



# Genetic Association Study of Adiposity and Melanocortin-4 Receptor (*MC4R*) Common Variants: Replication and Functional Characterization of Non-Coding Regions

Daniel S. Evans<sup>1,9</sup>, Melissa A. Calton<sup>2,9</sup>, Mee J. Kim<sup>3</sup>, Pui-Yan Kwok<sup>4</sup>, Iva Miljkovic<sup>5</sup>, Tamara Harris<sup>6</sup>, Annemarie Koster<sup>6</sup>, Yongmei Liu<sup>7</sup>, Gregory J. Tranah<sup>1</sup>, Nadav Ahituv<sup>3</sup>, Wen-Chi Hsueh<sup>2</sup>, Christian Vaisse<sup>2\*</sup>

**1** California Pacific Medical Center Research Institute, San Francisco, California, United States of America, **2** Diabetes Center and Department of Medicine, University of California, San Francisco, California, United States of America, **3** Department of Bioengineering and Therapeutic Sciences and Institute for Human Genetics, University of California, San Francisco, California, United States of America, **4** Cardiovascular Research Institute, Institute for Human Genetics, and Department of Dermatology, University of California, San Francisco, California, United States of America, **5** Department of Epidemiology, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America, **6** National Institute on Aging, Bethesda, Maryland, United States of America, **7** Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest University, Winston-Salem, North Carolina, United States of America

## Abstract

Common genetic variants 3' of *MC4R* within two large linkage disequilibrium (LD) blocks spanning 288 kb have been associated with common and rare forms of obesity. This large association region has not been refined and the relevant DNA segments within the association region have not been identified. In this study, we investigated whether common variants in the *MC4R* gene region were associated with adiposity-related traits in a biracial population-based study. Single nucleotide polymorphisms (SNPs) in the *MC4R* region were genotyped with a custom array and a genome-wide array and associations between SNPs and five adiposity-related traits were determined using race-stratified linear regression. Previously reported associations between lower BMI and the minor alleles of rs2229616/Val103Ile and rs52820871/Ile251Leu were replicated in white female participants. Among white participants, rs11152221 in a proximal 3' LD block (closer to *MC4R*) was significantly associated with multiple adiposity traits, but SNPs in a distal 3' LD block (farther from *MC4R*) were not. In a case-control study of severe obesity, rs11152221 was significantly associated. The association results directed our follow-up studies to the proximal LD block downstream of *MC4R*. By considering nucleotide conservation, the significance of association, and proximity to the *MC4R* gene, we identified a candidate *MC4R* regulatory region. This candidate region was sequenced in 20 individuals from a study of severe obesity in an attempt to identify additional variants, and the candidate region was tested for enhancer activity using *in vivo* enhancer assays in zebrafish and mice. Novel variants were not identified by sequencing and the candidate region did not drive reporter gene expression in zebrafish or mice. The identification of a putative insulator in this region could help to explain the challenges faced in this study and others to link SNPs associated with adiposity to altered *MC4R* expression.

**Citation:** Evans DS, Calton MA, Kim MJ, Kwok P-Y, Miljkovic I, et al. (2014) Genetic Association Study of Adiposity and Melanocortin-4 Receptor (*MC4R*) Common Variants: Replication and Functional Characterization of Non-Coding Regions. *PLoS ONE* 9(5): e96805. doi:10.1371/journal.pone.0096805

**Editor:** Marta Letizia Hribal, University of Catanzaro Magna Graecia, Italy

**Received:** December 24, 2013; **Accepted:** April 11, 2014; **Published:** May 12, 2014

**Copyright:** © 2014 Evans et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging (NIA), and by NIA contracts N01-AG-6-2101, N01-AG-6-2103, N01-AG-6-2106, NIA grant R01-AG028050, NINR grant R01-NR012459, and NIDDK grants 1R01DK090382 and R01DK060540. The genome-wide association study was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. DSE was supported by NIH training grant T32 DK007418. MAC was supported by NIH training grant T32 GM007175, Pharmaceutical Sciences and Pharmacogenomics. MJK was supported in part by NIH training grant T32 GM007175 and the Amgen Research Excellence in Bioengineering and Therapeutic Sciences Fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

**†** These authors contributed equally to this work.

\* E-mail: [vaisse@diabetes.ucsf.edu](mailto:vaisse@diabetes.ucsf.edu).

