported to show association with BMI or obesity and included *FTO*, *LEP*, *LEPR*, *MC4R*, *NPY2R* and *POMC*.
- Adolescent data were used to facilitate the identification of genetic loci that predispose to obesity early in life, as it is known that overweight/obese children have an elevated risk of becoming obese adults.

Key messages

- We found four SNPs were individually significantly associated with BMI: *FTO* rs17817449 ($p=0.022$), *LEP* rs10954174 ($p=0.0004$), *LEP* rs6966536 ($p=0.012$) and *MC4R* rs17782313 ($p=0.045$).
- Together the four SNPs account for 2.1% of the variation in log(BMI).
- We also demonstrated that an accumulation of risk alleles is linked to a significant increase in BMI—individuals with seven risk alleles had an 11.0% increase in median BMI compared with those with two risk alleles.

Strengths and limitations of this study

- This study provides the first preliminary evidence of the role of genetic variants in obesity risk in an adolescent black South African population.
- This study was only moderately powered to detect association with BMI, and not all genes were exhaustively investigated.
- TagSNP selection would have been enhanced if South African data were available for this approach.

INTRODUCTION

Being overweight or obese is a significant risk factor for the development of chronic diseases like type II diabetes and cardiovascular disease and is becoming increasingly common in low- and middle-income countries.¹ Both environmental and genetic factors influence body mass, with the heritability of body mass index (BMI, kg/m²) estimated at 40%–70%.²

The study of syndromic obesity and genome-wide association studies (GWAS) of common measures of adiposity (eg, BMI) have provided evidence for genetic risk loci for obesity.^{3–4} Current GWAS have primarily focused on populations of European origin, and only one GWAS for anthropomorphic traits, including BMI, has been performed in an indigenous African population. This study was carried out in a cohort of 1188 Nigerian subjects and replicated the association of *MC4R* with BMI, but no other significant associations were detected.⁵ However, the study was underpowered to detect modest effects. Although several replication studies have been performed in African–Americans,^{6–7} only a few are reported in well-characterised African populations^{8–9} and showed some supporting evidence of the role of *FTO* in a Nigerian population but not in the Gambia. Genetic association studies in African populations have the significant advantage that linkage disequilibrium (LD) generally exists over a shorter genomic distance, potentially increasing the efficiency of the identification of causal variants.¹⁰

Several of the likely causal genes for increased body fat mass are involved in appetite regulation. Specifically, constituents of the leptin–melanocortin neurotransmission pathway have been implicated in both monogenic and polygenic obesity. Recent evidence that *FTO* (the only gene consistently associated with measures of obesity) acts on the central nervous system-mediated control of food intake¹¹ lends further support for the hypothesis that obesity may be a heritable neurobehavioral disorder.

Obesity levels are rising at an alarming rate among black South Africans, with 29% and 57% of men and women, respectively, being overweight or obese. The major impact of obesity on the development of chronic diseases within South Africa is demonstrated by data showing that in the year 2000, 87% of type II diabetes cases and 68% of hypertensive disease were attributable to a BMI of ≥ 21 .¹ The multifactorial nature of these conditions creates a challenge, and understanding the genetic contribution to obesity in black South Africans may influence the nature of effective interventions. The aim of this study was therefore to assess the association of genetic variants with BMI in black South African adolescents. We focused on genes linked to appetite control that were previously reported to show association with BMI or obesity: *FTO*, *LEP*, *LEPR*, *MC4R*, *NPY2R* and *POMC*.

METHODS

Participants

The investigation focused on participants of the South African Birth to Twenty (Bt20) cohort study. Bt20 is

a longitudinal study of the health and well-being of children who were born in the metropolitan area of Soweto, Johannesburg, during a 6-week period in 1990. Participants in Bt20 were enrolled at birth, and detailed information has been collected from the participants and their care givers over the past 20 years.¹² Following informed consent, blood samples were collected and DNA extracted using a salting-out method.¹³ The Bt20 research programme has received clearance by the Human Research Ethics Committee: (Medical) of the University of the Witwatersrand (M010556). The Federal-Wide Assurance registration number of the committee is FWA00000715.

A subset of individuals from the Bt20 cohort (524 female and 466 male adolescents, total $n=990$) were randomly selected for this study. This represents 43% of the total cohort (2290 participants). The following calculation illustrates that this study was adequately powered. Group sizes of 748, 204 and 21 (total=973), corresponding to rs10954174 genotype frequencies observed in this study, will achieve 93% power to detect differences among the means of BMI using an F test with a 0.05 significance level, as is standard in linear regression models. The size of the variation in the means (say 18, 19 and 20) is represented by their SD, which is 0.47. The common SD within a group is assumed to be 4.

Adolescent data (participants were 13 years of age at the time measurements were done) were used to facilitate the identification of genetic loci that predispose to obesity early in life, as it is known that overweight/obese children have an elevated risk of becoming obese adults.¹⁴ Anthropometric measurements were obtained using standard methods.¹⁵ BMI was computed as weight (measured in kilogram) divided by the square of the height (measured in metres) of an individual. Pubertal stage was assessed using a validated self-assessment method drawing upon the Tanner Sexual Development Tool,¹⁶ and exact age at time of measurement was recorded.

