BMJ open

Candidate Genes Related to Appetite Regulation are Associated with Body-Mass Index in Black South African Adolescents

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-000873
Article Type:	Research
Date Submitted by the Author:	13-Jan-2012
Complete List of Authors:	Lombard, Zané; National Health Laboratory Service & University of the Witwatersrand, Division of Human Genetics Crowther, Nigel; National Health Laboratory Service & University of the Witwatersrand, Chemical Pathology van der Merwe, Lize; Medical Research Council of South Africa, Biostatistics Unit Pitamber, Punita; National Health Laboratory Service & University of the Witwatersrand, Division of Human Genetics Norris, Shane; University of the Witwatersrand, MRC/Wits Developmental Pathways for Health Research Unit Ramsay, Michele; National Health Laboratory Service & University of the Witwatersrand, Division of Human Genetics
Primary Subject Heading:	Dentistry and oral medicine
Secondary Subject Heading:	Anaesthesia
Keywords:	Cell biology < BASIC SCIENCES, Immunology < BASIC SCIENCES, Microbiology < BASIC SCIENCES

Candidate Genes Related to Appetite Regulation are Associated with Body-Mass Index in Black South African Adolescents

Zané Lombard[1*], Nigel J Crowther[2], Lize van der Merwe[3], Punita Pitamber[1], Shane A Norris[4], Michèle Ramsay[1]

[1] Division Human Genetics, School of Pathology, Faculty of Health Sciences, National Health Laboratory Service & University of the Witwatersrand, Johannesburg, 2000, South Africa

[2] Department of Chemical Pathology, School of Pathology, Faculty of Health Sciences, National Health Laboratory Service & University of the Witwatersrand, Johannesburg, 2000, South Africa

[3] Biostatistics Unit, Medical Research Council, Cape Town, 8001, South Africa and Department of Statistics, University of Western Cape, Bellville, South Africa

[4] MRC/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, 2000, South Africa

***Corresponding author:** Zané Lombard; Postal address: PO Box 1038, Johannesburg, 2000, South Africa; Email: zane.lombard@wits.ac.za; Telephone: +27114899344; Fax: +27114899226

Key words: Obesity, Body Mass Index, Genetic Association Studies, *FTO*, *LEP*, *MC4R*

Word count: 3984 words

ABSTRACT

Background: Genome-wide association studies have provided evidence for common obesity risk loci, but few studies have been performed on African populations. Several identified candidate genes for increased body mass index (BMI) and associated phenotypes are involved in appetite regulation.

Objective: To assess the association of genetic variants with BMI in black South Africans and focused on SNPs in the *FTO*, *LEP*, *LEPR*, *MC4R*, *NPY2R* and *POMC* genes.

Methods: We selected 44 SNPs previously associated with BMI (and including tagSNPs), as well as 18 ancestry informative markers (to assess genetic substructure), as the focus of this analysis. Genotyping was performed in 990 participants from the Birth to Twenty cohort (a longitudinal study of health and development in Africans), focusing on data collected during adolescence (13 years of age) to aid the identification of loci that predispose to obesity early in life.

Results: Gender, gender-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); *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 increase of 2.5% in BMI.

Conclusion: The association with the two SNPs in the 3' UTR of the *LEP* gene is novel. Future studies in Africans will further highlight the complexities of obesity risk across different populations and identify potential targets for therapeutic intervention.

INTRODUCTION

Being overweight or obese is a significant risk factor for the development of chronic diseases, and is becoming increasingly common in developing countries [1]. Both environmental and genetic factors influence body mass, with the heritability of BMI estimated at 40-70% [2].

The study of syndromic obesity, as well as genome-wide association studies (GWAS) of common measures of adiposity (e.g. BMI) has provided evidence for genetic risk loci for obesity [3 4]. Current GWAS have primarily focused on populations of European origin, and only one GWAS of obesity-associated SNPs 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]. Although replication studies of the GWAS findings for obesity risk have been performed in African Americans [6 7] only a few are reported in well characterized African populations [8 9]. Genetic association studies in African populations have the significant advantage that linkage disequilibrium (LD) generally exists over a shorter genomic distance, 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. The recent evidence that *FTO* (the only gene consistently associated with measures of obesity) acts on the central nervous system–mediated control of food intake [11] lends further support for the hypothesis that obesity may be a heritable neurobehavioral disorder.

The aim of this study was to assess the association of genetic variants with BMI in black South African adolescents. The investigation focused on participants of the South African Birth to Twenty (Bt20) cohort study. Bt20 is a longitudinal study of the health and wellbeing of children that were born in the metropolitan area of Soweto, Johannesburg, during a six-week period in 1990. Participants in Bt20 were enrolled at birth and detailed information has been collected from the participants and their caregivers over the past twenty years [12]. 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 [13]. 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

A subset of individuals from the Bt20 cohort (524 female and 466 male adolescents, total n=990) were randomly selected for this study. Weight, height, pubertal stage and exact age (days) at the time of assessment were measured at 13 years of age in participants. With this sample size, our study achieves 80% power to detect differences in BMI of $\beta \geq 0.67$ (BMI increase ≥0.67kg/m^2 per risk allele) at $\alpha = 0.05$ [14]. Following informed consent, blood samples were collected and DNA extracted using a salting-out method [15].

BMI distribution in study group

Anthropometric measurements (weight (kg) and height (m)) were obtained using standard methods [16]. BMI was computed as weight (measured in kg) divided by the square of the height (measured in meters) of an individual. Pubertal stage was assessed using a validated self-assessment method drawing upon the Tanner sexual development tool [17]. The Bt20 research programme, has received clearance by the Human Research Ethics Committee: (Medical) of the University of the Witwatersrand (M010556). The Federal Wise Assurance registration number of the Committee is FWA00000715.

Population substructure

A set of 18 ancestry informative markers (AIMs) were included to assess population substructure within the study group, using STRUCTURE V.2.3.3 software [18]. For comparative purposes, genotype data for the 18 AIMs were obtained from HapMap 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.

SNP selection

SNPs previously shown to be associated with BMI or obesity within the selected genes were identified following a literature review. In addition, tagSNPs were selected for *LEP*, *NPY2R* and *POMC*, to ensure that all known common polymorphisms are either directly assayed or associated with a tagSNP. The Tagger algorithm [19] was used to select tagSNPs in a multimarker approach at $r^2 > 0.80$ and with a minor allele frequency (MAF) > 0.05 among publicly available African data (YRI) from the HapMap dataset. The Illumina Assay Design Tool (Illumina, San Diego, CA) 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, San Diego, CA, USA). Quality control was performed on all raw genotype data according to the supplier's specifications using BeadStudio v2.0 software (Illumina, San Diego, CA) (summarized in [20]).

Statistical Analysis

Genotype and allele frequencies were calculated and Hardy–Weinberg equilibrium (HWE) was assessed. 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, starting by comparing BMI between genotype groups, enabling adjustment for confounders. As BMI correlated significantly with gender, gender-specific pubertal stage and exact age, all analyses were adjusted for these variables by including them in the linear models as covariates. Each genotype was modeled as the number of minor alleles. The modeled effect of each allele is converted to a percentage change (increase or decrease) in BMI. We also created and summarised a linear model containing all the variants that were individually significantly associated with BMI. In order to detect gender-specific effects, we modeled the interaction between gender and each variant and summarised the significant results. We then used these 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.

Results corresponding to *P*-values below 0.05 are described as statistically significant. We did not use the Bonferroni correction for multiple testing, as such a correction is considered over-conservative and one risks the rejection of important findings [21], and it is inappropriate to use this approach in a situation where there is *a priori* evidence that the genes are associated with obesity [22]. The programming environment R, and R package, genetics were used for all statistics (from www.r-project.org).

RESULTS

South African participants' characteristics

BMI distribution in study group

The study group consists of 524 (53%) female and 466 (47%) male adolescents, with a mean age of 13.7 years (SD=0.2 years). Females had a higher median BMI (19.4 [interquartile range (IQR) = 17.8-22.3]) than males (18.0 [IQR = 16.6-19.7]; P< 0.0001). BMI increased across the five pubertal stages in both males (*P*< 0.0324) and females (*P*<0.0001), but at significantly different rates. Exact age (days) correlated positively with BMI in males (*P*=0.027) but not females (*P*=0.305). In a joint model, gender, gender-specific pubertal stage and age each correlated independently with BMI and together explained 14.3% (*P*<0.001) of the variation in log(BMI). Pubertal effect when assessed alone had the greatest impact (10.7%) followed by gender (5.7%). As a result, all analyses were adjusted for age, gender and pubertal stage per gender. Unadjusted *P*-values of individual tests are presented for comparison.

Population substructure

Analysis of allele frequencies of AIMs in the South African cohort, and comparison to three other ethnic groups, 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 (i.e. higher values of K, where K indicates the number of parental populations) was shown to highlight the genetic distinction between these two African groups.

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 (results not shown). Table 2 presents the median and IQR of BMI for SNPs that were significantly associated with BMI following adjustment for gender, pubertal stage and age. Estimated effect sizes (percentage difference in BMI associated with each additional minor allele, compared to the major allele homozygote) adjusted for age, gender 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% in BMI (*P*< 0.001). Thus, for an adolescent of the same age, gender and pubertal stage, we would expect an individual with the A/G genotype at rs10954174 to have a BMI that is 4% less than that of a G/G homozygote; and the A/A homozygote is expected to have an 8% lower BMI compared to 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 whilst the G/G homozygotes would have a 6% higher BMI.

Gender specific effects

Prior research suggests that some loci show gender specific effects. In this study two SNPs – *MC4R* rs12970134 (P=0.044) and *NPY2R* (P=0.012) – showed significant gender differences in their effects on BMI. In females, each G-allele of the *NPY2R* rs11099992 SNP predicted a decrease in BMI by 3.7% (*P*-value=0.003) while in males the effect was not significant (P=0.523). In males, each A-allele of the *MC4R* rs12970134 SNP predicted an increase in BMI of 5.2% (P=0.011) while in females the effect was not significant (P=0.785). These SNPs were not associated with log(BMI) in the group as a whole.

Joint model

To assess the combined effect of the four SNPs that were individually associated with log(BMI) and the three confounders (gender, pubertal stage and age), a model was built based on 908 participants who had compete 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).

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 less than ten individuals each in the categories of one risk allele (seven individuals) and eight risk alleles (one individual), these were excluded from the figure. The estimated increase in BMI, for each risk allele, is 2.5%, independent of age, gender and gender-specific pubertal stage, and individuals with seven risk alleles had an 11% increase [(20.0-18.0)/18.0] in median BMI compared to a group of individuals with two risk alleles.

DISCUSSION

This study investigated the role of genetic variation on BMI in a cohort of adolescent South Africans, and is one of the largest analyses of obesity candidate genes undertaken in an African cohort. We confirmed that variants within *FTO* and *MC4R,* previously shown to be associated with BMI in Caucasians, also influence obesity risk in this African group. In addition, a novel association between the 3' UTR of *LEP* and increased BMI was observed. Furthermore, it was also demonstrated that an accumulation of risk alleles showed a significant increase in BMI (Figure 1). In our study, individuals with seven risk alleles had an 11% increase in median BMI compared to individuals with two risk alleles.

The role of *FTO* in obesity risk is now well established. Within the current study, seven *FTO* SNPs were analysed and rs17817449 was found be associated with BMI (*P*=0.022). It has been shown in several different European child- and adult cohorts [23 24] and in a meta-analysis [25] that this particular *FTO* SNP has a very strong association with BMI. Comparison of our data for *FTO* with that from other studies in African populations show that in Gambians, no association of BMI with polymorphisms in the *FTO* gene was observed [9]. However, the authors suggested that this lack of association may be related to the low BMI of the study population. The second study, performed in West Africans demonstrated that rs17817449 did not associate with BMI in this or an African American population [8]. This suggests different regions of the *FTO* gene may contain functional variants.

The rs17782313 SNP is one of several variants in *MC4R* linked to BMI, and has been shown to contribute to a significant increase in extreme obesity in adults [26]. In the present study this variant associated with BMI (*P*=0.045). This SNP is 3' to the gene and may therefore disrupt the function of *cis*-acting elements involved in the control of *MC4R* gene transcription. It is interesting to note that a recent study shows that the 3' SNPs, rs12970134 and rs17782313, do not represent a haplotype that extends into the coding region of the gene [27], thus suggesting that these SNPs are not in linkage disequilibrium with functional mutations in the exon of the *MC4R* gene.

Several studies have shown gender specific effects on BMI. The SNP rs11099992 in *NPY2R* was observed to associate with obesity in female subjects only (P=0.003). This

marker was recently shown to be associated with greater early infancy weight gain and growth in the ALSPAC longitudinal cohort, providing insights into the timing of weight changes leading to adult obesity [28]. The SNP lies 5' of the *NPY2R* gene and other SNPs in this region have also been linked to anthropometric variables [29]. It has been observed that SNPs in the *NPY2R* gene have effects on anthropometry in males only [30], as opposed to the female-only effect seen in the current study. However, one of these studies shows that a SNP in the peptide YY gene associates with obesity-related traits in females only [29]. Peptide YY is a ligand of the NPY2R. Thus, variants in genes of the NPY pathway seem to have gender specific effects on anthropometric variables, but the mechanism by which this occurs is currently not understood. One other SNP showed a gender specific effect in males only (P=0.011), where males with the *MC4R* A-allele (rs12970134) had a predicted 5.2% increase in BMI. No significant effect was detected in females. It is interesting to note that mutations in the *MC4R* gene that have been associated with morbid obesity have stronger effects in females than males [31]. These data suggest that the phenotypic effect of sequence variation in the *MC4R* gene is differentially modified by gender depending on the site of the variant.