## Introduction

Obesity has been increasing in prevalence worldwide and is a risk factor for many poor health outcomes [1,2]. Obesity results from the interaction between genetic and non-genetic factors. Studies of severe and common forms of obesity have demonstrated

that the Melanocortin-4 Receptor (*MC4R*) is an important regulator of obesity and adiposity [3]. *MC4R* belongs to a family of seven trans-membrane G-protein-coupled receptors (GPCR) and is expressed at low levels in hypothalamic nuclei involved in the regulation of food intake [4]. *MC4R* regulates food intake by integrating a satiety signal provided by its agonist  $\alpha$ -MSH and an

orexigenic signal provided by its antagonist Agouti-related protein (AGRP) [5,6]. These ligands are expressed in distinct neuronal populations of the arcuate nucleus of the hypothalamus and are regulated by the adipocyte-secreted hormone, leptin, to control food intake and maintain long-term energy homeostasis [7]. Mice lacking both alleles of *mc4r* (*mc4r*  $-/-$  mice) develop a maturity onset hyperphagic obesity syndrome by 10 weeks of age, while mice heterozygous for a *mc4r* deletion (*mc4r*  $+/-$  mice) show an intermediate obese phenotype [6].

Genetic variants within the *MC4R* coding region have been found to be associated with severe and common forms of obesity [3]. Rare mutations in the *MC4R* coding region account for a significant number of severe obesity cases [3,8,9,10]. More common, but still quite rare (minor allele frequency (MAF) < 5% in most populations) *MC4R* non-synonymous SNPs (nsSNPs) (rs2229616/Val103Ile and rs52820871/Ile251Leu) have been reproducibly associated with a protective effect from severe and common forms of obesity [11,12,13,14,15,16]. Functional studies indicate that the 251Leu allele increases MC4R basal activity and the 103Ile allele decreases MC4R antagonist potency while also increasing MC4R agonist potency [17]. These biochemical effects result in elevated MC4R function, which is consistent with the association between these variants and a lower body weight.

In addition to variants within the *MC4R* coding region, common variants outside of the coding region have been associated with common and severe forms of obesity. Meta-analyses of genome-wide association studies (GWAS) conducted in Caucasians have identified common variants in two large linkage disequilibrium (LD) blocks 3' of the *MC4R* coding region that are associated with adiposity and anthropometric traits [3,18,19,20,21,22,23]. The most significant association signal in the proximal 3' LD block (closer to *MC4R*) is rs17700633, and in the distal 3' LD block (farther from *MC4R*) is rs17782313 [19]. Multiple SNPs in high LD with rs17782313 (rs17700144, HapMap Phase 3 CEU  $r^2 = 0.83$ ; rs12970134, HapMap Phase 3 CEU  $r^2 = 0.84$ ) have also been associated with adiposity-related traits [18,22,23]. In addition to common forms of obesity, rs17782313 and rs17700144 have also been associated with early-onset severe obesity [16,24].

While recent GWAS efforts in populations of European descent have been very successful at identifying the 288 kb association region that encompasses both LD blocks located 21 kb 3' of *MC4R*, there has been little success at refining this association region or assigning a functional role to non-coding variants in these regions. Conditional analysis indicates that at least a small degree of dependence might exist between SNPs in the proximal and distal LD blocks, even though LD would suggest otherwise [19]. It has been argued that synthetic associations with *MC4R* nsSNPs are not likely to underlie the associations between SNPs 3' of *MC4R* and obesity [25]. Thus, the identity of causal variants that might underlie common SNP associations in the *MC4R* non-coding region remains unknown.

Refining the large association region 3' of *MC4R* and evaluating the biological role of DNA in this region could aid in the identification of causal risk alleles near *MC4R*. To this end, we investigated the association of common SNPs within and surrounding the *MC4R* gene with multiple adiposity-related traits in the Health ABC study, a biracial population-based cohort. SNP associations with severe obesity in an independent study were also examined. By considering nucleotide conservation, the significance of association, and proximity to the *MC4R* gene, we identified a candidate *MC4R* regulatory region. This candidate region was sequenced in 20 individuals from a study of severe obesity in an attempt to identify additional variants, and the candidate region

was tested for enhancer activity using *in vivo* enhancer assays in zebrafish and mice. Data from the ENCODE project [26] were used to gain further insight into the biological function of this DNA region.

## Materials and Methods

### Ethics Statement

The Health ABC study protocol was approved by the institutional review boards at the University of Pittsburgh and the University of Tennessee, Memphis, and written informed consent was obtained from all participants. The severe obesity study protocol was approved by the UCSF Committee on Human Research, and written informed consent was obtained from all participants. All animal work protocols were approved by the UCSF Institutional Animal Care and Use Committee (Approval Number: AN100466-01A).