Population substructure

A set of 18 ancestry informative markers (AIMs) were included to assess population substructure within the study group, based on previously published data¹⁷ and local unpublished data.¹⁸ Population substructure was assessed computationally using STRUCTURE V.2.3.3 software.¹⁹ For comparative purposes, genotypes of the AIMs were obtained from HapMap (release 24) for three additional populations—the Yoruba from Nigeria (YRI), the Han Chinese population (CHB) and Utah residents with European ancestry (CEU). STRUCTURE implements the value K to distinguish parental populations, with K referring to the number of inferred population clusters. The true value of K may not always be known, but the smallest value of K that captures the majority of structure in the data should be obtained. Ten iterations at $K=1$ to $K=4$ with a burn-in of 30 000 followed by 100 000 iterations were implemented. The K value with the highest average posterior likelihood score was calculated and used to identify the best cluster assignment.

Single nucleotide polymorphism selection

Single nucleotide polymorphisms (SNPs) previously shown to be associated with BMI or obesity in *FTO*, *LEP*, *LEPR*, *MC4R*, *NPY2R* and *POMC* were identified from the literature. In addition, tagSNPs were selected for *LEP*, *NPY2R* and *POMC* to ensure that known common polymorphisms are either directly assayed or associated with a tagSNP in these genes. SNPs from the literature were not force included in tagSNP selection. The Tagger algorithm²⁰ was used to select tagSNPs in a multimarker approach at $r^2 > 0.80$ and with a minor allele frequency > 0.05 among publicly available African data (YRI) from the HapMap data set (release 24). The Illumina Assay Design Tool (Illumina, San Diego, California, USA) was used to assess the SNPs' compatibility for inclusion in a custom-design GoldenGate VeraCode Assay, and 62 SNPs were selected for genotyping (supplementary table 1).

Genotyping

Genotyping was performed using the Illumina GoldenGate VeraCode assay (Illumina). Quality control was performed on all raw genotype data according to the supplier's specifications using BeadStudio V.2.0 software (Illumina) (summarised in Tindall *et al.*²¹). Standard quality control filters were applied to the data²²: minor allele frequency > 0.01 , SNP missingness rate < 0.05 , individual missingness rate < 0.2 and Hardy–Weinberg equilibrium (HWE) $< 1 \times 10^{-4}$.

LD plots

LD was visualised using Haploview V.4.2,²³ implementing the CI method of Gabriel *et al.*²⁴ to construct haplotype blocks. The chromosomal location of SNPs was obtained from NCBI (genome build 37.1).

Statistical analysis

The programming environment R, and R package genetics were used for all statistics (<http://www.r-project.org>).

Individual genotype associations with BMI

Genotype and allele frequencies were calculated through direct counting. The distribution of BMI was skewed to the right, and therefore, BMI was log-transformed to approximate normality for all analyses. Linear models were used for all analyses (individual association, joint association, sex-specific association and risk score association), starting by comparing BMI between genotype groups, enabling adjustment for confounders. As BMI correlated significantly with sex, sex-specific pubertal stage and exact age, some analyses were adjusted for these variables by including them in the linear models as covariates. All association p values are from linear models of $\log(\text{BMI})$. Exact tests of HWE were also performed. Because of its skewed distribution in this study group, BMI is summarised as median values and IQRs, with lower quartile and upper quartiles reported.

Furthermore, each genotype was modelled additively as the number of minor alleles present. Because we modelled $\log(\text{BMI})$, the modelled effect of each allele is reported as a percentage change (increase or decrease) in BMI.

Joint model and allele risk score of all associated SNPs

We also created and summarised a linear model containing all the variants that were individually significantly associated with BMI. We used all significantly associated variants to calculate a risk allele score by counting the number of alleles associated with increasing BMI in each adolescent. This risk allele score was also assessed with a linear model, independent of the three confounders.

Sex-specific effects

In order to detect sex-specific effects, we modelled the interaction between sex and each variant and summarised the significant results.

Correcting for multiple testing

Correcting for multiple testing is a contentious issue, and some approaches (such as Bonferroni correction) are considered overconservative and one risks the rejection of true findings.^{25 26} Given the strong prior information about the role of the variation tested here in obesity, we considered this a replication study, and therefore, p values below 0.05 were considered significant. For tagSNPs, tests of associations could be considered discovery rather than replication, but since these markers are correlated due to LD, the Bonferroni assumption of independence is not upheld.

RESULTS

Participants

The study group consists of 524 (53.0%) female and 466 (47.0%) male adolescents, with a mean age of 13.7 years (SD=0.2 years). Female adolescents had a higher median BMI (19.4 (IQR=17.8–22.3)) than male adolescents (18.0 (IQR=16.6–19.7); $p < 0.0001$). BMI increased across the five pubertal stages in both male ($p < 0.0324$) and female adolescents ($p < 0.0001$), but at significantly different rates. Exact age (days) correlated positively with BMI in male adolescents ($p = 0.027$) but not in female adolescents ($p = 0.305$). In a joint model, sex, sex-specific pubertal stage and age each correlated independently with BMI and together explained 14.3% ($p < 0.001$) of the variation in $\log(\text{BMI})$. Pubertal effect when assessed alone had the greatest impact (10.7%) followed by sex (5.7%). As a result, all analyses were adjusted for age, sex and pubertal stage per sex. Unadjusted p values of individual tests are presented for comparison (table 1).