We detected significant allelic effects for two SNPs in *LEP* (rs10954174, P=0.0004 and rs6966536, P=0.012), located in the 3'UTR of this gene. Leptin is an integral molecule in energy homeostasis, with circulating leptin primarily produced by adipocytes, and functioning 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 [32]**.** However, these coding sequence variations are rare in the general population and there is inconsistent evidence of the association of more common variants (predominantly SNPs in the 5'UTR) in *LEP* with body mass [33].

This is the first observation of association of these particular SNPs within the 3'UTR of the *LEP* gene with BMI in an African population. The *LEP* SNPs that have been most consistently associated with obesity are rs2167270 (also known as *LEP* 19G>A) in the untranslated first exon [34], and rs7799039 (LEP -2548G>A) in the 5'-region of the gene [35]. It is known from previous research that in Caucasians, strong LD extends from the 5'-region to the first intron of *LEP* but does not extend to SNPs in the genes distal and proximal to *LEP* [36]. Studies in a Brazilian population [37] and adult Samoans [38] have also shown obesity to be correlated to variants of the *LEP* 3' hypervariable region

(HVR). The 3'HVR is a $(TTTC)_n$ tetranucleotide repeat located approximately 2,000bp 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.

Individuals in this study who had a G-allele at *LEP* SNP rs6966536, showed an increase in BMI (Table 2). On assessment of the allelic frequency of rs6966536 in the HapMap database, it was found not to be polymorphic in the European and Asian populations (100% of individuals have the A-allele), with the Yoruba showing a minor allele frequency of 8% for the G-allele as opposed to the observed 14% in the Bt20 cohort.

The 3' *LEP* SNPs found to be associated with BMI in this study may be the actual causal SNPs that influence BMI, or may be in LD with a causative variant. There are several mechanisms by which a 3'UTR variant could impact on the functioning of a gene. MicroRNAs are evolutionarily conserved small non-coding RNAs known to inhibit the translation of proteins by binding to the target transcript in the 3' untranslated region. Functional polymorphisms in 3' UTRs of several genes have been reported to be associated with diseases by affecting gene expression [39]. Recently, a novel class of functional variants termed miRSNPs were reported and defined as a variant at or near a microRNA binding site that can affect gene expression [40]. Furthermore, the 3' untranslated regions of eukaryotic mRNAs play crucial roles in the posttranscriptional regulation of gene expression through the modulation of nucleocytoplasmic mRNA transport, translation efficiency, subcellular localization and message stability.

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 less than 0.1 at a second interacting polymorphism [41]. 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 [41]. 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 [42 43]. The AIM data obtained in this study highlights 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, eighteen 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 only interrogating a small number of markers [43].

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. It is also possible that the effects of some polymorphisms on BMI are observed only later in life. 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.

This study shows evidence of the role of genetic factors 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 (1 to 8) 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, gender and gender-specific pubertal stage. Recent GWAS in populations of European ancestry have identified 32 common loci associated with BMI, explaining only 1.45% of the inter-individual variation in BMI [4].

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 [44]. 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 characterization 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 [45].

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. This is reflected in the national burden of disease by its impact on the increase in chronic disease [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.

LEGENDS TO FIGURES:

Figure 1: Observed cumulative effect of high risk alleles on 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 the interquartile range of BMI.

TABLES:

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 13 years of age

					Frequency						
			RefSNP alleles*		Genotype			Allele		*P*-values	
Gene	dbSNP ID	n	A	B	AA	AB	BB	Minor	Major	Unadjusted	Adjusted**
FTO	rs9939973	977	A	G	0.14	0.45	0.40	0.37	0.63	0.109	0.354
	rs9940128	985	A	G	0.12	0.47	0.41	0.35	0.65	0.105	0.379
	rs1421085	986	T	C	0.89	0.11	0.00	0.06	0.94	0.728	0.709
	rs1121980	974	T	C	0.20	0.51	0.29	0.45	0.55	0.125	0.294
	rs17817449	977	T	G	0.39	0.48	0.14	0.37	0.63	**0.008**	**0.022**
	rs8050136	982	A	C	0.17	0.51	0.32	0.42	0.58	**0.021**	0.057
	rs9939609	981	A	T	0.20	0.53	0.27	0.47	0.53	**0.029**	0.071
LEP	rs1349419	986	A	G	0.96	0.04	0.00	0.02	0.98	0.388	0.612
	rs12535708	975	A	C	0.03	0.32	0.65	0.19	0.81	0.755	0.911
	rs11770725	985	T	C	0.91	0.09	0.00	0.05	0.95	0.640	0.993
	rs12535747	977	A	C	0.01	0.18	0.81	0.10	0.90	0.258	0.339
	rs2167270	964	A	G	0.24	0.51	0.25	0.50	0.50	0.557	0.255
	rs2278815	980	A	G	0.00	0.08	0.92	0.04	0.96	0.918	0.887
	rs2122627	984	T	C	0.01	0.12	0.88	0.06	0.94	0.197	0.102
	rs4236625	985	A	T	0.78	0.21	0.02	0.12	0.88	0.804	0.706
	rs12706832	979	A	G	0.80	0.18	0.01	0.10	0.90	0.787	0.908
	rs10244329	963	A	T	0.41	0.45	0.14	0.37	0.63	**0.034**	0.161
	rs7791621	984	A	C	0.86	0.13	0.01	0.07	0.93	0.761	0.792
	rs7795794	987	A	G	0.01	0.12	0.87	0.07	0.93	0.185	0.084
	rs10954174	973	A	G	0.02	0.21	0.77	0.13	0.87	**0.003**	**<0.001**
	rs3828942	962	A	G	0.03	0.28	0.69	0.17	0.83	0.922	0.852
	rs17151919	969	A	G	0.02	0.26	0.71	0.15	0.85	0.699	0.829
	rs17151922	974	T	G	0.15	0.50	0.35	0.40	0.60	0.239	0.378
	rs6966536	972	A	G	0.73	0.25	0.02	0.14	0.86	**0.014**	**0.012**
	rs10954173	977	A	G	0.02	0.26	0.72	0.15	0.85	0.342	0.525
	rs11761556	984	A	C	0.01	0.19	0.80	0.11	0.89	0.798	0.659
LEPR	rs1137100	984	A	G	0.89	0.11	0.00	0.06	0.94	0.602	0.602
	rs1137101	965	A	G	0.20	0.52	0.28	0.45	0.55	0.650	0.905
MC4R	rs17782313	969	T	C	0.55	0.39	0.06	0.25	0.75	0.070	**0.045**
	rs12970134	981	A	G	0.01	0.16	0.83	0.09	0.91	0.166	0.116
NPY2R	rs2880416	985	C	G	0.00	0.08	0.92	0.04	0.96	0.094	0.186
	rs2342676	953	A	G	0.08	0.41	0.51	0.29	0.71	0.697	0.942
	rs12649641	988	A	C	0.48	0.43	0.08	0.30	0.70	0.601	0.786
	rs11099992	964	A	G	0.58	0.36	0.06	0.24	0.76	0.085	0.073
	rs33977152	986	A	G	0.01	0.09	0.90	0.05	0.95	0.508	0.301

	rs12507396	982	A	T	0.95	0.05	0.00	0.02	0.98	0.858	0.957
	rs6857530	979	A	G	0.05	0.36	0.58	0.23	0.77	0.666	0.707
	rs10461238	971	C	G	0.32	0.49	0.19	0.43	0.57	0.640	0.430
	rs2342674	985	A	G	0.00	0.13	0.87	0.07	0.93	0.530	0.476
	rs1047214	977	T	C	0.88	0.12	0.01	0.06	0.94	0.926	0.919
	rs2880415	982	A	G	0.55	0.40	0.05	0.25	0.75	0.458	0.292
POMC	rs6713532	970	T	C	0.29	0.49	0.22	0.46	0.54	0.368	0.270
	rs7565877	972	A	G	0.37	0.49	0.14	0.39	0.61	0.944	0.967
	rs7565427	987	A	G	0.01	0.20	0.79	0.11	0.89	0.633	0.889

* RefSNP allele according to NCBI dbSNP Build 134
** *P*-values are shown adjusted for age, gender and gender-specific pubertal stage

Table 2: Summary BMI statistics for SNPs that were individually associated with BMI

				Median (IQR) of BMI per genotype			
Gene	**SNP**	**MA**	**MAF**	**mm***	**mM***	**MM***	**Effect**
FTO	rs17817449	G	0.37	18.5 (16.9-20.7)	18.7 (17.2-21.1)	19.1 (17.6-22.5)	1.9
LEP	rs10954174	A	0.13	18.8 (17.1-21.3)	18.4 (17.2-20.5)	17.4 (16.8-19.1)	-4.0
LEP	rs6966536	G	0.14	18.6 (17.1-20.8)	18.9 (17.1-21.5)	19.2 (18.5-21.6)	2.9
MC4R	rs17782313	C	0.25	18.9 (17.4-21.2)	18.5 (17.0-20.9)	19.0 (16.5-20.6)	-1.8

MA = minor allele; MAF= minor allele frequency; IQR=interquartile range;

Effect=estimated individual % change in BMI of each minor allele, independent of age, gender and gender-specific pubertal stage.

*m denotes major allele; M is minor allele

Table 3: Summary of the model explaining 16.35% of the variation in log(BMI), and containing the four variants that were individually significantly associated with BMI, with adjustment for confounders

Gene	SNP	MA	Effect	*P*-value
FTO	rs17817449	G	2.2	0.007
LEP	rs10954174	A	-3.2	0.005
LEP	rs6966536	G	2.7	0.020
MC4R	rs17782313	C	-2.0	0.028

MA=minor allele

Effect=estimated % change in BMI of each minor allele, independent of other confounding variables i.e. age, gender and gender-specific pubertal stage.

ACKNOWLEDGEMENTS

The authors would like to 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.

COMPETING INTERESTS

None declared.

FUNDING

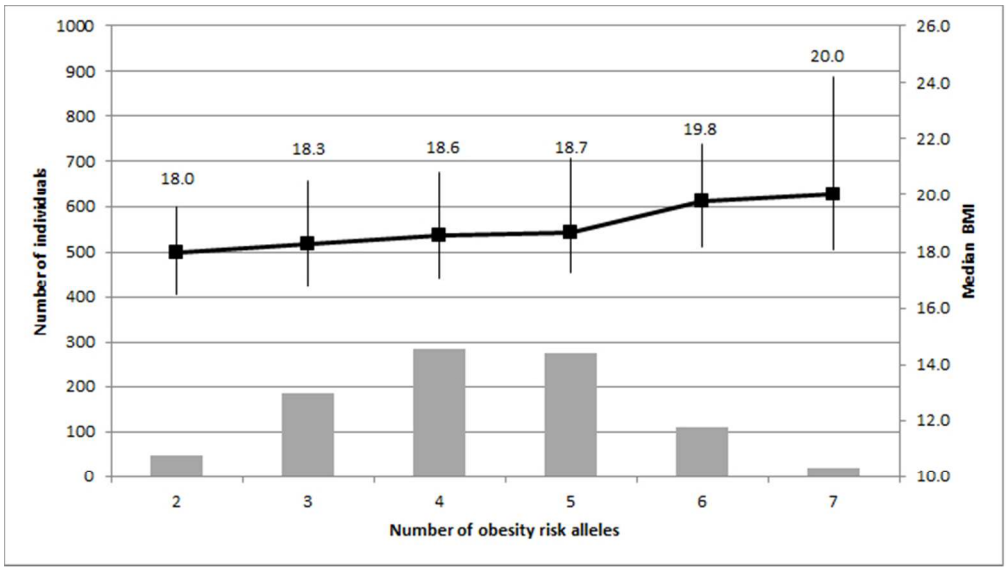
This work was supported by the Wellcome Trust (grant number 080535/Z/06/Z to SAN); the South African Medical Research Council and the University of the Witwatersrand for the collection of the samples for the Bt20 cohort and genotyping. In addition genotyping was funded by the South African National Research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