### Study populations

The Health, Aging, and Body Composition (Health ABC) study is a population-based prospective study of 3,075 men and women (48.5% male; 41.7% African American) aged 70 to 79 years at baseline, residing in Pittsburgh, PA and Memphis, TN. All participants were well-functioning at the time of entry into the study; they reported no difficulty walking a quarter of a mile or walking up 10 steps without resting. Data used in the present study were obtained from the baseline examination, during 1997–1998. Adiposity-related measures in these participants have been described previously [27]. Briefly, percentage of total body fat was assessed by DXA and abdominal visceral fat area (visceral fat) and abdominal subcutaneous fat area (subcutaneous fat) ( $\text{cm}^2$ ) were assessed using the computed tomography scan image measured at the L4-L5 disk space. Serum leptin was measured by radioimmunoassay (Linco Research Inc, St Charles, MO) in the morning from participants who fasted overnight.

Severely obese participants were selected from an ongoing UCSF study, as previously described [28].

### TagSNP selection

In order to capture the genetic variation in the *MC4R* coding region and flanking non-coding DNA, SNPs were selected using HapMap Phase 2 (release 20) project data ([www.hapmap.org/](http://www.hapmap.org/)). By considering conservation between human and mouse genomes, non-coding regions up to 32.5 kb from the 5' end (base pair 56,223,516, Reference assembly, NCBI genome build 36.3) and 21.3 kb from the 3' end (base pair 56,168,229) of *MC4R* were used to select tagSNPs, using the program Tagger [29]. TagSNPs were chosen based on having known or predicted alterations of gene or protein function and having strong LD ( $r^2 \geq 0.8$ ) with other SNPs. In order to select tagSNPs appropriate for use with a biracial cohort such as the Health ABC Study, tagSNPs were first selected using CEU (Caucasian/European) SNP genotypes, and these tagSNPs were then added to the tagSNP selection using YRI (Yoruban) SNP genotypes. Thirty-seven SNPs in or near *MC4R* were selected for genotyping using the Illumina Golden Gate Assay.

### Genotyping and quality control

All tagSNPs except for rs17782313 were genotyped using the Illumina Golden Gate Assay (Illumina, San Diego, CA, USA) from DNA isolated from participants of the Health ABC Study and the UCSF obesity study. Rs17782313 was genotyped with a Taqman assay from ABI using a stock kit. All samples that produced a genotype for rs17782313 were used to analyze that SNP. For the

Illumina Golden Gate Assay, samples with a missing call rate greater than 10% were excluded. Five percent of the DNA samples were genotyped in duplicate to estimate genotyping error rate, and SNPs with more than one discrepancy between duplicate samples (0.7% error rate) were excluded from analysis; none of the SNPs were excluded based on this criterion. SNPs were also removed if the HWE *P*-value in white participants was  $<0.001$  (Bonferroni corrected *P*-value of 0.05 corrected for 38 SNPs); none of the SNPs were excluded based on this criterion. 13 SNPs in Health ABC white participants and 5 SNPs in Health ABC black participants were excluded from analysis based on MAF  $<0.05$ .

To extend our SNP coverage into both LD blocks located 3' to *MC4R*, we examined genome-wide SNP genotypes that were previously assayed in Health ABC participants using the Illumina Human 1M-Duo array. Genotypes were called using Illumina BeadStudio. Samples were excluded from the dataset for the reasons of sample failure, genotypic sex mismatch, and first-degree relative of an included individual based on genotype data. SNPs were excluded if the call rate was  $<97\%$ , HWE *P*-value  $<10^{-6}$ , or MAF  $<0.01$ .

Individual level genetic data from the genome-wide SNP dataset from Health ABC is available through controlled access from dbGaP (dbGaP Study accession phs000169.v1.p1). Individual level Health ABC genetic data from the *MC4R* candidate gene SNP genotype data is available through Health ABC's coordinating center website (<http://www.keepertrack.ucsf.edu/>).