Population substructure

One AIM failed quality control measures and was subsequently excluded from analysis. Analysis of allele frequencies of AIMS in the South African cohort

Table 1 Candidate SNP allele and genotype frequencies in a black South African population, and p values for additive allelic association with log(BMI) in adolescents aged 13 years

Gene	dbSNP ID	n	RefSNP alleles*		Frequency				Allele				p Values	
					Genotype			Allele		Unadjusted	Adjusted†	HWE		
			A	B	AA	AB	BB	Minor	Major					
FTO	rs9939973	977	A	G	0.14	0.45	0.40	0.37	0.63	0.109	0.354	0.450		
	rs9940128	985	A	G	0.12	0.47	0.41	0.35	0.65	0.105	0.379	0.403		
	rs1421085	986	T	C	0.89	0.11	0.00	0.06	0.94	0.728	0.709	0.764		
	rs1121980	974	T	C	0.20	0.51	0.29	0.45	0.55	0.125	0.294	0.518		
	rs17817449	977	T	G	0.39	0.48	0.14	0.37	0.63	0.008	0.022	0.539		
	rs8050136	982	A	C	0.17	0.51	0.32	0.42	0.58	0.021	0.057	0.170		
	rs9939609	981	A	T	0.20	0.53	0.27	0.47	0.53	0.029	0.071	0.055		
	rs1349419	986	A	G	0.96	0.04	0.00	0.02	0.98	0.388	0.612	1.000		
	rs12535708	975	A	C	0.03	0.32	0.65	0.19	0.81	0.755	0.911	0.092		
	rs11770725	985	T	C	0.91	0.09	0.00	0.05	0.95	0.640	0.993	0.491		
LEP	rs12535747	977	A	C	0.01	0.18	0.81	0.10	0.90	0.258	0.339	0.724		
	rs2167270	964	A	G	0.24	0.51	0.25	0.50	0.50	0.557	0.255	0.562		
	rs2278815	980	A	G	0.00	0.08	0.92	0.04	0.96	0.918	0.887	1.000		
	rs2122627	984	T	C	0.01	0.12	0.88	0.06	0.94	0.197	0.102	0.282		
	rs4236625	985	A	T	0.78	0.21	0.02	0.12	0.88	0.804	0.706	0.366		
	rs12706832	979	A	G	0.80	0.18	0.01	0.10	0.90	0.787	0.908	0.607		
	rs10244329	963	A	T	0.41	0.45	0.14	0.37	0.63	0.034	0.161	0.447		
	rs7791621	984	A	C	0.86	0.13	0.01	0.07	0.93	0.761	0.792	1.000		
	rs7795794	987	A	G	0.01	0.12	0.87	0.07	0.93	0.185	0.084	0.308		
	rs10954174	973	A	G	0.02	0.21	0.77	0.13	0.87	0.003	<0.001	0.111		
LEPR	rs3828942	962	A	G	0.03	0.28	0.69	0.17	0.83	0.922	0.852	0.567		
	rs17151919	969	A	G	0.02	0.26	0.71	0.15	0.85	0.699	0.829	1.000		
	rs17151922	974	T	G	0.15	0.50	0.35	0.40	0.60	0.239	0.378	0.143		
	rs6966536	972	A	G	0.73	0.25	0.02	0.14	0.86	0.014	0.012	0.896		
	rs10954173	977	A	G	0.02	0.26	0.72	0.15	0.85	0.342	0.525	0.620		
	rs11761556	984	A	C	0.01	0.19	0.80	0.11	0.89	0.798	0.659	0.737		
	rs1137100	984	A	G	0.89	0.11	0.00	0.06	0.94	0.602	0.602	1.000		
	rs1137101	965	A	G	0.20	0.52	0.28	0.45	0.55	0.650	0.905	0.271		
	rs17782313	969	T	C	0.55	0.39	0.06	0.25	0.75	0.070	0.045	0.444		
	rs12970134	981	A	G	0.01	0.16	0.83	0.09	0.91	0.166	0.116	0.838		
MC4R														

Continued

Continued

Table 1 Continued

Gene	dbSNP ID	n	RefSNP alleles*	Frequency				Allele			p Values		
				Genotype				AA	AB	BB	Minor	Major	HWE
				A	B	A	B						
NPY2R	rs2880416	985	C	G	G	0.00	0.08	0.92	0.04	0.96	0.094	0.186	1.000
	rs2342676	953	A	G	G	0.08	0.41	0.51	0.29	0.71	0.697	0.942	0.937
	rs12649641	988	A	C	C	0.48	0.43	0.08	0.30	0.70	0.601	0.786	0.255
	rs11099992	964	A	G	G	0.58	0.36	0.06	0.24	0.76	0.085	0.073	1.000
	rs33977152	986	A	G	G	0.01	0.09	0.90	0.05	0.95	0.508	0.301	0.181
	rs12507396	982	A	T	T	0.95	0.05	0.00	0.02	0.98	0.858	0.957	0.417
	rs6857530	979	A	G	G	0.05	0.36	0.58	0.23	0.77	0.666	0.707	0.790
	rs10461238	971	C	G	G	0.32	0.49	0.19	0.43	0.57	0.640	0.430	0.948
	rs2342674	985	A	G	G	0.00	0.13	0.87	0.07	0.93	0.530	0.476	0.308
	rs1047214	977	T	C	C	0.88	0.12	0.01	0.06	0.94	0.926	0.919	0.581
POMC	rs2880415	982	A	G	G	0.55	0.40	0.05	0.25	0.75	0.458	0.292	0.061
	rs6713532	970	T	C	C	0.29	0.49	0.22	0.46	0.54	0.368	0.270	0.796
	rs7565877	972	A	G	G	0.37	0.49	0.14	0.39	0.61	0.944	0.967	0.742
	rs7565427	987	A	G	G	0.01	0.20	0.79	0.11	0.89	0.633	0.889	0.418

Significant p values are shown in bold.