REFERENCES

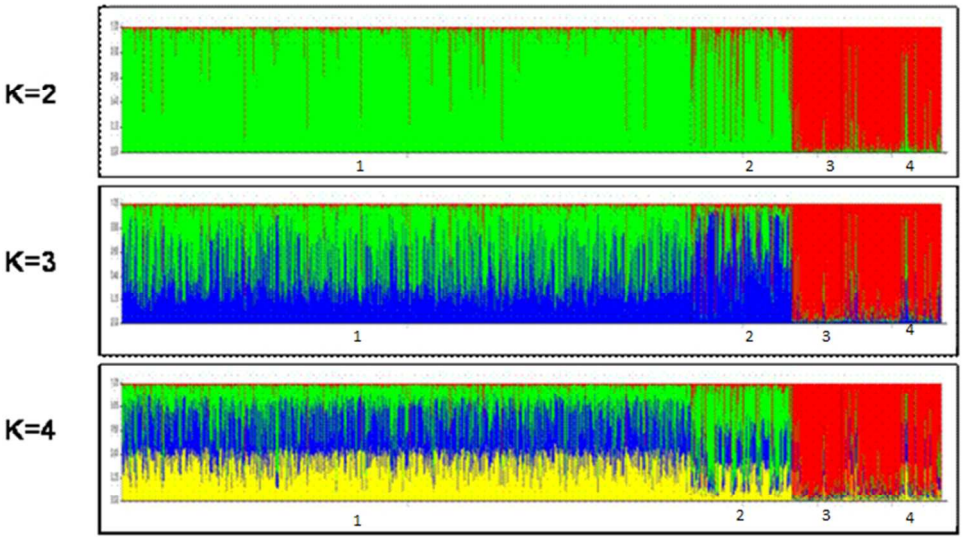
1. Joubert J, Norman R, Bradshaw D, Goedecke JH, Steyn NP, Puoane T. Estimating the burden of disease attributable to excess body weight in South Africa in 2000. S Afr Med J 2007;**97**(8 Pt 2):683-90.
2. Bodurtha JN, Mosteller M, Hewitt JK, et al. Genetic analysis of anthropometric measures in 11-year-old twins: the Medical College of Virginia Twin Study. Pediatr Res 1990;**28**(1):1-4.
3. 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**(8):753-60.
4. 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**(11):937-48.
5. 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**(13):2725-38.
6. Bressler J, Kao WH, Pankow JS, Boerwinkle E. 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**(5):e10521.
7. 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**(14):2907-16.
8. Adeyemo A, Chen G, Zhou J, et al. FTO genetic variation and association with obesity in West Africans and African Americans. Diabetes 2010;**59**(6):1549-54.
9. 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.
10. Teo YY, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. Nat Rev Genet 2010;**11**(2):149-60.
11. Church C, Moir L, McMurray F, et al. Overexpression of Fto leads to increased food intake and results in obesity. Nat Genet 2010;**42**(12):1086-92.
12. Richter L, Norris S, Pettifor J, Yach D, Cameron N. Cohort Profile: Mandela's children: the 1990 Birth to Twenty study in South Africa. Int J Epidemiol 2007;**36**(3):504-11.
13. Stettler N, Iotova V. Early growth patterns and long-term obesity risk. Curr Opin Clin Nutr Metab Care 2010;**13**(3):294-9.
14. QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies [program]. http://hydra.usc.edu/gxe, 2006.
15. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;**16**(3):1215.
16. Cameron N, De Wet T, Ellison GT, Bogin B. Growth in height and weight of South African urban infants from birth to five years: The Birth to Ten study. American Journal of Human Biology 1998;**10**(4):495-504.
17. Norris SA, Richter LM. Validation of Tanner pubertal self-rating to urban black adolescents in South Africa. Journal of Research on Adolescence 2005;**15**:609-24.
18. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;**155**(2):945-59.
19. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet 2005;**37**(11):1217-23.

20. Tindall EA, Petersen DC, Nikolaysen S, Miller W, Schuster SC, Hayes VM. Interpretation of custom designed Illumina genotype cluster plots for targeted association studies and next-generation sequence validation. BMC Res Notes 2010;**3**:39.
21. Perneger TV. What's wrong with Bonferroni adjustments. BMJ 1998;**316**(7139):1236-8.
22. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. Am J Hum Genet 2004;**74**(4):765-9.
23. Wang K, Li WD, Zhang CK, et al. A genome-wide association study on obesity and obesity-related traits. PLoS One 2011;**6**(4):e18939.
24. Dina C, Meyre D, Gallina S, et al. Variation in FTO contributes to childhood obesity and severe adult obesity. Nat Genet 2007;**39**(6):724-6.
25. Peng S, Zhu Y, Xu F, Ren X, Li X, Lai M. FTO gene polymorphisms and obesity risk: a meta-analysis. BMC Med 2011;**9**:71.
26. 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**(6):768-75.
27. Scherag A, Jarick I, Grothe J, et al. Investigation of a genome wide association signal for obesity: synthetic association and haplotype analyses at the melanocortin 4 receptor gene locus. PLoS One 2010;**5**(11):e13967.
28. 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**(5):e1000284.
29. 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**(5):387-98.
30. 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**(5):1598-602.
31. 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**(10):795-800.
32. Farooqi IS. Genetic aspects of severe childhood obesity. Pediatr Endocrinol Rev 2006;**3 Suppl 4**:528-36.
33. Cieslak J, Bartz M, Stachowiak M, et al. Effect of three common SNPs in 5'-flanking region of LEP and ADIPOQ genes on their expression in Polish obese children and adolescents. Mol Biol Rep 2011.
34. Mattevi VS, Zembrzuski 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**(9):1179-85.
35. Constantin A, Costache G, Sima AV, Glavce CS, Vladica M, Popov DL. Leptin G-2548A and leptin receptor Q223R gene polymorphisms are not associated with obesity in Romanian subjects. Biochem Biophys Res Commun 2010;**391**(1):282-6.
36. Jiang Y, Wilk JB, Borecki I, et al. Common variants in the 5' region of the leptin gene are associated with body mass index in men from the National Heart, Lung, and Blood Institute Family Heart Study. Am J Hum Genet 2004;**75**(2):220-30.
37. 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**(3):282-8.

38. 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**(6):783-8.
39. Mishra PJ, Banerjee D, Bertino JR. MiRSNPs or MiR-polymorphisms, new players in microRNA mediated regulation of the cell: Introducing microRNA pharmacogenomics. Cell Cycle 2008;**7**(7):853-8.
40. Mishra PJ, Humeniuk R, Longo-Sorbello GS, Banerjee D, Bertino JR. A miR-24 microRNA binding-site polymorphism in dihydrofolate reductase gene leads to methotrexate resistance. Proc Natl Acad Sci U S A 2007;**104**(33):13513-8.
41. Greene CS, Penrod NM, Williams SM, Moore JH. Failure to replicate a genetic association may provide important clues about genetic architecture. PLoS One 2009;**4**(6):e5639.
42. 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**(2):786-91.
43. Tishkoff SA, Reed FA, Friedlaender FR, et al. The genetic structure and history of Africans and African Americans. Science 2009;**324**(5930):1035-44.
44. 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.
45. Kissebah AH, Krakower GR. Regional adiposity and morbidity. Physiol Rev 1994;**74**(4):761-811.



Observed cumulative effect of high risk alleles on 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 the interquartile range of BMI.



STRUCTURE analysis with 18 AIMs illustrating no significant population substructure in the Bt20 cohort. The values of K indicate the number of hypothetical ancestral populations. The clustering at K = 2 distinctly separated the African ((1) South African and (2) YRI) from the Non-African ((3) CHB and (4) CEU) populations. Higher order clustering (K=3 to K=4) was found to resolve the Bt20 cohort from all other populations, indicating that these 18 SNPs have the power to detect major differences in population structure. Population designations: (1) South African cohort; (2) HapMap YRI; (3) HapMap CHB; (4) HapMap CEU.

Supplementary Figure Legend:

Supplementary Figure 1: STRUCTURE analysis with 18 AIMs illustrating no significant population substructure in the Bt20 cohort. The values of K indicate the number of hypothetical ancestral populations. The clustering at K = 2 distinctly separated the African ((1) South African and (2) YRI) from the Non-African ((3) CHB and (4) CEU) populations. Higher order clustering (K=3 to K=4) was found to resolve the Bt20 cohort from all other populations, indicating that these 18 SNPs have the power to detect major differences in population structure. Population designations: (1) South African cohort; (2) HapMap YRI; (3) HapMap CHB; (4) HapMap CEU.

Supplementary Tables:

Table 1: Summary of SNP information genotyped in candidate genes

Gene	SNP	Genomic location *	Chrom	RefSNP Alleles **	Ancestral allele **	BMI assoc SNP	TagSNP	AIM
FTO	rs9939973	53800568	16	A/G	G	x		
	rs9940128	53800754		A/G	G	x		
	rs1421085	53800954		C/T	T	x		
	rs1121980	53809247		C/T	T	x		
	rs17817449	53813367		G/T	G	x		
	rs8050136	53816275		A/C	A	x		
	rs9939609	53820527		A/T	A	x		
LEP	rs1349419	127877213	7	A/G	A	x		
	rs12535708	127878098		A/C	C	x		
	rs11770725	127878267		C/T	T	x		
	rs12535747	127878335		A/C	C	x		
	rs2167270	127881349		A/G	G		x	
	rs2278815	127881851		A/G	G		x	
	rs2122627	127883323		C/T	C		x	
	rs4236625	127883695		A/T	T		x	
	rs12706832	127887139		A/G	A		x	
	rs10244329	127888689		A/T	T		x	
	rs7791621	127889696		A/C	A		x	
	rs7795794	127890151		A/G	C		x	
	rs10954173	127891440		A/G	G		x	
	rs3828942	127894305		A/G	G		x	
	rs17151919	127894592		A/G	G		x	
	rs17151922	127895216		G/T	G		x	
	rs6966536	127896059		A/G	A		x	
	rs10954174	127896536		A/G	G		x	

	rs11761556	127897069		A/C	A		x	
LEPR	rs1137100	66036441	1	A/G	A	x		
	rs1137101	66058513		A/G	A	x		
MC4R	rs17782313	57851097	18	C/T	T	x		
	rs12970134	57884750		A/G	G	x		
NPY2R	rs2880416	156124543	4	C/G	C		x	
	rs2342676	156124704		A/G	A	x		
	rs12649641	156125333		A/C	A		x	
	rs11099992	156125510		A/G	A		x	
	rs33977152	156128589		A/G	G		x	
	rs12507396	156129044		A/T	A		x	
	rs6857530	156129154		A/G	A		x	
	rs10461238	156132216		C/G	G	x		
	rs2342674	156135250		A/G	G		x	
	rs1047214	156135676		C/T	T	x		
	rs2880415	156136027		A/G	A		x	
POMC	rs6713532	25384833	2	C/T	C		x	
	rs7565427	25385638		A/G	G		x	
	rs7565877	25386064		A/G	A		x	
Ancestry informative markers	rs723854	192511012	1	C/G	G			x
	rs1876482	17362568	2	C/T	C			x
	rs952718	215888624	2	A/C	A			x
	rs1344870	21307401	3	A/C	A			x
	rs720096	179551071	4	A/C	C			x
	rs1363448	140783596	5	C/T	T			x
	rs217538	108483470	6	C/G	C			x
	rs65264	28545611	7	C/T	T			x
	rs679047	12883664	9	A/T	T			x
	rs2077559	36014850	9	C/T	T			x
	rs714857	15974389	11	C/T	T			x
	rs953386	110943692	13	A/G	A			x
	rs722869	97277005	14	C/G	C			x

	rs735612	34076642	15	G/T	G			x
	rs2089740	36310531	15	G/T	G			x
	rs1823718	74147244	15	C/T	T			x
	rs1858465	51142920	17	A/T	A			x
	rs2112527	9603751	19	A/G	G			x

* Chromosomal location of SNPs obtained from NCBI genome build 37.1 (hg19).
** RefSNP allele and ancestral alleles according to NCBI dbSNP Build 134.

BMJ open

Appetite Regulation Genes are Associated with Body-Mass Index in Black South African Adolescents: A Genetic Association Study

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-000873.R1
Article Type:	Research
Date Submitted by the Author:	26-Mar-2012
Complete List of Authors:	Lombard, Zané; National Health Laboratory Service & University of the Witwatersrand, Division of Human Genetics Crowther, Nigel; National Health Laboratory Service & University of the Witwatersrand, Chemical Pathology van der Merwe, Lize; Medical Research Council of South Africa, Biostatistics Unit Pitamber, Punita; National Health Laboratory Service & University of the Witwatersrand, Division of Human Genetics Norris, Shane; University of the Witwatersrand, MRC/Wits Developmental Pathways for Health Research Unit Ramsay, Michele; National Health Laboratory Service & University of the Witwatersrand, Division of Human Genetics
Primary Subject Heading:	Genetics and genomics
Secondary Subject Heading:	Nutrition and metabolism, Public health
Keywords:	Cell biology < BASIC SCIENCES, Genetics < TROPICAL MEDICINE, Nutrition < TROPICAL MEDICINE

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], Punita Pitamber[1], Shane A Norris[5], Michèle Ramsay[1]

[1] Division Human Genetics, School of Pathology, Faculty of Health Sciences, National Health Laboratory Service & University of the Witwatersrand, Johannesburg, 2000, South Africa

[2] Current address: Wits Bioinformatics, School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg, 2000, South Africa

[3] Department of Chemical Pathology, School of Pathology, Faculty of Health Sciences, National Health Laboratory Service & University of the Witwatersrand, Johannesburg, 2000, South Africa

[4] Biostatistics Unit, Medical Research Council, Cape Town, 8001, South Africa and Department of Statistics, University of Western Cape, Bellville, South Africa

[5] MRC/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, 2000, South Africa

***Corresponding author:** Zané Lombard; Postal address: Wits Bioinformatics, Private Bag 3, 2050 Wits, Johannesburg, South Africa; Email: zane.lombard@wits.ac.za; Telephone: +27117176393; Fax: +27867656304

Key words: Obesity, Body Mass Index, Genetic Association Studies, *FTO*, *LEP*, *MC4R*

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 BMI many of which are found within genes involved in appetite regulation. Currently, genetic association data for obesity is lacking in Africans - a single GWAS and a few replication studies have been published in West Africans, but none have been performed in a South African population.
Objectives: To assess the association of candidate with BMI in black South Africans. We focused on 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: We genotyped 44 SNPs within the six candidate genes that included known BMI associated SNPs and tagSNPs based on linkage disequilibrium in an African populations for FTO, LEP and NPY2R. To assess population substructure, we included 18 ancestry informative markers. Weight, height, sex, sex-specific pubertal stage and exact age collected during adolescence (13 years of age) 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); 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 FTO and MC4R as potential genetic markers of obesity risk in South Africans. The association with two SNPs in the 3' UTR 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 to individuals 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've 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 BMI estimated at 40-70%[2].