### Sequencing putative *MC4R* regulatory region

Among the severely obese patients from the UCSF study that were homozygous for the rs11152221 C allele (major allele) or the T allele (minor allele), ten CC homozygotes and ten TT homozygotes were randomly selected for sequencing using the R function sample, which employs the Mersenne-Twister pseudo-random number generator. The forward primer 5'-GGCTGCTGCTGGGGTCAACA-3' and reverse primer 5'-ACCCACCATCCCATCTGTGCGA-3' were used in PCR to amplify the 1.25 kb region of interest (NCBI build 36: chromosome 18: 56,168,229–56,169,479). The sequencing reaction was performed with the BigDye terminator kit (Applied Biosystems, Foster City, CA) under the standard manufacturer's conditions. Sequencing was performed on an ABI PRISM 3700 automated DNA sequencer (Applied Biosystems). Sanger sequencing data can be fully reconstructed from the description in the results.

### Cloning, transgenics, and enhancer assays

These studies were carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health and all efforts were made to minimize suffering. The protocols were approved by the UCSF Institutional Animal Care and Use Committee (Approval Number: AN100466-01A).

We PCR amplified the same DNA region that was sequenced (NCBI build 36: chromosome 18: 56,168,229–56,169,479) using the forward primer 5'-AACTCGAGGGCTGCTGCTGGGGTCAACA-3' and reverse primer: 5'-GGCTCGAGACCCACCATCCCATCTGTGCGA-3' from genomic DNA of a patient from the UCSF study who was homozygous for the rs11152221 C allele. The sequence was sequence verified for having the proper allele. For zebrafish transgenics, the PCR product was cut with XhoI and ligated into the E1B-GFP-Tol2 enhancer assay vector [30]. The plasmid DNA was cleaned for endotoxins using the Qiagen EndoFree Plasmid Midi kit.

Zebrafish injections were performed using standard procedures as previously described [31]. The injection mix contained 1  $\mu$ L

125 ng/ $\mu$ L endotoxin-free plasmid DNA, 1  $\mu$ L 175 ng/ $\mu$ L Tol2 RNA, 2  $\mu$ L sterile water, and 1  $\mu$ L 2% Phenol red. Embryo injections were performed four independent times and at least 50 embryos were injected each time. Zebrafish were examined at 24 hrs, 48 hrs, and 72 hrs post-fertilization for GFP expression, and at least 85 healthy surviving embryos were analyzed at each time point. Imaging of zebrafish was done using a Lumar V12 Stereomicroscope (Carl Zeiss) with Axio Vision Rel. 4.4 (Carl Zeiss).

For mouse transgenics, the candidate enhancer region was PCR amplified from human genomic DNA, as described above, and digested with XhoI and SmaI. The region was cloned into the Hsp68-promoter-LacZ reporter vector [32]. Transgenic mouse embryos were generated through Cyagen Biosciences, Inc. using standard procedures [33] and embryos at day 15 were stained for LacZ expression as in [34]. The embryos were then processed and imbedded in paraffin, sectioned (7  $\mu$ m thickness) and counterstained with neutral fast red for visualization by light microscopy (Carl Zeiss) with Axio Vision Rel. 4.4.

### Statistical Analysis

For the BMI, percent body fat, and leptin outcomes, the effects of age, sex, recruitment site, prevalent diabetes status, weekly levels of calculated physical activity, smoking and drinking habits, and education levels were adjusted for in the regression analysis. Leptin, visceral fat and subcutaneous fat were transformed by taking the square-root to produce normal distributions. To identify associations with leptin independent of percent body fat, leptin was subsequently analyzed adjusting for the percent body fat. To adjust for overall body size, baseline height and weight were included as covariates for abdominal visceral and subcutaneous fat.

For association analysis using tagSNPs from the Illumina Golden Gate assay, the appropriate mode of inheritance was determined by examining parameter estimates from a genotypic 2df test. All SNPs are modeled with an additive mode of inheritance, except for rs11152221, which was modeled as dominant, and rs1943225, which was modeled as recessive. To avoid population stratification, all analyses were performed in whites and blacks separately. The first two principal components determined from principal component analysis (PCA) using genome-wide SNP data did not impact tagSNP association effect estimates. *P*-values less than 0.05 were deemed significant for sex-interactions and associations between replication SNPs (rs52820871, rs2229616, and rs17782313) and adiposity-related traits. Significance of associations using tagSNPs was corrected for multiple hypothesis testing by obtaining empirical *P*-values through permutation testing by the min*P* procedure using 100,000 replicates [35]. Logistic regression models that adjusted for the effects of age, sex, recruitment site, prevalent diabetes status, weekly levels of calculated physical activity, smoking and drinking habits, and education levels were used to determine the association between rs11152221 and obesity using cases (BMI  $\geq 30$ ) and controls (BMI  $< 30$ ) identified from participants in the Health ABC study.