*RefSNP allele according to NCBI dbSNP Build 134.

†p Values are shown adjusted for age, sex and sex-specific pubertal stage.

BMI, body mass index; SNP, single nucleotide polymorphism.

(supplementary table 2), and comparison to three other ethnic groups (supplementary table 3), confirmed that there was no significant population substructure within the study cohort (supplementary figure 1). The South African population and the YRI group showed distinctive clustering away from the Caucasian and Asian populations. Although the South African cohort and the YRI had similar clustering patterns, higher order clustering of all individuals included in this analysis (ie, higher values of K, where K indicates the number of parental populations) was shown to highlight the genetic distinction between these two African groups.

LD plot

To illustrate the unique patterns of LD in the South African population, an LD plot of the gene most significantly associated with BMI (*LEP*) was constructed (supplementary figure 2). Evidence of three LD blocks covering the gene is observed, with the 3' end of the gene in particular exhibiting high LD.

Statistical analysis

Individual genotype associations with BMI

The genotype and allele frequencies of all candidate SNPs, as well as p values for tests of additive allelic association with log(BMI) are shown in table 1. All SNPs were in HWE (table 1). Table 2 presents the median and IQR of BMI for SNPs that were significantly associated with BMI following adjustment for sex, pubertal stage and age. Estimated effect sizes (percentage difference in BMI associated with each additional minor allele compared with the major allele homozygote) adjusted for age, sex and pubertal stage are also presented. For two SNPs, *LEP* rs6966536 and *FTO* rs17817449, the minor allele (G in both cases) is associated with increased BMI, whereas the major allele is associated with increased BMI for the other two SNPs, *LEP* rs10954174 and *MC4R* rs17782313 (G and T, respectively). When referring to 'risk' allele, it is the allele that is associated with increased BMI.

For *LEP* rs6966536, each G-allele was associated with a 2.9% higher BMI (p=0.012), whereas for *LEP* rs10954174, each A-allele is associated with an estimated decrease of 4.0% in BMI (p<0.001). Thus, for an adolescent of the same age, sex and pubertal stage, we would expect an individual with the A/G genotype at

rs10954174 to have a BMI that is 4.0% less than that of a G/G homozygote; and the A/A homozygote is expected to have an 8.0% lower BMI compared with the G/G homozygote. Adolescents who are A/G heterozygotes for *LEP* rs6966536 are expected to have a 2.9% higher BMI than A/A homozygotes, while the G/G homozygotes would have a 5.8% higher BMI.

Sex-specific effects

Two SNPs—*MC4R* rs12970134 and *NPY2R* rs11099992—showed significant sex differences in their effects on BMI (results not shown). In female adolescents, each G-allele of the *NPY2R* rs11099992 SNP predicted a decrease in BMI by 3.7% (p=0.003), while in male adolescents, the effect was not significant (p=0.523). In male adolescents, each A-allele of the *MC4R* rs12970134 SNP predicted an increase in BMI of 5.2% (p=0.011), while in female adolescents, the effect was not significant (p=0.785). These SNPs were not associated with log(BMI) in the group as a whole (table 1).

Joint model

To assess the combined effect of the four SNPs that were individually associated with log(BMI) and the three confounders (sex, pubertal stage and age), a model was built based on 908 participants who had complete information. The model explained 16.4% of the variation in log(BMI), so that the four variants contributed 2.1% to the variation, independent of the three confounders (table 3). The two *LEP* SNPs had the largest effects, followed by the *FTO* and then the *MC4R* SNPs, which each had similar effect sizes (2.2 and -2.0, respectively). The seven *FTO* SNPs in this study explain 0.6% of variation in log(BMI) after adjusting for age-, gender- and sex-specific pubertal stage and 1.4% unadjusted.

Allelic risk score

Figure 1 shows the number of adolescents with each number of additive risk alleles (the allele associated with higher BMI—four loci and therefore a maximum of eight risk alleles in an individual) as well as the median and IQR of BMI for each group. Since there were <10 individuals each in the categories of one risk allele (seven individuals) and eight risk alleles (one

Table 2 Summary BMI statistics for SNPs that were individually associated with BMI									
Gene	SNP	Typed	Modelled	MA	MAF	Median (IQR) of BMI per genotype			Effect
						mm*	mM*	MM*	
<i>FTO</i>	rs17817449	977	952	G	0.37	18.5 (16.9–20.7)	18.7 (17.2–21.1)	19.1 (17.6–22.5)	1.9
<i>LEP</i>	rs10954174	973	948	A	0.13	18.8 (17.1–21.3)	18.4 (17.2–20.5)	17.4 (16.8–19.1)	-4.0
<i>LEP</i>	rs6966536	972	948	G	0.14	18.6 (17.1–20.8)	18.9 (17.1–21.5)	19.2 (18.5–21.6)	2.9
<i>MC4R</i>	rs17782313	969	945	C	0.25	18.9 (17.4–21.2)	18.5 (17.0–20.9)	19.0 (16.5–20.6)	-1.8

Effect sizes are from individual models, estimated individual percentage change in BMI for each minor allele, independent of age, sex and sex-specific pubertal stage. Effect sizes correspond to p values (adjusted) in table 1.
*m denotes major allele; M is minor allele.
BMI, body mass index; MA, minor allele; MAF, minor allele frequency; SNP, single nucleotide polymorphism.