The study of syndromic obesity and genome-wide association studies (GWAS) of common measures of adiposity (e.g. 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 1,188 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 characterized 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 intake[11] 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 2 diabetes cases and 68% of hypertensive disease were attributable to a BMI ≥ 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 wellbeing of children that were born in the metropolitan area of Soweto, Johannesburg, during a six-week period in 1990. Participants in Bt20 were enrolled at birth and detailed information has been collected from the participants and their caregivers over the past twenty 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 Wise 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 (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 standard deviation which is 0.47. The common standard deviation 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 kg) divided by the square of the height (measured in meters) 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 were 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 data[17] 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.

SNP selection

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 algorithm[20] was used to select tagSNPs in a multimarker approach at $r^2 > 0.80$ and with a minor allele frequency (MAF) > 0.05 among publicly available African data (YRI) from the HapMap dataset (release 24). The Illumina Assay Design Tool (Illumina, San Diego, CA) 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, San Diego, CA, USA). Quality control was performed on all raw genotype data according to the supplier's specifications using BeadStudio v2.0 software (Illumina, San Diego, CA) (summarized in[21]). Standard quality control filters were applied to the data[22]: minor allele frequency (MAF) >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 visualized using Haploview v4.2[23], implementing the confidence interval method of Gabriel et al.[24] to construct haplotype blocks. Chromosomal location of SNPs obtained from ncbi genome build 37.1.

Statistical Analysis

The programming environment R, and R package genetics were used for all statistics (from 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(BMI). Exact tests of HWE were also performed. Because of its skewed distribution in this study group, BMI is summarized as median values as well as inter-quartile ranges (IQR), with lower quartile and upper quartiles reported.

Furthermore, each genotype was modeled additively as the number of minor alleles present. Because we modelled log(BMI), the modeled 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 summarized 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 modeled the interaction between sex and each variant and summarized the significant results.

Correcting for multiple testing

Correcting for multiple testing is a contentious issue, and some approaches (such as Bonferroni correction) is considered over-conservative 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 linkage disequilibrium, 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). Females had a higher median BMI (19.4 [IQR = 17.8-22.3]) than males (18.0 [IQR = 16.6-19.7]; P< 0.0001). BMI increased across the five pubertal stages in both males (*P*< 0.0324) and females (*P*<0.0001), but at significantly different rates. Exact age (days) correlated positively with BMI in males (*P*=0.027) but not females (*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(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 were subsequently excluded from analysis. Analysis of allele frequencies of AIMs in the South African cohort (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 (i.e. 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 was constructed (Supplementary Figure 2). Evidence of LD structure in the 3'UTR of *LEP* can be observed, which is different to the haplotype structure observed in the HapMap European and Nigerian populations.

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 to 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 to 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 whilst 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 females, each G-allele of the *NPY2R* rs11099992 SNP predicted a decrease in BMI by 3.7% (*P*-value=0.003) while in males the effect was not significant (*P*=0.523). In males, each A-allele of the *MC4R* rs12970134 SNP predicted an increase in BMI of 5.2% (*P*=0.011) while in females 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 compete 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 less than ten individuals each in the categories of one risk allele (seven individuals) and eight risk alleles (one 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 to 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 inter-individual 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 to individuals with two risk alleles.

We provide preliminary evidence that SNPs in the 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 population[34,36] and adult Samoans[37] have also shown obesity to be correlated to variants of the *LEP* 3' hypervariable region (HVR). The 3'HVR is a $(TTTC)_n$ tetranucleotide repeat located approximately 2,000bp 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 localization and message stability.[38].

Variants in *FTO* (the only gene consistently associated with obesity) have been associated with adiposity measures in Europeans,[4,39] Asians[40,41] and African Americans[7], with the strongest effect seen with SNP rs9939609. Our study showed a modest association 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 linkage disequilibrium 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 males (P=0.011). It has been observed that SNPs in the *NPY2R* gene have effects on obesity in males 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 females only [45]. Furthermore, mutations in the *MC4R* gene that have been associated with morbid obesity have stronger effects in females than males[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 less than 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 highlights 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, eighteen 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 only 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 characterization 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].

LEGENDS TO FIGURES:

Figure 1: Observed cumulative effect of high risk alleles on 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 the interquartile range of BMI.

TABLES:

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 13 years of age. Significant *P*-values are shown in bold.

					Frequency							
			RefSNP alleles*		Genotype			Allele		*P*-values		
Gene	dbSNP ID	n	A	B	AA	AB	BB	Minor	Major	Unadjusted	Adjusted**	HWE
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
LEP	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
	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
	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
LEPR	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
MC4R	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
NPY2R	rs2880416	985	C	G	0.00	0.08	0.92	0.04	0.96	0.094	0.186	1.000
	rs2342676	953	A	G	0.08	0.41	0.51	0.29	0.71	0.697	0.942	0.937
	rs12649641	988	A	C	0.48	0.43	0.08	0.30	0.70	0.601	0.786	0.255
	rs11099992	964	A	G	0.58	0.36	0.06	0.24	0.76	0.085	0.073	1.000
	rs33977152	986	A	G	0.01	0.09	0.90	0.05	0.95	0.508	0.301	0.181
	rs12507396	982	A	T	0.95	0.05	0.00	0.02	0.98	0.858	0.957	0.417
	rs6857530	979	A	G	0.05	0.36	0.58	0.23	0.77	0.666	0.707	0.790
	rs10461238	971	C	G	0.32	0.49	0.19	0.43	0.57	0.640	0.430	0.948
	rs2342674	985	A	G	0.00	0.13	0.87	0.07	0.93	0.530	0.476	0.308
	rs1047214	977	T	C	0.88	0.12	0.01	0.06	0.94	0.926	0.919	0.581
	rs2880415	982	A	G	0.55	0.40	0.05	0.25	0.75	0.458	0.292	0.061
POMC	rs6713532	970	T	C	0.29	0.49	0.22	0.46	0.54	0.368	0.270	0.796
	rs7565877	972	A	G	0.37	0.49	0.14	0.39	0.61	0.944	0.967	0.742

	rs7565427	987	A	G	0.01	0.20	0.79	0.11	0.89	0.633	0.889	0.418

* RefSNP allele according to NCBI dbSNP Build 134
** *P*-values are shown adjusted for age, sex and sex-specific pubertal stage

Table 2: Summary BMI statistics for SNPs that were individually associated with BMI

		Typed	Modeled			Median (IQR) of BMI per genotype			
Gene	**SNP**			**MA**	**MAF**	**mm***	**mM***	**MM***	**Effect**
FTO	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
LEP	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
LEP	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
MC4R	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

MA = minor allele; MAF= minor allele frequency; IQR=interquartile range;

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

Table 3: **SNPs individually associated with log(BMI) and three confounders (sex, pubertal stage and age). The model explains 16.4% of the covariation in log(BMI)**

Gene	SNP	MA	Effect	*P*-value
FTO	rs17817449	G	2.2	0.007
LEP	rs10954174	A	-3.2	0.005
LEP	rs6966536	G	2.7	0.020
MC4R	rs17782313	C	-2.0	0.028

MA=minor allele

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).

ACKNOWLEDGEMENTS

The authors would like to 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.

COMPETING INTERESTS

None declared.

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).

REFERENCES

1. 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.

2. 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.

3. 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.

4. 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.

5. 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.

6. 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.

7. 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.

8. 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 doi: db09-1252.

9. 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.

10. Teo YY, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. Nat Rev Genet 2010;11:149-60.

11. 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.

12. 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.

13. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16:1215.

14. Stettler N, Iotova V. Early growth patterns and long-term obesity risk. Curr Opin Clin Nutr Metab Care 2010;13:294-9.

15. 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. American Journal of Human Biology 1998;10:495-504.

16. Norris SA, Richter LM. Validation of Tanner pubertal self-rating to urban black adolescents in South Africa. Journal of Research on Adolescence 2005;15:609-24.

17. 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.

18. Schlebusch CM. Genetic variation in Khoisan-speaking populations from southern Africa. Unpublished PhD thesis. University of the Witwatersrand, Johannesburg, South Africa, 2011.

19. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155:945-59.

20. de Bakker PI, Yelensky R, Pe'er I, et al. Efficiency and power in genetic association studies. Nat Genet 2005;37:1217-23.

21. Tindall EA, Petersen DC, Nikolaysen 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.

22. Clarke GM, Anderson CA, Pettersson FH, et al. Basic statistical analysis in genetic case-control studies. Nature protocols 2011;6:121-33.

23. Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.

24. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225-9.

25. Perneger TV. What's wrong with Bonferroni adjustments. BMJ 1998;316:1236-8.

26. Elsäßer A, Victor A, Hommel G. Multiple Testing in Candidate Gene Situations: A Comparison of Classical, Discrete, and Resampling-Based Procedures. Statistical Applications in Genetics and Molecular Biology 2011;10:e51.

27. Farooqi IS. Genetic aspects of severe childhood obesity. Pediatr Endocrinol Rev 2006;3 Suppl 4:528-36.

28. 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.

29. Mattevi VS, Zembrzuski 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.

30. 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.

31. 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.

32. 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.

33. 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.

34. 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.

35. 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.

36. 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.

37. 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.

38. Mishra PJ, Banerjee D, Bertino JR. MiRSNPs or MiR-polymorphisms, new players in microRNA mediated regulation of the cell: Introducing microRNA pharmacogenomics. Cell Cycle 2008;7:853-8.

39. Wang K, Li WD, Zhang CK, et al. A genome-wide association study on obesity and obesity-related traits. PLoS One 2011;6:e18939.

40. 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.

41. 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.

42. 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.

43. Chen JM, Ferec 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.

44. 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.

45. 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.

46. 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.

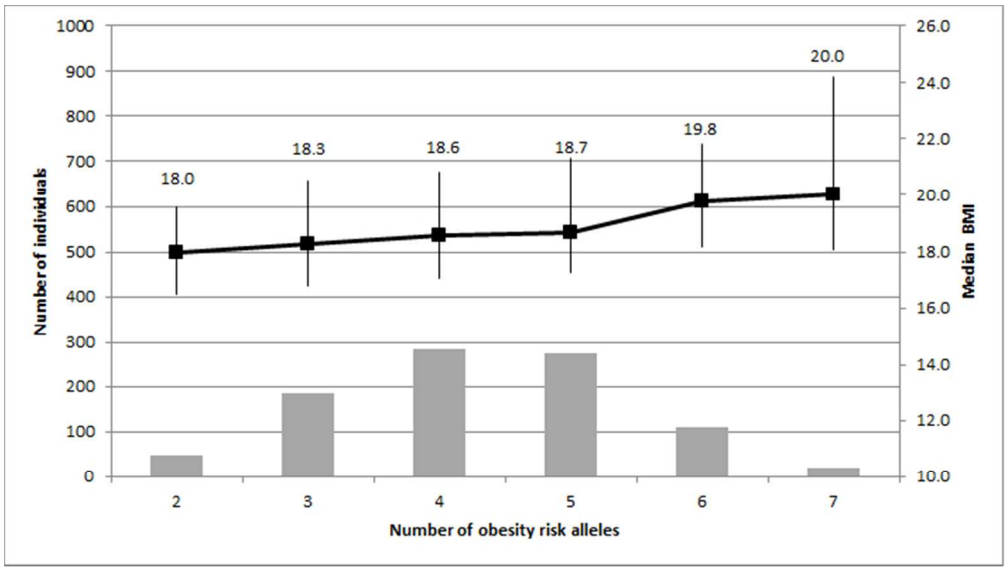
55. Greene CS, Penrod NM, Williams SM, et al. Failure to replicate a genetic association may provide important clues about genetic architecture. PLoS One 2009;4:e5639.