Association between rs11152221 and obesity was also examined using cases (BMI  $\geq 30$ ) from self-identified Caucasian subjects from the UCSF study of severe obesity and controls (BMI  $< 30$ ) from white participants of the Health ABC study. Logistic regression models included sex as a covariate.

For the SNP association analysis of the extended *MC4R* region, directly genotyped and imputed SNPs on chromosome 18 from position 55,850,000–56,230,000 (NCBI build 36) were selected from genome-wide genotyped and imputed SNP data that were

previously obtained in Health ABC participants. Genotype imputation was performed using MACH (v. 1.0.16) with the HapMap CEU Phase 2 release 22 build 36 haplotypes in Health ABC white participants and a 1:1 mixture of HapMap CEU:YRI Phase 2 release 22 build 36 haplotypes in Health ABC black participants. Imputed SNPs with an MAF < 0.05 or an observed:expected variance ratio < 0.3 were removed. To adjust the significance threshold for multiple testing of 304 genotyped and imputed SNPs in the *MC4R* region, a Bonferroni correction was applied using the number of independent SNPs, which was determined using Tagger to select tagSNPs in the region of interest from HapMap phase 2 release 24 CEU and YRI genotypes (SNPs with MAF < 0.05 excluded, pair-wise tagging with an  $r^2$  threshold of 0.8) [29]. There were 55 tagSNPs for CEU genotypes and 119 tagSNPs for YRI genotypes, resulting in significance thresholds of  $9 \times 10^{-4}$  and  $4 \times 10^{-4}$  in Health ABC white and black participants, respectively. Selection of tagSNPs using HapMap phase 3 ASW (individuals of African ancestry in Southwest USA) genotypes yielded 97 tagSNPs, but in an effort to be conservative, the significance threshold based on YRI genotypes was adopted for Health ABC black participants. The first two principal components determined from PCA of genome-wide SNP data were included in regression models. Covariates used were the same as in the analysis of tagSNPs from the Illumina Golden Gate assay. LD of imputed allele dosages was visualized by constructing a correlation matrix (Pearson's  $r^2$ ) of the imputed allele dosage data for SNPs in the region of interest that passed QC, then plotting the correlation matrix as a heatmap using the LDheatmap R package. Nucleotide conservation between human and mouse genomes was obtained using the VISTA browser [36]. All regression analyses were performed using R software ([www.r-project.org](http://www.r-project.org)).

Publicly available expression quantitative trait loci (eQTL) data from CEU and YRI HapMap3 lymphoblastoid cell lines were accessed using Genevar [37,38]. On chromosome 18, position 55,850,000–56,230,000 (NCBI build 36), SNP associations with *MC4R* expression were determined using Spearman's rank correlation coefficient and association significance was assessed using a *t*-statistic and a *t*-distribution with  $n-2$  degrees of freedom. The eQTL significance level was adjusted for multiple testing using the Bonferroni correction by dividing 0.05 by the number of independent SNPs in the chromosome 18 region between positions 55,850,000–56,230,000. The number of independent SNPs was determined by selecting tagSNPs from HapMap3 (release 27) CEU and YRI genotypes in the region using Tagger (tagSNP  $r^2 \geq 0.8$ , 45 CEU and 106 YRI tagSNPs selected).

## Results

Association analysis was performed using adiposity-related traits (BMI, plasma leptin levels, percentage total body fat mass, abdominal subcutaneous fat, and abdominal visceral fat) measured in participants from the population-based Health ABC study. The study population contained white and black participants, and all outcomes and covariates except for physical activity estimates were significantly different by race (Table 1).

### Replication analysis

We first attempted to replicate the previous finding that the minor alleles of two nsSNPs (rs52820871 and rs2229616) have a protective effect on obesity that is stronger in females [12,39]. While the association between the minor alleles of these two nsSNPs and lower BMI did not reach statistical significance among Health ABC white participants (Table 2), the association did reach statistical significance among white females (rs52820871

$P$ -value = 0.03, rs2229616  $P$ -value = 0.01) (Table 3). Rs2229616 did show a significant interaction with sex ( $P$ -value<sub>INT</sub> = 0.002), but rs52820871 did not. The minor alleles of the two nsSNPs were also associated with lower BMI in black participants and black females, but the association did not reach statistical significance (Table S1 and Table S2). Among black participants in sex-stratified analysis, the only significant association between either of the nsSNPs and adiposity-related traits was between the 251Leu allele of rs52820871 and higher abdominal visceral fat in black females ( $\beta \pm SE = 3.16 \pm 1.39$ ,  $P$ -value = 0.02) (Table S1 and Table S2).