Table 3 Joint model of SNPs individually associated with log(BMI)

Gene	SNP	MA	Effect	p Value
<i>FTO</i>	rs17817449	G	2.2	0.007
<i>LEP</i>	rs10954174	A	-3.2	0.005
<i>LEP</i>	rs6966536	G	2.7	0.020
<i>MC4R</i>	rs17782313	C	-2.0	0.028

The model explains 16.4% of the covariation in log(BMI). p Values are from the joint model, so they are adjusted for age, sex and sex-specific pubertal stage and all other SNPs in the model. Effect = estimated % change in BMI of each minor allele, independent of other confounding variables (age, sex and sex-specific pubertal stage) and all other SNPs in the model (908 adolescents included in model). BMI, body mass index; MA, minor allele; SNP, single nucleotide polymorphism.

individual), these were excluded from the figure. The estimated increase in BMI, for each risk allele, is 2.5%, independent of age, sex and sex-specific pubertal stage, and individuals with seven risk alleles had an 11.0% increase ((20.0-18.0)/18.0) in median BMI compared with a group of individuals with two risk alleles.

DISCUSSION

This study provides preliminary evidence of the role of genetic variants in obesity risk in an adolescent black South African group. Individual SNPs in *FTO*, *LEP* and *MC4R* genes were shown to be associated with an increase in BMI, and the cumulative effect of the risk alleles (one to eight alleles) in *FTO*, *LEP* and *MC4R* was demonstrated. Together they contribute 2.1% to the variation observed in log(BMI) in this group of subjects, after adjusting for age, sex and sex-specific pubertal stage. Recent GWAS in populations of European ancestry have identified 32 common loci associated with BMI, explaining only 1.5% of the interindividual variation in BMI.⁴ We also demonstrated that an accumulation of risk alleles is linked to a significant increase in BMI (figure 1). In our study, individuals with seven risk

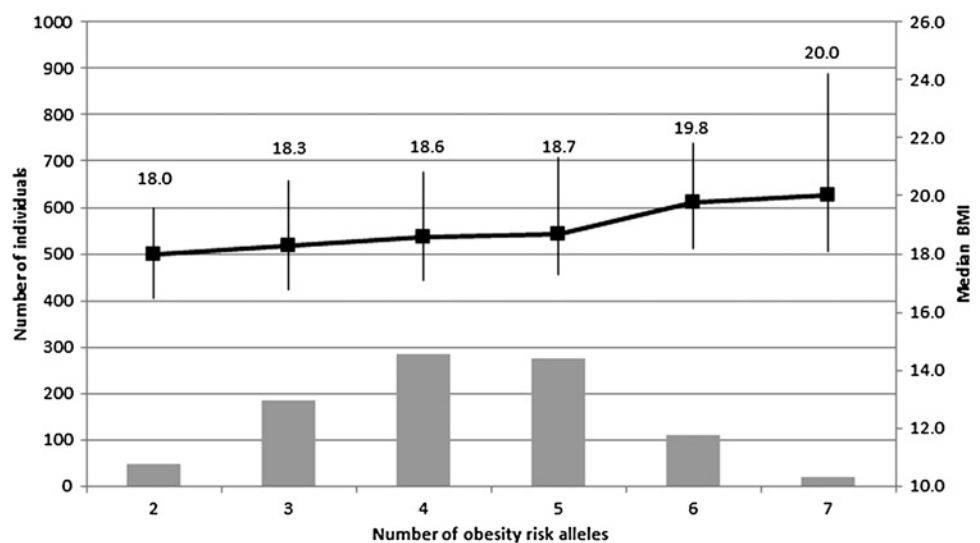
alleles had an 11.0% increase in median BMI compared with individuals with two risk alleles.

We provide preliminary evidence that SNPs in the 3' untranslated region (3'UTR) of *LEP* (rs10954174, p=0.0004 and rs6966536, p=0.012) are associated with BMI in South Africans. Leptin is an integral molecule in energy homeostasis, and circulating leptin is primarily produced by adipocytes. The level of circulating leptin acts as a satiety signal. Coding or splice site mutations in the *LEP* gene render the leptin protein unable to signal through its receptor, which leads to severe childhood-onset obesity.²⁷⁻²⁸ These mutations are rare in the general population, and their impact on the common obesity phenotype remains unclear.

The *LEP* SNPs that have been most consistently associated with obesity are rs2167270 (also known as *LEP* 19G>A) in the untranslated first exon,²⁹⁻³¹ and rs7799039 (*LEP* -2548G>A) in the 5'-region of the gene.³²⁻³⁵ Studies in a Brazilian population³⁴⁻³⁶ and adult Samoans³⁷ have also shown obesity to be correlated with variants of the *LEP* 3' hypervariable region (HVR). The 3' HVR is a (TTTC)_n tetranucleotide repeat located approximately 2000 bp downstream from *LEP*. The 3'UTR SNPs associated with BMI in this study do not share a haplotype block with the 3' HVR in Africans and could be a manifestation of a uniquely African association to BMI and obesity. There are several plausible mechanisms by which a 3'UTR variant could impact on the functioning of a gene. They could reside in regions of miRNA binding sites or may affect gene expression in other ways as 3'UTRs have been shown to be involved in posttranscriptional regulation of gene expression through the modulation of nucleocytoplasmic mRNA transport, translation efficiency, subcellular localisation and message stability.³⁸