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.

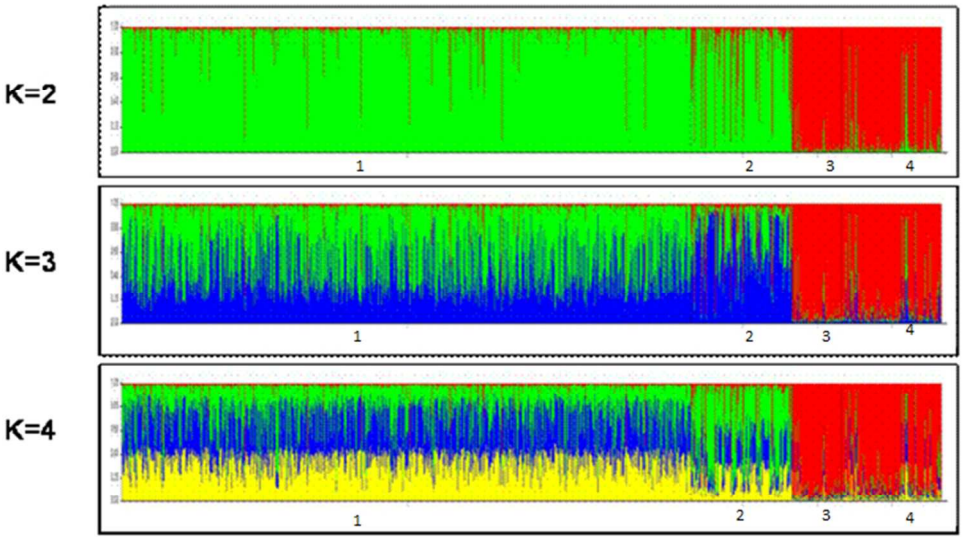
57. Tishkoff SA, Reed FA, Friedlaender FR, et al. The genetic structure and history of Africans and African Americans. Science 2009;324:1035-44.

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.

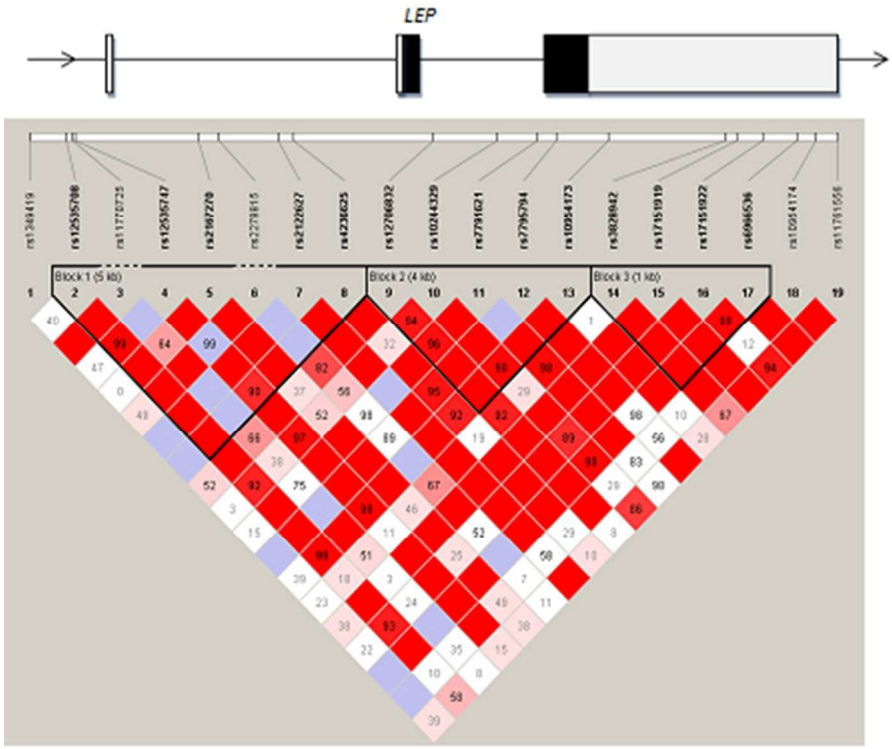
59. Kissebah AH, Krakower GR. Regional adiposity and morbidity. Physiol Rev 1994;74:761-811.



Observed cumulative effect of high risk alleles on 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 the interquartile range of BMI.



STRUCTURE analysis with 18 AIMs illustrating no significant population substructure in the Bt20 cohort. The values of K indicate the number of hypothetical ancestral populations. The clustering at K = 2 distinctly separated the African ((1) South African and (2) YRI) from the Non-African ((3) CHB and (4) CEU) populations. Higher order clustering (K=3 to K=4) was found to resolve the Bt20 cohort from all other populations, indicating that these 18 SNPs have the power to detect major differences in population structure. Population designations: (1) South African cohort; (2) HapMap YRI; (3) HapMap CHB; (4) HapMap CEU.



Supplementary Figure Legend:

Supplementary Figure 1: STRUCTURE analysis with 18 AIMs illustrating no significant population substructure in the Bt20 cohort. The values of K indicate the number of hypothetical ancestral populations. The clustering at K = 2 distinctly separated the African ((1) South African and (2) YRI) from the Non-African ((3) CHB and (4) CEU) populations. Higher order clustering (K=3 to K=4) was found to resolve the Bt20 cohort from all other populations, indicating that these 18 SNPs have the power to detect major differences in population structure. Population designations: (1) South African cohort; (2) HapMap YRI; (3) HapMap CHB; (4) HapMap CEU.

Supplementary Tables:

Supplementary table 1: Summary of SNP information genotyped in candidate genes

Gene	SNP	Genomic location *	Chrom	RefSNP Alleles **	Ancestral allele **	BMI assoc SNP	TagSNP	AIM
FTO	rs9939973	53800568	16	A/G	G	x		
	rs9940128	53800754		A/G	G	x		
	rs1421085	53800954		C/T	T	x		
	rs1121980	53809247		C/T	T	x		
	rs17817449	53813367		G/T	G	x		
	rs8050136	53816275		A/C	A	x		
	rs9939609	53820527		A/T	A	x		
LEP	rs1349419	127877213	7	A/G	A	x		
	rs12535708	127878098		A/C	C	x		
	rs11770725	127878267		C/T	T	x		
	rs12535747	127878335		A/C	C	x		
	rs2167270	127881349		A/G	G		x	
	rs2278815	127881851		A/G	G		x	
	rs2122627	127883323		C/T	C		x	
	rs4236625	127883695		A/T	T		x	
	rs12706832	127887139		A/G	A		x	
	rs10244329	127888689		A/T	T		x	
	rs7791621	127889696		A/C	A		x	
	rs7795794	127890151		A/G	C		x	
	rs10954173	127891440		A/G	G		x	
	rs3828942	127894305		A/G	G		x	
	rs17151919	127894592		A/G	G		x	
	rs17151922	127895216		G/T	G		x	
	rs6966536	127896059		A/G	A		x	
	rs10954174	127896536		A/G	G		x	

	rs11761556	127897069		A/C	A		x	
LEPR	rs1137100	66036441	1	A/G	A	x		
	rs1137101	66058513		A/G	A	x		
MC4R	rs17782313	57851097	18	C/T	T	x		
	rs12970134	57884750		A/G	G	x		
NPY2R	rs2880416	156124543	4	C/G	C		x	
	rs2342676	156124704		A/G	A	x		
	rs12649641	156125333		A/C	A		x	
	rs11099992	156125510		A/G	A		x	
	rs33977152	156128589		A/G	G		x	
	rs12507396	156129044		A/T	A		x	
	rs6857530	156129154		A/G	A		x	
	rs10461238	156132216		C/G	G	x		
	rs2342674	156135250		A/G	G		x	
	rs1047214	156135676		C/T	T	x		
	rs2880415	156136027		A/G	A		x	
POMC	rs6713532	25384833	2	C/T	C		x	
	rs7565427	25385638		A/G	G		x	
	rs7565877	25386064		A/G	A		x	
Ancestry informative markers	rs723854	192511012	1	C/G	G			x
	rs1876482	17362568	2	C/T	C			x
	rs952718	215888624	2	A/C	A			x
	rs1344870	21307401	3	A/C	A			x
	rs720096	179551071	4	C/G	C			x
	rs1363448	140783596	5	C/T	T			x
	rs217538	108483470	6	C/G	C			x
	rs65264	28545611	7	C/T	T			x
	rs679047	12883664	9	A/T	T			x
	rs2077559	36014850	9	C/T	T			x
	rs714857	15974389	11	C/T	T			x
	rs953386	110943692	13	A/G	A			x
	rs722869	97277005	14	C/G	C			x

	rs735612	34076642	15	G/T	G			x
	rs2089740	36310531	15	G/T	G			x
	rs1823718	74147244	15	C/T	T			x
	rs1858465	51142920	17	A/T	A			x
	rs2112527	9603751	19	A/G	G			x

* Chromosomal location of SNPs obtained from NCBI genome build 37.1 (hg19).
** RefSNP allele and ancestral alleles according to NCBI dbSNP Build 134.

Supplementary table 2: Allele- and genotype frequencies of AIMs in the South African cohort

				Frequency				
		RefSNP alleles*		Genotype			Allele	
SNP	n	A	B	AA	AB	BB	A	B
rs723854	974	C	G	0.05	0.31	0.65	0.20	0.80
rs1876482	983	T	C	0.00	0.00	1.00	0.00	1.00
rs952718	973	A	C	0.49	0.42	0.09	0.70	0.30
rs1344870	977	A	C	0.90	0.10	0.00	0.95	0.05
rs720096	970	C	G	0.64	0.32	0.05	0.79	0.21
rs1363448	981	T	C	1.00	0.00	0.00	1.00	0.00
rs217538	984	C	G	0.00	0.03	0.97	0.02	0.98
rs65264	982	T	C	0.64	0.33	0.03	0.80	0.20
rs679047	978	A	T	0.08	0.43	0.50	0.29	0.71
rs2077559	972	T	C	0.44	0.45	0.10	0.67	0.33
rs953386	988	A	G	0.21	0.50	0.30	0.46	0.54
rs722869	986	C	G	0.79	0.20	0.01	0.89	0.11
rs735612	979	T	G	0.21	0.51	0.28	0.47	0.53
rs2089740	970	T	G	0.11	0.46	0.42	0.35	0.65
rs1823718	972	T	C	0.62	0.35	0.03	0.79	0.21
rs1858465	972	A	T	0.02	0.26	0.72	0.15	0.85
rs2112527	984	A	G	0.43	0.47	0.11	0.66	0.34

* RefSNP allele and ancestral alleles according to NCBI dbSNP Build 134.

Supplementary table 3: Allele frequencies of AIMs in the South African cohort (SAB) and three HapMap populations

			Allele Frequencies							
	RefSNP alleles*		YRI		CEU		CHB		SAB	
SNP ID	A	B	A	B	A	B	A	B	A	B
rs723854	C	G	0.19	0.81	0.62	0.38	0.62	0.38	0.20	0.80
rs1876482	T	C	0.00	1.00	0.03	0.98	0.70	0.30	0.00	1.00
rs952718	A	C	0.74	0.26	0.11	0.89	0.04	0.96	0.70	0.30
rs1344870	A	C	0.95	0.05	0.98	0.02	0.73	0.27	0.95	0.05
rs720096	C	G	0.04	0.96	0.53	0.48	0.01	0.99	0.79	0.21
rs1363448	T	C	0.86	0.14	0.42	0.58	0.35	0.65	1.00	0.00
rs217538	C	G	0.97	0.03	0.68	0.32	0.41	0.59	0.02	0.98
rs65264	T	C	0.87	0.13	0.67	0.33	0.42	0.58	0.80	0.20
rs679047	A	T	0.22	0.78	0.08	0.92	0.09	0.91	0.29	0.71
rs2077559	T	C	0.77	0.23	0.88	0.12	0.73	0.27	0.67	0.33
rs953386	A	G	0.46	0.54	0.15	0.85	0.09	0.91	0.46	0.54
rs722869	C	G	0.93	0.07	0.88	0.12	0.14	0.86	0.89	0.11
rs735612	T	G	0.42	0.58	0.61	0.39	0.98	0.02	0.47	0.53
rs2089740	T	G	0.29	0.71	0.06	0.94	0.32	0.69	0.35	0.65
rs1823718	T	C	0.77	0.23	0.54	0.46	0.99	0.01	0.79	0.21
rs1858465	A	T	0.90	0.10	0.18	0.82	0.34	0.66	0.15	0.85
rs2112527	A	G	0.55	0.45	0.93	0.07	0.97	0.03	0.66	0.34

* RefSNP allele and ancestral alleles according to NCBI dbSNP Build 134

YRI – HapMap Yoruba from Nigeria; CEU – HapMap Utah residents with European ancestry ; CHB – HapMap Han Chinese population; SAB – South African black population (Bt20).