The previously identified top hit in the distal LD block 3' of *MC4R*, rs17782313, was associated with higher BMI in all race and sex stratified analyses, but the association failed to reach statistical significance (Table 2, Table 3, and Table S2). Our study did not have sufficient power (Power = 0.25, 2-sided  $\alpha = 0.05$ ) to detect the reported effect size (0.22 BMI units) for the association between rs17782313 and BMI [19]. The SNP rs17782313 was significantly associated with lower abdominal visceral fat in white and white male participants ( $P$ -value = 0.006 and 0.004, respectively) (Table 2 and Table 3). In addition, rs17782313 was significantly associated with higher leptin levels in black and black female participants ( $P$ -value = 0.03 and 0.05, respectively) (Table S2).

### Genetic associations in the *MC4R* gene region

TagSNPs were selected to capture common genetic variation within the *MC4R* gene and the surrounding non-coding DNA. Two SNPs, rs11152221 and rs1943225, remained significantly associated with adiposity-related traits after correction for multiple testing in race stratified or race and sex stratified analysis (Figure 1, Figure S1). Rs11152221 was in high LD with the previously reported top association signal in the proximal LD block, rs17700633 (HapMap CEU  $r^2 = 0.79$ ). After correction for multiple testing, rs11152221 coded with a dominant mode of inheritance was significantly associated with higher BMI ( $P$ -value<sub>nom</sub> =  $2 \times 10^{-5}$ ,  $P$ -value<sub>emp</sub> =  $5 \times 10^{-4}$ ), percentage body fat ( $P$ -value<sub>nom</sub> = 0.003,  $P$ -value<sub>emp</sub> = 0.05), and leptin levels ( $P$ -value<sub>nom</sub> =  $4 \times 10^{-4}$ ,  $P$ -value<sub>emp</sub> = 0.008) among white participants, but not black participants (Table 2, Table S2, and Figure 1). Conditional analysis demonstrated that the association between rs11152221 and BMI was not dependent on rs17782313 or the two nsSNPs within the *MC4R* gene (rs52820871 and rs2229616) (data not shown). While rs11152221 was more significantly associated with BMI in white females ( $P$ -value<sub>nom</sub> =  $2 \times 10^{-4}$ ,  $P$ -value<sub>emp</sub> = 0.004) than white males ( $P$ -value<sub>nom</sub> = 0.02,  $P$ -value<sub>emp</sub> > 0.05), the effect estimates were not significantly different as evidenced by the lack of significance of a sex interaction term and the overlap of the 95% confidence intervals for the estimates in these two groups (Table 3). Coded with an additive mode of inheritance, rs11152221 was significantly associated in white participants with BMI ( $\beta \pm SE = 0.53 \pm 0.15$ ,  $P$ -value =  $5 \times 10^{-4}$ ), percentage body fat ( $\beta \pm SE = 0.42 \pm 0.19$ ,  $P$ -value = 0.03), and leptin levels ( $\beta \pm SE = 0.13 \pm 0.05$ ,  $P$ -value = 0.005), and the association with BMI passed multiple test correction.

Association between rs11152221 and obesity was tested among the 296 obesity cases and 1303 obesity controls selected from Health ABC white participants. The rs11152221 T allele was significantly associated with increased odds of obesity (Dominant coding: OR = 1.76, 95% CI = 1.34–2.30,  $P$ -value =  $4 \times 10^{-5}$ , Additive coding: OR = 1.46, 95% CI = 1.20–1.78,  $P$ -value =  $2 \times 10^{-4}$ , Table S3).

The association between rs11152221 and obesity was also examined using cases selected from a cohort of severely obese