Variants in *FTO* (the only gene consistently associated with obesity) have been associated with adiposity measures in Europeans,⁴⁻³⁹ Asians⁴⁰⁻⁴¹ and African-Americans,⁷ with the strongest effect seen with SNP rs9939609. Our study showed a modest association

Figure 1 Observed cumulative effect of high-risk alleles on body mass index (BMI) in a black South African adolescent cohort. The four risk alleles included in this analysis are: *LEP* rs6966536 (G-allele), *LEP* rs10954174 (G-allele), *MC4R* rs17782313 (G-allele) and *FTO* rs17817449 (T-allele). Bar chart (left axis) shows number of adolescents with specific number of obesity risk alleles. Points show corresponding median (right axis) and vertical lines represent the IQR of BMI.



between the *FTO* SNP, rs17817449, and BMI after adjustment for sex, age and sex-specific pubertal stage. An estimated effect size of a 1.9% increase in BMI for each minor allele of rs17817449 was observed. Inconsistent associations between *FTO* variants and obesity measures have been observed in native African populations. In Gambians, no associations between *FTO* and weight-for-height z scores were observed.⁹ rs17817449 was directly assayed in that study but showed no association. It should be noted that the Gambia has a very low obesity rate, which could possibly explain the null result. In West Africans, rs17817449 was not associated with BMI, but several other *FTO* variants were found to be associated.⁸ The role of genetic variation at the *FTO* locus in predisposing to obesity in African populations warrants further investigation, specifically in relation to the epidemiological transition and access to a calorie-rich diet.

The rs17782313 SNP is one of several variants in *MC4R* linked to BMI in European populations and has been shown to contribute to a significant increase in extreme obesity in adults.⁴² In the present study, this variant showed a borderline association with BMI ($p=0.045$), suggesting that this SNP is not the functional variant. It is located 3' to the gene, and it is therefore possible that it may be in LD with another polymorphism that disrupts the function of cis-acting elements involved in the control of *MC4R* gene transcription.⁴³

In this study, the SNP rs11099992 in *NPY2R* was observed to associate with BMI in female subjects only ($p=0.003$), whereas, *MC4R* rs12970134 showed a sex-specific effect in male subjects ($p=0.011$). It has been observed that SNPs in the *NPY2R* gene have effects on obesity in male subjects only.^{44–46} One of these studies also showed that a SNP in *PYY* (encoding a ligand of *NPY2R*) associates with obesity-related traits in female subjects only.⁴⁵ Furthermore, mutations in the *MC4R* gene that have been associated with morbid obesity have stronger effects in female subjects than male subjects.⁴⁷ These data suggest that the phenotypic effect of sequence variation in the *MC4R* gene is differentially modified by sex depending on the site of the variant. Thus, genetic variants relevant to appetite-regulatory pathways seem to have sex-specific effects on body composition, but the mechanism by which this occurs is still poorly understood.

The data in this study are derived from a cohort of adolescents in the midst of puberty. It is therefore possible that the effects on weight of some polymorphisms may have been masked by puberty-associated changes in body fat mass.^{48–50} It is also possible that the effects of some polymorphisms on BMI are observed only later in life.⁵¹ However, it has been shown that the *FTO* and *MC4R* polymorphisms do influence anthropometry in adults and children and that other polymorphisms identified in GWAS in adults also have observable effects on childhood measures of adiposity.^{42 52–54}

Replication has become the gold standard for assessing the validity of statistically significant results

from association studies. However, a true result can fail to replicate in another independent sample for numerous reasons including differences in allele frequencies and haplotype structure that could influence the indirect detection of functional polymorphisms. A recent study showed that the power to replicate the statistically significant independent main effect of one polymorphism can drop dramatically with a change of allele frequency of <0.1 at a second interacting polymorphism. Differences in allele frequency could also result in a reversal of allelic effects where a putative protective allele becomes associated with increased risk in a replication study.⁵⁵ These are particularly important considerations when undertaking genomic research and replication studies in African populations, with known high levels of genome diversity and population substructure between different African groups.^{56 57} The AIM data obtained in this study highlight that there are not high levels of heterogeneity within the South African group, but that the group differs from other African tribes, such as the Yoruba. Despite these differences, 18 AIMS cannot resolve the detailed population differences between this South African population and other African groups, as the Bantu expansion occurred rather recently in history (~ 4000 years ago), and one would expect that only major differences would be observable when interrogating a small number of markers.⁵⁷

Further studies on other candidate obesity loci in African populations will provide a better understanding of the role of variants in these genes in a population with a different LD structure to that of Caucasians.⁵⁸ In addition, a thorough examination of other measures of obesity is needed to better understand the complexities of obesity risk across different populations.³ Such studies will be enhanced by more detailed phenotypic characterisation of study cohorts that includes body fat distribution, as this is known to be a significant risk factor for other chronic diseases, such as cardiovascular disease and type II diabetes.⁵⁹

Author affiliations

¹Division of Human Genetics, School of Pathology, Faculty of Health Sciences, National Health Laboratory Service and University of the Witwatersrand, Johannesburg, South Africa

²Wits Bioinformatics, University of the Witwatersrand, Johannesburg, South Africa

³Department of Chemical Pathology, School of Pathology, Faculty of Health Sciences, National Health Laboratory Service and University of the Witwatersrand, Johannesburg, South Africa

⁴Biostatistics Unit, Medical Research Council, Cape Town, South Africa

⁵Department of Statistics, University of Western Cape, Cape Town, South Africa

⁶MRC/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

Acknowledgements The authors thank the participants in the Birth to Twenty study and acknowledge funding from the Wellcome Trust, the South African MRC, the South African National Research Foundation and University of the Witwatersrand for the collection and genotyping of samples.