Appetite Regulation Genes are Associated with Body-Mass Index in Black South African Adolescents: A Genetic Association Study

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-000873.R2
Article Type:	Research
Date Submitted by the Author:	19-Apr-2012
Complete List of Authors:	Lombard, Zané; National Health Laboratory Service & University of the Witwatersrand, Faculty of Health Sciences, Division of Human Genetics; University of the Witwatersrand, Wits Bioinformatics, School of Molecular and Cell Biology Crowther, Nigel; National Health Laboratory Service & University of the Witwatersrand, Faculty of Health Sciences, Department of Chemical Pathology van der Merwe, Lize; Medical Research Council of South Africa, Biostatistics Unit Pitamber, Punita; National Health Laboratory Service & University of the Witwatersrand, Faculty of Health Sciences, Division of Human Genetics Norris, Shane; University of the Witwatersrand, Faculty of Health Sciences, MRC/Wits Developmental Pathways for Health Research Unit Ramsay, Michele; National Health Laboratory Service & University of the Witwatersrand, Faculty of Health Sciences, Division of Human Genetics
Primary Subject Heading:	Genetics and genomics
Secondary Subject Heading:	Nutrition and metabolism, Public health
Keywords:	Cell biology < BASIC SCIENCES, Genetics < TROPICAL MEDICINE, Nutrition < TROPICAL MEDICINE

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], Punita Pitamber[1], Shane A Norris[5], Michèle Ramsay[1]

[1] Division Human Genetics, School of Pathology, Faculty of Health Sciences, National Health Laboratory Service & University of the Witwatersrand, Johannesburg, 2000, South Africa

[2] Current address: Wits Bioinformatics, School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg, 2000, South Africa

[3] Department of Chemical Pathology, School of Pathology, Faculty of Health Sciences, National Health Laboratory Service & University of the Witwatersrand, Johannesburg, 2000, South Africa

[4] Biostatistics Unit, Medical Research Council, Cape Town, 8001, South Africa and Department of Statistics, University of Western Cape, Bellville, South Africa

[5] MRC/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, 2000, South Africa

***Corresponding author:** Zané Lombard; Postal address: Wits Bioinformatics, School of Molecular and Cell Biology, Private Bag 3, 2050 Wits, Johannesburg, South Africa; Email: zane.lombard@wits.ac.za; Telephone: +27117176393; Fax: +27867656304

Key words: Obesity, Body Mass Index, Genetic Association Studies, *FTO*, *LEP*, *MC4R*

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 BMI many of which are found within genes involved in appetite regulation. Currently, genetic association data for obesity is lacking in Africans - a single GWAS and a few replication studies have been published in West Africans, but none have been performed in a South African population.
Objectives: To assess the association of candidate with BMI in black South Africans. We focused on 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: We genotyped 44 SNPs within the six candidate genes that included known BMI associated SNPs and tagSNPs based on linkage disequilibrium in an African populations for *FTO*, *LEP* and *NPY2R*. To assess population substructure, we included 18 ancestry informative markers. Weight, height, sex, sex-specific pubertal stage and exact age collected during adolescence (13 years of age) 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); *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 *FTO* and *MC4R* as potential genetic markers of obesity risk in South Africans. The association with two SNPs in the 3' UTR 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 to individuals 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've 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 BMI estimated at 40-70%[2].

The study of syndromic obesity and genome-wide association studies (GWAS) of common measures of adiposity (e.g. 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 1,188 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 characterized 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 intake[11] 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 2 diabetes cases and 68% of hypertensive disease were attributable to a BMI ≥ 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 wellbeing of children that were born in the metropolitan area of Soweto, Johannesburg, during a six-week period in 1990. Participants in Bt20 were enrolled at birth and detailed information has been collected from the participants and their caregivers over the past twenty 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 Wise 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 (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 standard deviation which is 0.47. The common standard deviation 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 kg) divided by the square of the height (measured in meters) 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 were 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 data[17] 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.

SNP selection

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 algorithm[20] was used to select tagSNPs in a multimarker approach at $r^2 > 0.80$ and with a minor allele frequency (MAF) > 0.05 among publicly available African data (YRI) from the HapMap dataset (release 24). The Illumina Assay Design Tool (Illumina, San Diego, CA) 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, San Diego, CA, USA). Quality control was performed on all raw genotype data according to the supplier's specifications using BeadStudio v2.0 software (Illumina, San Diego, CA) (summarized in[21]). Standard quality control filters were applied to the data[22]: minor allele frequency (MAF) >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 visualized using Haploview v4.2[23], implementing the confidence interval 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 (from 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(BMI). Exact tests of HWE were also performed. Because of its skewed distribution in this study group, BMI is summarized as median values as well as inter-quartile ranges (IQR), with lower quartile and upper quartiles reported.

Furthermore, each genotype was modeled additively as the number of minor alleles present. Because we modelled log(BMI), the modeled 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 summarized 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 modeled the interaction between sex and each variant and summarized the significant results.

Correcting for multiple testing

Correcting for multiple testing is a contentious issue, and some approaches (such as Bonferroni correction) are considered over-conservative 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 linkage disequilibrium, 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). Females had a higher median BMI (19.4 [IQR = 17.8-22.3]) than males (18.0 [IQR = 16.6-19.7]; P< 0.0001). BMI increased across the five pubertal stages in both males (*P*< 0.0324) and females (*P*<0.0001), but at significantly different rates. Exact age (days) correlated positively with BMI in males (*P*=0.027) but not females (*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(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 (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 (i.e. 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 to 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 to 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 whilst 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 females, each G-allele of the *NPY2R* rs11099992 SNP predicted a decrease in BMI by 3.7% (*P*-value=0.003) while in males the effect was not significant (*P*=0.523). In males, each A-allele of the *MC4R* rs12970134 SNP predicted an increase in BMI of 5.2% (*P*=0.011) while in females 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 compete 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 less than ten individuals each in the categories of one risk allele (seven individuals) and eight risk alleles (one 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 to 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 inter-individual 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 to individuals with two risk alleles.

We provide preliminary evidence that SNPs in the 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 population[34,36] and adult Samoans[37] have also shown obesity to be correlated to variants of the *LEP* 3' hypervariable region (HVR). The 3'HVR is a $(TTTC)_n$ tetranucleotide repeat located approximately 2,000bp 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 localization and message stability.[38].

Variants in *FTO* (the only gene consistently associated with obesity) have been associated with adiposity measures in Europeans,[4,39] Asians[40,41] and African Americans[7], with the strongest effect seen with SNP rs9939609. Our study showed a modest association 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 linkage disequilibrium 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 males (P=0.011). It has been observed that SNPs in the *NPY2R* gene have effects on obesity in males 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 females only [45]. Furthermore, mutations in the *MC4R* gene that have been associated with morbid obesity have stronger effects in females than males[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 less than 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 highlights 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, eighteen 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 only 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 characterization 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].

LEGENDS TO FIGURES:

Figure 1: Observed cumulative effect of high risk alleles on 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 the interquartile range of BMI.

TABLES:

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 13 years of age. Significant *P*-values are shown in bold.

					Frequency							
			RefSNP alleles*		Genotype			Allele		*P*-values		
Gene	dbSNP ID	n	A	B	AA	AB	BB	Minor	Major	Unadjusted	Adjusted**	HWE
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
LEP	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
	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
	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
LEPR	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
MC4R	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
NPY2R	rs2880416	985	C	G	0.00	0.08	0.92	0.04	0.96	0.094	0.186	1.000
	rs2342676	953	A	G	0.08	0.41	0.51	0.29	0.71	0.697	0.942	0.937
	rs12649641	988	A	C	0.48	0.43	0.08	0.30	0.70	0.601	0.786	0.255
	rs11099992	964	A	G	0.58	0.36	0.06	0.24	0.76	0.085	0.073	1.000
	rs33977152	986	A	G	0.01	0.09	0.90	0.05	0.95	0.508	0.301	0.181
	rs12507396	982	A	T	0.95	0.05	0.00	0.02	0.98	0.858	0.957	0.417
	rs6857530	979	A	G	0.05	0.36	0.58	0.23	0.77	0.666	0.707	0.790
	rs10461238	971	C	G	0.32	0.49	0.19	0.43	0.57	0.640	0.430	0.948
	rs2342674	985	A	G	0.00	0.13	0.87	0.07	0.93	0.530	0.476	0.308
	rs1047214	977	T	C	0.88	0.12	0.01	0.06	0.94	0.926	0.919	0.581
	rs2880415	982	A	G	0.55	0.40	0.05	0.25	0.75	0.458	0.292	0.061
POMC	rs6713532	970	T	C	0.29	0.49	0.22	0.46	0.54	0.368	0.270	0.796
	rs7565877	972	A	G	0.37	0.49	0.14	0.39	0.61	0.944	0.967	0.742

	rs7565427	987	A	G	0.01	0.20	0.79	0.11	0.89	0.633	0.889	0.418

* RefSNP allele according to NCBI dbSNP Build 134
** *P*-values are shown adjusted for age, sex and sex-specific pubertal stage

Table 2: Summary BMI statistics for SNPs that were individually associated with BMI

		Typed	Modeled			Median (IQR) of BMI per genotype			
Gene	**SNP**			**MA**	**MAF**	**mm***	**mM***	**MM***	**Effect**
FTO	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
LEP	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
LEP	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
MC4R	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

MA = minor allele; MAF= minor allele frequency; IQR=interquartile range;

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

Table 3: **SNPs individually associated with log(BMI) and three confounders (sex, pubertal stage and age). The model explains 16.4% of the covariation in log(BMI)**

Gene	SNP	MA	Effect	*P*-value
FTO	rs17817449	G	2.2	0.007
LEP	rs10954174	A	-3.2	0.005
LEP	rs6966536	G	2.7	0.020
MC4R	rs17782313	C	-2.0	0.028

MA=minor allele

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).

ACKNOWLEDGEMENTS

The authors would like to 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.

COMPETING INTERESTS

None declared.

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).

REFERENCES

1. 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.

2. 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.

3. 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.

4. 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.

5. 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.

6. 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.

7. 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.

8. 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 doi: db09-1252.

9. 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.

10. Teo YY, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. Nat Rev Genet 2010;11:149-60.

11. 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.

12. 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.

13. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16:1215.

14. Stettler N, Iotova V. Early growth patterns and long-term obesity risk. Curr Opin Clin Nutr Metab Care 2010;13:294-9.

15. 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. American Journal of Human Biology 1998;10:495-504.

16. Norris SA, Richter LM. Validation of Tanner pubertal self-rating to urban black adolescents in South Africa. Journal of Research on Adolescence 2005;15:609-24.

17. 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.

18. Schlebusch CM. Genetic variation in Khoisan-speaking populations from southern Africa. Unpublished PhD thesis. University of the Witwatersrand, Johannesburg, South Africa, 2011.

19. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155:945-59.

20. de Bakker PI, Yelensky R, Pe'er I, et al. Efficiency and power in genetic association studies. Nat Genet 2005;37:1217-23.

21. Tindall EA, Petersen DC, Nikolaysen 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.

22. Clarke GM, Anderson CA, Pettersson FH, et al. Basic statistical analysis in genetic case-control studies. Nature protocols 2011;6:121-33.

23. Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.

24. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225-9.

25. Perneger TV. What's wrong with Bonferroni adjustments. BMJ 1998;316:1236-8.

26. Elsäßer A, Victor A, Hommel G. Multiple Testing in Candidate Gene Situations: A Comparison of Classical, Discrete, and Resampling-Based Procedures. Statistical Applications in Genetics and Molecular Biology 2011;10:e51.

27. Farooqi IS. Genetic aspects of severe childhood obesity. Pediatr Endocrinol Rev 2006;3 Suppl 4:528-36.

28. 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.

29. Mattevi VS, Zembrzuski 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.

30. 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.

31. 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.

32. 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.

33. 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.

34. 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.

35. 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.

36. 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.

37. 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.

38. Mishra PJ, Banerjee D, Bertino JR. MiRSNPs or MiR-polymorphisms, new players in microRNA mediated regulation of the cell: Introducing microRNA pharmacogenomics. Cell Cycle 2008;7:853-8.

39. Wang K, Li WD, Zhang CK, et al. A genome-wide association study on obesity and obesity-related traits. PLoS One 2011;6:e18939.

40. 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.

41. 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.

42. 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.

43. Chen JM, Ferec 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.

44. 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.

45. 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.

46. 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.