**Table 1.** Health ABC participant characteristics by race.

Characteristic	White		Black		P<0.05
	n	Mean ±SD or n (%)	n	Mean ±SD or n (%)	
Age (y)	1655	73.79±2.86	1175	73.45±2.89	*
Sex (males)		873 (53%)		501 (43%)	*
Height (m)	1655	1.67±0.09	1175	1.65±0.09	*
Weight (kg)	1655	74.23±14.38	1175	78.26±15.65	*
BMI (kg/m <sup>2</sup> )	1655	26.57±4.15	1175	28.70±5.46	*
Body Fat (%)	1592	34.75±7.19	1136	35.51±8.64	*
Leptin (ng/ml)	1635	12.81±11.58	1152	17.31±13.92	*
VAT (cm <sup>2</sup> )	1590	152.01±69.03	1128	130.38±61.59	*
SAT (cm <sup>2</sup> )	1558	265.77±102.72	1074	315.60±139.75	*
Physical Activity (Kcal/kg/week)	1655	84.02±62.83	1175	79.91±75.94	
Prevalent Diabetes	-	319 (20%)	-	347 (31%)	*
Education					*
Not high school graduate	-	203 (12%)	-	507 (43%)	
High school graduate	-	564 (34%)	-	360 (31%)	
Postsecondary education	-	886 (54%)	-	303 (26%)	
Smoking habits					*
Never smoker	-	709 (43%)	-	518 (44%)	
Former smoker	-	836 (51%)	-	465 (40%)	
Current smoker	-	108 (7%)	-	189 (16%)	
Drinking habits					*
Never drinker	-	405 (25%)	-	378 (32%)	
Former drinker	-	278 (17%)	-	342 (29%)	
Current drinker	-	964 (59%)	-	450 (38%)	

VAT = abdominal visceral adipose tissue. SAT = abdominal subcutaneous adipose tissue.

\**p*<0.05 between races by t-test for continuous traits and by Chi-squared test for categorical traits.

doi:10.1371/journal.pone.0096805.t001

patients from a UCSF study and controls selected from non-obese white participants from the Health ABC study (Table S4). Fewer covariates were included in the analysis of obesity association using cases from the UCSF severe obesity study compared to cases from the Health ABC study, resulting in fewer controls being excluded due to incomplete covariate information. The rs11152221 T allele was significantly associated with increased odds of obesity (OR = 1.28, 95% CI = 1.00–1.64, *P*-value = 0.05, Table S5) under an additive mode of inheritance.

A SNP (rs1943225) located in the non-coding DNA 5' to *MC4R* remained significantly associated with adiposity-related traits after correction for multiple testing (Figure S1). This SNP was significantly associated with higher leptin levels adjusted for percentage body fat in white females (*P*-value<sub>nom</sub> =  $2 \times 10^{-4}$ , *P*-value<sub>emp</sub> = 0.005) (Table 3). This association was nominally significant in white participants (*P*-value<sub>nom</sub> = 0.004), but did not remain significant after correction for multiple testing (Table 2). In white females, rs1943225 was associated with lower percentage body fat but higher plasma leptin levels (Table 3). As percentage body fat and leptin levels are strongly correlated among Health ABC white participants (Pearson's *r* = 0.77), it is expected that an association with lower body fat could mask an association with higher leptin levels. Thus, when leptin levels are adjusted for percentage body fat, the association between rs1943225 and higher leptin levels was observed to be highly significant in white females (Table 3). This association was only observed when

rs1943225 was coded with a recessive mode of inheritance. When the six genotyped tagSNPs (rs1943217, rs1943218, rs8093815, rs9965495, rs17066879, and rs17773774) that were in LD with rs1943225 (*r*<sup>2</sup>>0.6) were also coded with a recessive mode of inheritance, three of them remained significantly associated after multiple test correction with leptin levels adjusted for percentage body fat in white females (rs1943217, rs1943218, and rs9965495) (Figure S1).

The location of the most significantly associated SNP, rs11152221, in the proximal 3' LD block previously identified in GWAS meta-analyses of anthropometric traits compelled us to examine SNPs in the proximal and distal 3' LD blocks using genome-wide genotype data in Health ABC participants. Multiple SNPs in the proximal but not the distal LD block were associated with BMI in white participants after multiple test correction (Figure 2). Two SNPs in the distal LD block (rs17782313 and rs12970134) that were previously reported to be significantly associated with BMI were not associated with BMI (*P*-value > 0.05) in Health ABC white participants (Figure 2), confirming the association results found from Taqman genotyping of rs17782313 in Health ABC participants (Table 2). We next attempted to take advantage of the shorter haplotypes present in Health ABC black participants in this region, however, no SNPs were significantly associated with BMI after multiple test correction in this population (Figure 3).