Contributors SAN, MR, ZL and NJC conceived the idea for the study. SAN provided access to all phenotypic data used in this study. ZL and PP performed the SNP selection and laboratory work. PP completed the STRUCTURE analysis, ZL the Haploview plots and NJC the allelic risk score computation. LvdM performed all other statistical analyses. ZL wrote the first draft of the manuscript, and all authors contributed to, read and approved the final manuscript.

Funding This work was supported by the Wellcome Trust (grant number 080535/Z/06/Z to SAN); the South African Medical Research Council and the University of the Witwatersrand for the collection of the samples for the Bt20 cohort and genotyping. In addition, genotyping was funded by the South African National Research Foundation (grant number IFR2011062100004 to MR).

Competing interests None.

Ethics approval Ethics approval was provided by Human Research Ethics Committee: (Medical) of the University of the Witwatersrand.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

- Joubert J, Norman R, Bradshaw D, *et al.* Estimating the burden of disease attributable to excess body weight in South Africa in 2000. *S Afr Med J* 2007;97:683–90.
- Bodurtha JN, Mosteller M, Hewitt JK, *et al.* Genetic analysis of anthropometric measures in 11-year-old twins: the Medical College of Virginia Twin Study. *Pediatr Res* 1990;28:1–4.
- Kilpelainen TO, Zillikens MC, Stancakova A, *et al.* Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nat Genet* 2011;43:753–60.
- Speliotes EK, Willer CJ, Berndt SI, *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010;42:937–48.
- Kang SJ, Chiang CW, Palmer CD, *et al.* Genome-wide association of anthropometric traits in African- and African-derived populations. *Hum Mol Genet* 2010;19:2725–38.
- Bressler J, Kao WH, Pankow JS, *et al.* Risk of type 2 diabetes and obesity is differentially associated with variation in FTO in whites and African-Americans in the ARIC study. *PLoS One* 2010;5:e10521.
- Hassanein MT, Lyon HN, Nguyen TT, *et al.* Fine mapping of the association with obesity at the FTO locus in African-derived populations. *Hum Mol Genet* 2010;19:2907–16.
- Adeyemo A, Chen G, Zhou J, *et al.* FTO genetic variation and association with obesity in West Africans and African Americans. *Diabetes* 2010;59:1549–54.
- Hennig BJ, Fulford AJ, Sirugo G, *et al.* FTO gene variation and measures of body mass in an African population. *BMC Med Genet* 2009;10:21.
- Teo YY, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. *Nat Rev Genet* 2010;11:149–60.
- Church C, Moir L, McMurray F, *et al.* Overexpression of Fto leads to increased food intake and results in obesity. *Nat Genet* 2010;42:1086–92.
- Richter L, Norris S, Pettifor J, *et al.* Cohort Profile: Mandela's children: the 1990 Birth to Twenty study in South Africa. *Int J Epidemiol* 2007;36:504–11.
- Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988;16:1215.
- Stettler N, Iotova V. Early growth patterns and long-term obesity risk. *Curr Opin Clin Nutr Metab Care* 2010;13:294–9.
- Cameron N, De Wet T, Ellison GT, *et al.* Growth in height and weight of South African urban infants from birth to five years: the birth to ten study. *Am J Hum Biol* 1998;10:495–504.
- Norris SA, Richter LM. Validation of Tanner pubertal self-rating to urban black adolescents in South Africa. *J Res Adolesc* 2005;15:609–24.
- Lao O, van Duijn K, Kersbergen P, *et al.* Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. *Am J Hum Genet* 2006;78:680–90.
- Schlebusch CM. *Genetic Variation in Khoisan-speaking Populations from Southern Africa. Unpublished PhD thesis.* Johannesburg, South Africa: University of the Witwatersrand, 2011.
- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;155:945–59.
- de Bakker PI, Yelensky R, Pe'er I, *et al.* Efficiency and power in genetic association studies. *Nat Genet* 2005;37:1217–23.
- Tindall EA, Petersen DC, Nikolayzen S, *et al.* Interpretation of custom designed Illumina genotype cluster plots for targeted association studies and next-generation sequence validation. *BMC Res Notes* 2010;3:39.
- Clarke GM, Anderson CA, Pettersson FH, *et al.* Basic statistical analysis in genetic case-control studies. *Nat Protoc* 2011;6:121–33.
- Barrett JC, Fry B, Maller J, *et al.* Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–5.
- Gabriel SB, Schaffner SF, Nguyen H, *et al.* The structure of haplotype blocks in the human genome. *Science* 2002;296:2225–9.
- Perneer TV. What's wrong with Bonferroni adjustments. *BMJ* 1998;316:1236–8.
- Elsäßer A, Victor A, Hommel G. Multiple testing in candidate gene Situations: a comparison of classical, discrete, and resampling-based Procedures. *Stat Appl Genet Mol Biol* 2011;10:e51.
- Farooqi IS. Genetic aspects of severe childhood obesity. *Pediatr Endocrinol Rev* 2006;3(Suppl 4):528–36.
- Montague CT, Farooqi IS, Whitehead JP, *et al.* Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature* 1997;387:903–8.
- Mattevi VS, Zembrzinski VM, Hutz MH. Association analysis of genes involved in the leptin-signaling pathway with obesity in Brazil. *Int J Obes Relat Metab Disord* 2002;26:1179–85.
- Karvonen MK, Pesonen U, Heinonen P, *et al.* Identification of new sequence variants in the leptin gene. *J Clin Endocrinol Metab* 1998;83:3239–42.
- Lucantoni R, Ponti E, Berselli ME, *et al.* The A19G polymorphism in the 5' untranslated region of the human obese gene does not affect leptin levels in severely obese patients. *J Clin Endocrinol Metab* 2000;85:3589–91.
- Constantin A, Costache G, Sima AV, *et al.* Leptin G-2548A and leptin receptor Q223R gene polymorphisms are not associated with obesity in Romanian subjects. *Biochem Biophys Res Commun* 2010;391:282–6.
- Portoles O, Sorli JV, Frances F, *et al.* Effect of genetic variation in the leptin gene promoter and the leptin receptor gene on obesity risk in a population-based case-control study in Spain. *Eur J Epidemiol* 2006;21:605–12.
- Hinuy HM, Hirata MH, Sampaio MF, *et al.* Relationship between variants of the leptin gene and obesity and metabolic biomarkers in Brazilian individuals. *Arq Bras Endocrinol Metabol* 2010;54:282–8.
- Wang TN, Huang MC, Chang WT, *et al.* G-2548A polymorphism of the leptin gene is correlated with extreme obesity in Taiwanese aborigines. *Obesity (Silver Spring)* 2006;14:183–7.
- Hinuy HM, Hirata MH, Sampaio MF, *et al.* LEP 3'HVR is associated with obesity and leptin levels in Brazilian individuals. *Mol Genet Metab* 2006;89:374–80.
- McGarvey ST, Forrest W, Weeks DE, *et al.* Human leptin locus (LEP) alleles and BMI in Samoans. *Int J Obes Relat Metab Disord* 2002;26:783–8.
- Mishra PJ, Banerjee D, Bertino JR. MiRNAs or MiR-polymorphisms, new players in microRNA mediated regulation of the cell: Introducing microRNA pharmacogenomics. *Cell Cycle* 2008;7:853–8.
- Wang K, Li WD, Zhang CK, *et al.* A genome-wide association study on obesity and obesity-related traits. *PLoS One* 2011;6:e18939.
- Liu Y, Liu Z, Song Y, *et al.* Meta-analysis added power to identify variants in FTO associated with type 2 diabetes and obesity in the Asian population. *Obesity (Silver Spring)* 2010;18:1619–24.
- Dorajoo R, Blakemore AI, Sim X, *et al.* Replication of 13 obesity loci among Singaporean Chinese, Malay and Asian-Indian populations. *Int J Obes (Lond)* 2012;36:159–63.
- Loos RJ, Lindgren CM, Li S, *et al.* Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* 2008;40:768–75.
- Chen JM, Ferenc C, Cooper DN. A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes I: general principles and overview. *Hum Genet* 2006;120:1–21.
- Ma L, Tataranni PA, Hanson RL, *et al.* Variations in peptide YY and Y2 receptor genes are associated with severe obesity in Pima Indian men. *Diabetes* 2005;54:1598–602.
- Friedlander Y, Li G, Fornage M, *et al.* Candidate molecular pathway genes related to appetite regulatory neural network, adipocyte homeostasis and obesity: results from the CARDIA Study. *Ann Hum Genet* 2010;74:387–98.
- Hung CC, Pirie F, Luan J, *et al.* Studies of the peptide YY and neuropeptide Y2 receptor genes in relation to human obesity and obesity-related traits. *Diabetes* 2004;53:2461–6.