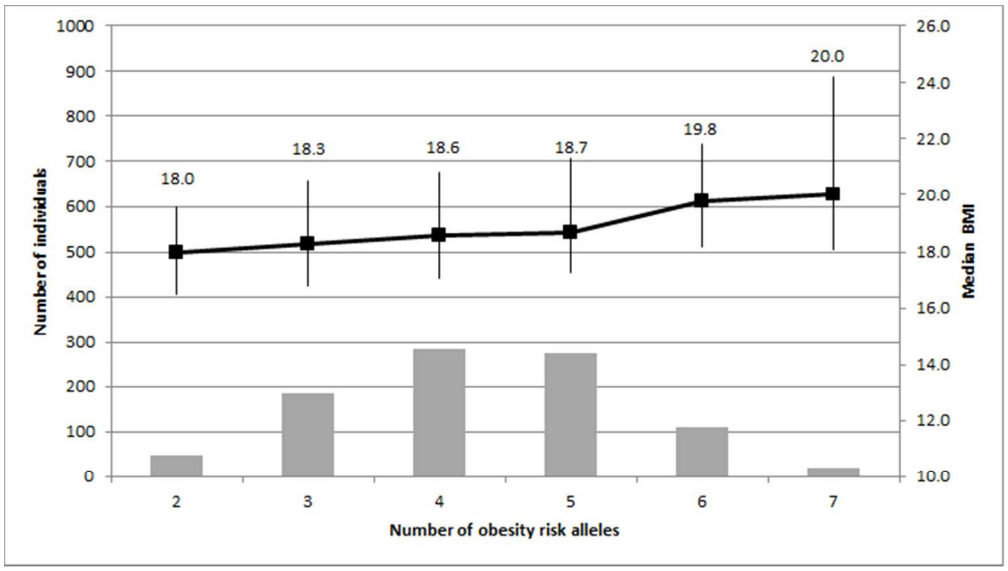
55. Greene CS, Penrod NM, Williams SM, et al. Failure to replicate a genetic association may provide important clues about genetic architecture. PLoS One 2009;4:e5639.

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.



Observed cumulative effect of high risk alleles on 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 the interquartile range of BMI.

Supplementary Figure Legend:

Supplementary Figure 1: STRUCTURE analysis with 18 AIMs illustrating no significant population substructure in the Bt20 cohort. The values of K indicate the number of hypothetical ancestral populations. The clustering at K = 2 distinctly separated the African ((1) South African and (2) YRI) from the Non-African ((3) CHB and (4) CEU) populations. Higher order clustering (K=3 to K=4) was found to resolve the Bt20 cohort from all other populations, indicating that these 18 SNPs have the power to detect major differences in population structure. Population designations: (1) South African cohort; (2) HapMap YRI; (3) HapMap CHB; (4) HapMap CEU.

Supplementary Figure 2: LD across *LEP* in a South African black population. The top section of the figure illustrates the gene and represents the location of the SNPs in *LEP* that were genotyped in this study. The lower section of the figure shows the output of Haploview – each square (with D' values depicted within the box) represents a pair-wise LD relationship between two SNPs.

Supplementary Tables:

Supplementary table 1: Summary of SNP information genotyped in candidate genes

Gene	SNP	Genomic location *	Chrom	RefSNP Alleles **	Ancestral allele **	BMI assoc SNP	TagSNP	AIM
FTO	rs9939973	53800568	16	A/G	G	x		
	rs9940128	53800754		A/G	G	x		
	rs1421085	53800954		C/T	T	x		
	rs1121980	53809247		C/T	T	x		
	rs17817449	53813367		G/T	G	x		
	rs8050136	53816275		A/C	A	x		
	rs9939609	53820527		A/T	A	x		
LEP	rs1349419	127877213	7	A/G	A	x		
	rs12535708	127878098		A/C	C	x		
	rs11770725	127878267		C/T	T	x		
	rs12535747	127878335		A/C	C	x		
	rs2167270	127881349		A/G	G		x	
	rs2278815	127881851		A/G	G		x	
	rs2122627	127883323		C/T	C		x	
	rs4236625	127883695		A/T	T		x	
	rs12706832	127887139		A/G	A		x	
	rs10244329	127888689		A/T	T		x	
	rs7791621	127889696		A/C	A		x	
	rs7795794	127890151		A/G	C		x	
	rs10954173	127891440		A/G	G		x	
	rs3828942	127894305		A/G	G		x	
	rs17151919	127894592		A/G	G		x	
	rs17151922	127895216		G/T	G		x	
	rs6966536	127896059		A/G	A		x	
	rs10954174	127896536		A/G	G		x	

	rs11761556	127897069		A/C	A		x	
LEPR	rs1137100	66036441	1	A/G	A	x		
	rs1137101	66058513		A/G	A	x		
MC4R	rs17782313	57851097	18	C/T	T	x		
	rs12970134	57884750		A/G	G	x		
NPY2R	rs2880416	156124543	4	C/G	C		x	
	rs2342676	156124704		A/G	A	x		
	rs12649641	156125333		A/C	A		x	
	rs11099992	156125510		A/G	A		x	
	rs33977152	156128589		A/G	G		x	
	rs12507396	156129044		A/T	A		x	
	rs6857530	156129154		A/G	A		x	
	rs10461238	156132216		C/G	G	x		
	rs2342674	156135250		A/G	G		x	
	rs1047214	156135676		C/T	T	x		
	rs2880415	156136027		A/G	A		x	
POMC	rs6713532	25384833	2	C/T	C		x	
	rs7565427	25385638		A/G	G		x	
	rs7565877	25386064		A/G	A		x	
Ancestry informative markers	rs723854	192511012	1	C/G	G			x
	rs1876482	17362568	2	C/T	C			x
	rs952718	215888624	2	A/C	A			x
	rs1344870	21307401	3	A/C	A			x
	rs720096	179551071	4	C/G	C			x
	rs1363448	140783596	5	C/T	T			x
	rs217538	108483470	6	C/G	C			x
	rs65264	28545611	7	C/T	T			x
	rs679047	12883664	9	A/T	T			x
	rs2077559	36014850	9	C/T	T			x
	rs714857	15974389	11	C/T	T			x
	rs953386	110943692	13	A/G	A			x
	rs722869	97277005	14	C/G	C			x

	rs735612	34076642	15	G/T	G			x
	rs2089740	36310531	15	G/T	G			x
	rs1823718	74147244	15	C/T	T			x
	rs1858465	51142920	17	A/T	A			x
	rs2112527	9603751	19	A/G	G			x

* Chromosomal location of SNPs obtained from NCBI genome build 37.1 (hg19).
** RefSNP allele and ancestral alleles according to NCBI dbSNP Build 134.

Supplementary table 2: Allele- and genotype frequencies of AIMs in the South African cohort

				Frequency				
		RefSNP alleles*		Genotype			Allele	
SNP	n	A	B	AA	AB	BB	A	B
rs723854	974	C	G	0.05	0.31	0.65	0.20	0.80
rs1876482	983	T	C	0.00	0.00	1.00	0.00	1.00
rs952718	973	A	C	0.49	0.42	0.09	0.70	0.30
rs1344870	977	A	C	0.90	0.10	0.00	0.95	0.05
rs720096	970	C	G	0.64	0.32	0.05	0.79	0.21
rs1363448	981	T	C	1.00	0.00	0.00	1.00	0.00
rs217538	984	C	G	0.00	0.03	0.97	0.02	0.98
rs65264	982	T	C	0.64	0.33	0.03	0.80	0.20
rs679047	978	A	T	0.08	0.43	0.50	0.29	0.71
rs2077559	972	T	C	0.44	0.45	0.10	0.67	0.33
rs953386	988	A	G	0.21	0.50	0.30	0.46	0.54
rs722869	986	C	G	0.79	0.20	0.01	0.89	0.11
rs735612	979	T	G	0.21	0.51	0.28	0.47	0.53
rs2089740	970	T	G	0.11	0.46	0.42	0.35	0.65
rs1823718	972	T	C	0.62	0.35	0.03	0.79	0.21
rs1858465	972	A	T	0.02	0.26	0.72	0.15	0.85
rs2112527	984	A	G	0.43	0.47	0.11	0.66	0.34

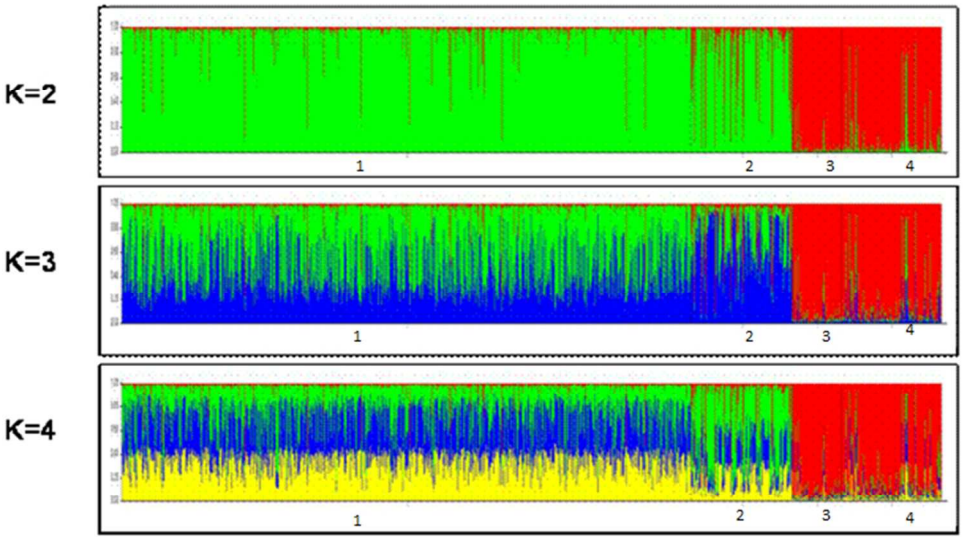
* RefSNP allele and ancestral alleles according to NCBI dbSNP Build 134.

Supplementary table 3: Allele frequencies of AIMs in the South African cohort (SAB) and three HapMap populations

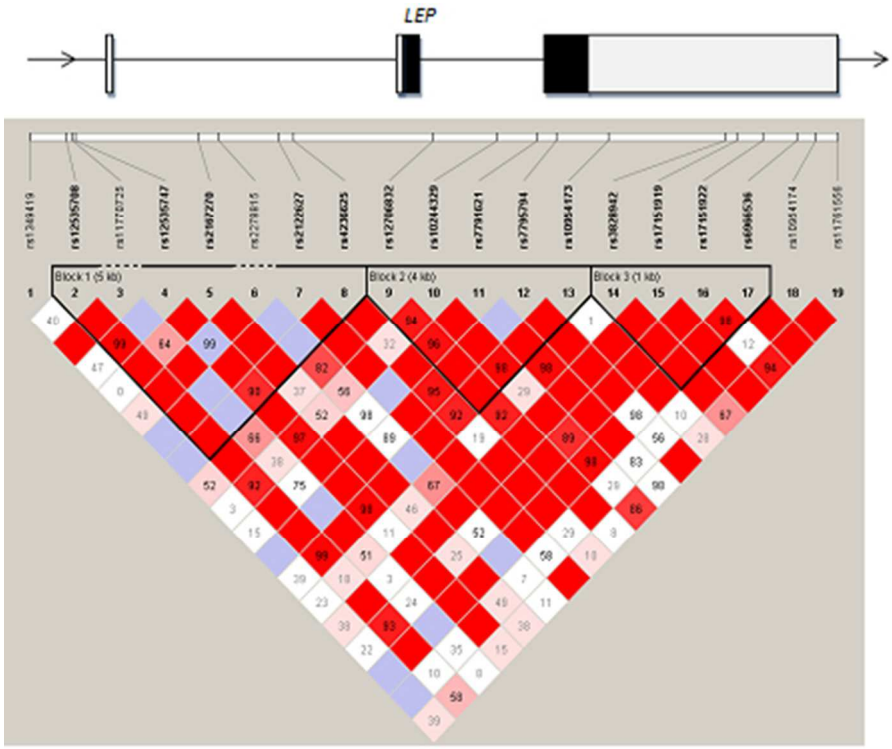
			Allele Frequencies							
	RefSNP alleles*		YRI		CEU		CHB		SAB	
SNP ID	A	B	A	B	A	B	A	B	A	B
rs723854	C	G	0.19	0.81	0.62	0.38	0.62	0.38	0.20	0.80
rs1876482	T	C	0.00	1.00	0.03	0.98	0.70	0.30	0.00	1.00
rs952718	A	C	0.74	0.26	0.11	0.89	0.04	0.96	0.70	0.30
rs1344870	A	C	0.95	0.05	0.98	0.02	0.73	0.27	0.95	0.05
rs720096	C	G	0.04	0.96	0.53	0.48	0.01	0.99	0.79	0.21
rs1363448	T	C	0.86	0.14	0.42	0.58	0.35	0.65	1.00	0.00
rs217538	C	G	0.97	0.03	0.68	0.32	0.41	0.59	0.02	0.98
rs65264	T	C	0.87	0.13	0.67	0.33	0.42	0.58	0.80	0.20
rs679047	A	T	0.22	0.78	0.08	0.92	0.09	0.91	0.29	0.71
rs2077559	T	C	0.77	0.23	0.88	0.12	0.73	0.27	0.67	0.33
rs953386	A	G	0.46	0.54	0.15	0.85	0.09	0.91	0.46	0.54
rs722869	C	G	0.93	0.07	0.88	0.12	0.14	0.86	0.89	0.11
rs735612	T	G	0.42	0.58	0.61	0.39	0.98	0.02	0.47	0.53
rs2089740	T	G	0.29	0.71	0.06	0.94	0.32	0.69	0.35	0.65
rs1823718	T	C	0.77	0.23	0.54	0.46	0.99	0.01	0.79	0.21
rs1858465	A	T	0.90	0.10	0.18	0.82	0.34	0.66	0.15	0.85
rs2112527	A	G	0.55	0.45	0.93	0.07	0.97	0.03	0.66	0.34

* RefSNP allele and ancestral alleles according to NCBI dbSNP Build 134

YRI – HapMap Yoruba from Nigeria; CEU – HapMap Utah residents with European ancestry ; CHB – HapMap Han Chinese population; SAB – South African black population (Bt20).