47. Dempfle A, Hinney A, Heinzel-Gutenbrunner M, *et al.* Large quantitative effect of melanocortin-4 receptor gene mutations on body mass index. *J Med Genet* 2004;41:795–800.
48. Comings DE, Gade R, Muhleman D, *et al.* The LEP gene and age of menarche: maternal age as a potential cause of hidden stratification in association studies. *Mol Genet Metab* 2001;73:204–10.
49. Mumby HS, Elks CE, Li S, *et al.* Mendelian randomisation study of childhood BMI and early menarche. *J Obes* 2011;2011:180729.
50. Elks CE, Perry JR, Sulem P, *et al.* Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat Genet* 2010;42:1077–85.
51. Mei H, Chen W, Jiang F, *et al.* Longitudinal replication studies of GWAS risk SNPs influencing body mass index over the Course of childhood and Adulthood. *PLoS One* 2012;7:e31470.
52. Dina C, Meyre D, Gallina S, *et al.* Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet* 2007;39:724–6.
53. Elks CE, Loos RJ, Sharp SJ, *et al.* Genetic markers of adult obesity risk are associated with greater early infancy weight gain and growth. *PLoS Med* 2010;7:e1000284.
54. Bollepalli S, Dolan LM, Deka R, *et al.* Association of FTO gene variants with adiposity in African-American adolescents. *Obesity* 2010;18:1959–63.
55. Greene CS, Penrod NM, Williams SM, *et al.* Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS One* 2009;4:e5639.
56. Bryc K, Auton A, Nelson MR, *et al.* Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc Natl Acad Sci U S A* 2010;107:786–91.
57. Tishkoff SA, Reed FA, Friedlaender FR, *et al.* The genetic structure and history of Africans and African Americans. *Science* 2009;324:1035–44.
58. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet* 2008;9:403–33.
59. Kissebah AH, Krakower GR. Regional adiposity and morbidity. *Physiol Rev* 1994;74:761–811.