STRUCTURE analysis with 18 AIMs illustrating no significant population substructure in the Bt20 cohort. The values of K indicate the number of hypothetical ancestral populations. The clustering at K = 2 distinctly separated the African ((1) South African and (2) YRI) from the Non-African ((3) CHB and (4) CEU) populations. Higher order clustering (K=3 to K=4) was found to resolve the Bt20 cohort from all other populations, indicating that these 18 SNPs have the power to detect major differences in population structure. Population designations: (1) South African cohort; (2) HapMap YRI; (3) HapMap CHB; (4) HapMap CEU.



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], Punita Pitamber[1], Shane A Norris[5], Michèle Ramsay[1]

[1] Division Human Genetics, School of Pathology, Faculty of Health Sciences, National Health Laboratory Service & University of the Witwatersrand, Johannesburg, 2000, South Africa

[2] Current address: Wits Bioinformatics, School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg, 2000, South Africa

[3] Department of Chemical Pathology, School of Pathology, Faculty of Health Sciences, National Health Laboratory Service & University of the Witwatersrand, Johannesburg, 2000, South Africa

[4] Biostatistics Unit, Medical Research Council, Cape Town, 8001, South Africa and Department of Statistics, University of Western Cape, Bellville, South Africa

[5] MRC/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, 2000, South Africa

***Corresponding author:** Zané Lombard; Postal address: Wits Bioinformatics, Private Bag 3, 2050 Wits, Johannesburg, South Africa; Email: zane.lombard@wits.ac.za; Telephone: +27117176393; Fax: +27867656304

Key words: Obesity, Body Mass Index, Genetic Association Studies, *FTO*, *LEP*, *MC4R*

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 BMI many of which are found within genes involved in appetite regulation. Currently, genetic association data for obesity is lacking in Africans - a single GWAS and a few replication studies have been published in West Africans, but none have been performed in a South African population.
Objectives: To assess the association of candidate with BMI in black South Africans. We focused on 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: We genotyped 44 SNPs within the six candidate genes that included known BMI associated SNPs and tagSNPs based on linkage disequilibrium in an African populations for FTO, LEP and NPY2R. To assess population substructure, we included 18 ancestry informative markers. Weight, height, sex, sex-specific pubertal stage and exact age collected during adolescence (13 years of age) 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); 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 FTO and MC4R as potential genetic markers of obesity risk in South Africans. The association with two SNPs in the 3' UTR 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 to individuals 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've 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 BMI estimated at 40-70%[2].

The study of syndromic obesity and genome-wide association studies (GWAS) of common measures of adiposity (e.g. 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 1,188 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 characterized 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 intake[11] 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 2 diabetes cases and 68% of hypertensive disease were attributable to a BMI ≥ 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 wellbeing of children that were born in the metropolitan area of Soweto, Johannesburg, during a six-week period in 1990. Participants in Bt20 were enrolled at birth and detailed information has been collected from the participants and their caregivers over the past twenty 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 Wise 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 (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 standard deviation which is 0.47. The common standard deviation 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 kg) divided by the square of the height (measured in meters) 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 were 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 data[17] 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.

SNP selection

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 algorithm[20] was used to select tagSNPs in a multimarker approach at $r^2 > 0.80$ and with a minor allele frequency (MAF) > 0.05 among publicly available African data (YRI) from the HapMap dataset (release 24). The Illumina Assay Design Tool (Illumina, San Diego, CA) 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, San Diego, CA, USA). Quality control was performed on all raw genotype data according to the supplier's specifications using BeadStudio v2.0 software (Illumina, San Diego, CA) (summarized in[21]). Standard quality control filters were applied to the data[22]: minor allele frequency (MAF) >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 visualized using Haploview v4.2[23], implementing the confidence interval 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 (from 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(BMI). Exact tests of HWE were also performed. Because of its skewed distribution in this study group, BMI is summarized as median values as well as inter-quartile ranges (IQR), with lower quartile and upper quartiles reported.

Furthermore, each genotype was modeled additively as the number of minor alleles present. Because we modelled log(BMI), the modeled 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 summarized 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 modeled the interaction between sex and each variant and summarized the significant results.

Correcting for multiple testing

Correcting for multiple testing is a contentious issue, and some approaches (such as Bonferroni correction) are considered over-conservative 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 linkage disequilibrium, 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). Females had a higher median BMI (19.4 [IQR = 17.8-22.3]) than males (18.0 [IQR = 16.6-19.7]; P< 0.0001). BMI increased across the five pubertal stages in both males (*P*< 0.0324) and females (*P*<0.0001), but at significantly different rates. Exact age (days) correlated positively with BMI in males (*P*=0.027) but not females (*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(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 (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 (i.e. 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 to 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 to 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 whilst 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 females, each G-allele of the *NPY2R* rs11099992 SNP predicted a decrease in BMI by 3.7% (*P*-value=0.003) while in males the effect was not significant (*P*=0.523). In males, each A-allele of the *MC4R* rs12970134 SNP predicted an increase in BMI of 5.2% (*P*=0.011) while in females 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 compete 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 less than ten individuals each in the categories of one risk allele (seven individuals) and eight risk alleles (one 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 to 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 inter-individual 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 to individuals with two risk alleles.

We provide preliminary evidence that SNPs in the 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 population[34,36] and adult Samoans[37] have also shown obesity to be correlated to variants of the *LEP* 3' hypervariable region (HVR). The 3'HVR is a $(TTTC)_n$ tetranucleotide repeat located approximately 2,000bp 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 localization and message stability.[38].

Variants in *FTO* (the only gene consistently associated with obesity) have been associated with adiposity measures in Europeans,[4,39] Asians[40,41] and African Americans[7], with the strongest effect seen with SNP rs9939609. Our study showed a modest association 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 linkage disequilibrium 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 males (P=0.011). It has been observed that SNPs in the *NPY2R* gene have effects on obesity in males 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 females only [45]. Furthermore, mutations in the *MC4R* gene that have been associated with morbid obesity have stronger effects in females than males[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 less than 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 highlights 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, eighteen 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 only 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 characterization 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].

LEGENDS TO FIGURES:

Figure 1: Observed cumulative effect of high risk alleles on 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 the interquartile range of BMI.

TABLES:

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 13 years of age. Significant *P*-values are shown in bold.

					Frequency							
			RefSNP alleles*		Genotype			Allele		*P*-values		
Gene	**dbSNP ID**	**n**	**A**	**B**	**AA**	**AB**	**BB**	**Minor**	**Major**	**Unadjusted**	**Adjusted****	**HWE**
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
LEP	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
	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
	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
LEPR	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
MC4R	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
NPY2R	rs2880416	985	C	G	0.00	0.08	0.92	0.04	0.96	0.094	0.186	1.000
	rs2342676	953	A	G	0.08	0.41	0.51	0.29	0.71	0.697	0.942	0.937
	rs12649641	988	A	C	0.48	0.43	0.08	0.30	0.70	0.601	0.786	0.255
	rs11099992	964	A	G	0.58	0.36	0.06	0.24	0.76	0.085	0.073	1.000
	rs33977152	986	A	G	0.01	0.09	0.90	0.05	0.95	0.508	0.301	0.181
	rs12507396	982	A	T	0.95	0.05	0.00	0.02	0.98	0.858	0.957	0.417
	rs6857530	979	A	G	0.05	0.36	0.58	0.23	0.77	0.666	0.707	0.790
	rs10461238	971	C	G	0.32	0.49	0.19	0.43	0.57	0.640	0.430	0.948
	rs2342674	985	A	G	0.00	0.13	0.87	0.07	0.93	0.530	0.476	0.308
	rs1047214	977	T	C	0.88	0.12	0.01	0.06	0.94	0.926	0.919	0.581
	rs2880415	982	A	G	0.55	0.40	0.05	0.25	0.75	0.458	0.292	0.061
POMC	rs6713532	970	T	C	0.29	0.49	0.22	0.46	0.54	0.368	0.270	0.796
	rs7565877	972	A	G	0.37	0.49	0.14	0.39	0.61	0.944	0.967	0.742

	rs7565427	987	A	G	0.01	0.20	0.79	0.11	0.89	0.633	0.889	0.418

* RefSNP allele according to NCBI dbSNP Build 134
** *P*-values are shown adjusted for age, sex and sex-specific pubertal stage

Table 2: Summary BMI statistics for SNPs that were individually associated with BMI

		Typed	Modeled			Median (IQR) of BMI per genotype			
Gene	**SNP**			**MA**	**MAF**	**mm***	**mM***	**MM***	**Effect**
FTO	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
LEP	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
LEP	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
MC4R	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

MA = minor allele; MAF= minor allele frequency; IQR=interquartile range;

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

Table 3: **SNPs individually associated with log(BMI) and three confounders (sex, pubertal stage and age). The model explains 16.4% of the covariation in log(BMI)**

Gene	SNP	MA	Effect	*P*-value
FTO	rs17817449	G	2.2	0.007
LEP	rs10954174	A	-3.2	0.005
LEP	rs6966536	G	2.7	0.020
MC4R	rs17782313	C	-2.0	0.028

MA=minor allele

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).

ACKNOWLEDGEMENTS

The authors would like to 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.

COMPETING INTERESTS

None declared.

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).

REFERENCES

1. 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.

2. 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.

3. 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.

4. 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.

5. 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.

6. 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.

7. 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.

8. 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 doi: db09-1252.

9. 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.

10. Teo YY, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. Nat Rev Genet 2010;11:149-60.

11. 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.

12. 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.

13. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16:1215.

14. Stettler N, Iotova V. Early growth patterns and long-term obesity risk. Curr Opin Clin Nutr Metab Care 2010;13:294-9.

15. 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. American Journal of Human Biology 1998;10:495-504.

16. Norris SA, Richter LM. Validation of Tanner pubertal self-rating to urban black adolescents in South Africa. Journal of Research on Adolescence 2005;15:609-24.

17. 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.

18. Schlebusch CM. Genetic variation in Khoisan-speaking populations from southern Africa. Unpublished PhD thesis. University of the Witwatersrand, Johannesburg, South Africa, 2011.

19. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155:945-59.

20. de Bakker PI, Yelensky R, Pe'er I, et al. Efficiency and power in genetic association studies. Nat Genet 2005;37:1217-23.

21. Tindall EA, Petersen DC, Nikolaysen 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.

22. Clarke GM, Anderson CA, Pettersson FH, et al. Basic statistical analysis in genetic case-control studies. Nature protocols 2011;6:121-33.

23. Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.

24. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. Science 2002;296:2225-9.

25. Perneger TV. What's wrong with Bonferroni adjustments. BMJ 1998;316:1236-8.

26. Elsäßer A, Victor A, Hommel G. Multiple Testing in Candidate Gene Situations: A Comparison of Classical, Discrete, and Resampling-Based Procedures. Statistical Applications in Genetics and Molecular Biology 2011;10:e51.

27. Farooqi IS. Genetic aspects of severe childhood obesity. Pediatr Endocrinol Rev 2006;3 Suppl 4:528-36.

28. 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.

29. Mattevi VS, Zembrzuski 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.

30. 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.

31. 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.

32. 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.

33. 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.

34. 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.

35. 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.

36. 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.

37. 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.

38. Mishra PJ, Banerjee D, Bertino JR. MiRSNPs or MiR-polymorphisms, new players in microRNA mediated regulation of the cell: Introducing microRNA pharmacogenomics. Cell Cycle 2008;7:853-8.

39. Wang K, Li WD, Zhang CK, et al. A genome-wide association study on obesity and obesity-related traits. PLoS One 2011;6:e18939.

40. 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.

41. 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.

42. 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.

43. Chen JM, Ferec 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.

44. 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.

45. 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.

46. 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.

STROBE 2007 (v4) Statement—Checklist of items that should be included in reports of *cross-sectional studies*

Section/Topic	Item #	Recommendation	Reported on page #
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	6
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-7
Bias	9	Describe any efforts to address potential sources of bias	7,9
Study size	10	Explain how the study size was arrived at	6
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6-7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8
		(*b*) Describe any methods used to examine subgroups and interactions	8-9
		(*c*) Explain how missing data were addressed	7
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	NA
		(*e*) Describe any sensitivity analyses	NA
Results			

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	10
		(b) Give reasons for non-participation at each stage	NA
		(c) Consider use of a flow diagram	NA
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	10, Table 2
		(b) Indicate number of participants with missing data for each variable of interest	Table 1, 2
Outcome data	15*	Report numbers of outcome events or summary measures	10
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	Table 1, 2
		(*b*) Report category boundaries when continuous variables were categorized	Table 2
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	10-12
Discussion			
Key results	18	Summarise key results with reference to study objectives	13
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13-16
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	13-16
Generalisability	21	Discuss the generalisability (external validity) of the study results	15-16
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	24

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.