**Table 2.** SNP associations with adiposity-related traits in Health ABC white participants.

SNP <sup>a</sup>	Alleles	CAF	BMI (n = 1599–1682)		% Body Fat (n = 1538–1616)		Leptin (n = 1580–1663)		Leptin <sup>b</sup> (n = 1519–1597)		VAT (n = 1538–1620)		SAT (n = 1506–1586)	
			β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>
rs52820871	A/C	0.01	-1.07±0.75	NS	-0.28±0.96	NS	-0.16±0.24	NS	-0.14±0.17	NS	-0.45±0.36	NS	0.29±0.31	NS
rs2229616	G/A	0.02	-0.46±0.55	NS	-0.65±0.68	NS	-0.15±0.17	NS	-0.04±0.12	NS	0.09±0.26	NS	-0.14±0.22	NS
rs17782313	T/C	0.22	0.24±0.17	NS	0.25±0.21	NS	0.05±0.05	NS	0.01±0.04	NS	-0.22±0.08	0.006	0.03±0.07	NS
rs11152221	C/T	0.31	0.84±0.20	2 × 10 <sup>-5*</sup>	0.75±0.25	0.003*	0.22±0.06	4 × 10 <sup>-4*</sup>	0.07±0.05	NS	0.07±0.10	NS	0.02±0.08	NS
rs1943225	T/G	0.23	0.61±0.45	NS	-0.08±0.56	NS	0.29±0.14	0.04	0.29±0.10	0.004	-0.27±0.22	NS	-0.21±0.18	NS

Alleles listed as reference allele/coded allele. CAF = coded allele frequency. VAT = abdominal visceral adipose tissue. SAT = abdominal subcutaneous adipose tissue.

<sup>a</sup>Additive coding = rs52820871/1251L, rs2229616/V103I, and rs17782313; dominant coding = rs11152221; recessive coding = rs1943225.

<sup>b</sup>Leptin outcome adjusted for percentage of body fat.

<sup>c</sup>Unadjusted P-value. P-values >0.05 shown as NS (not significant).

\*Empirical P-value ≤ 0.05.

doi:10.1371/journal.pone.0096805.t002

**Functional characterization of non-coding regions**

We investigated whether non-coding variants located 3' of *MC4R* were associated with *MC4R* expression by using publicly available eQTL data from HapMap CEU and YRI lymphoblastoid cell lines. In the *MC4R* 288 kb gene region encompassing the proximal and distal LD blocks, no SNPs were significantly associated with *MC4R* expression after correction for multiple testing in CEU or YRI cell lines (Figure 2 and Figure 3). While lymphoblastoid cell lines are convenient for high-throughput gene expression studies, these cell lines might not accurately reflect gene expression in hypothalamic tissue. Thus, we selected a DNA region near rs11152221 in the proximal LD block to search for potential causal variants by sequencing and subsequent *in vivo* enhancer assays.

DNA regions in the human genome near rs11152221 are conserved with the mouse genome (Figure 2 and Figure 3). This SNP is 704 bp 3' to a 357 bp stretch of DNA that is 70.6% conserved with mouse DNA and 1091 bp 3' to a 156 bp DNA region that is 70.4% conserved. These areas of conservation, in their entirety, were sequenced in twenty severely obese patients (ten rs11152221CC homozygotes and ten rs11152221 TT homozygotes) from an ongoing UCSF study (see Materials and Methods). The twenty obese patients were all female Caucasians without diabetes, and patient characteristics did not differ by rs11152221 genotype (Table S6). Given that the rs11152221 T allele frequency was 0.31, we hypothesized that potential causal variants tagged by rs11152221 would also be common and could be detected in ten homozygous patients. However, we were unable to detect any novel homozygous variants in this region in our small sample set of severely obese patients homozygous for the rs11152221 T allele. One patient homozygous for the rs11152221 C allele (major allele) was homozygous for the minor allele of rs11872889, and one patient homozygous for the rs11152221 T allele (minor allele) was heterozygous for the minor allele of rs72973926. No association was found between rs11872889 and BMI in Health ABC white participants (n = 1613, β±SE = -0.31±0.31, P-value = 0.31, coded allele (A) frequency = 0.08, MACH r<sup>2</sup> imputation quality = 0.70). The SNP rs11872889 was not imputed in Health ABC black participants. The SNP rs72973926 was not imputed in the Health ABC cohort, and its allele frequency from the 1000 genomes project was reported for the YRI population (C allele frequency = 0.06), but not for populations of European descent.

We next tested the conserved DNA region of interest for enhancer activity using both zebrafish and mouse enhancer assays. In zebrafish embryos examined up to 72 hours post-fertilization, the 1.25 kb conserved region amplified from a patient homozygous for the rs11152221 major allele (C) that was associated with lower values of adiposity-related traits was negative for reporter expression in the midbrain (24 hpf: midbrain expression in 1.6% of 126 examined embryos; 48 hpf: 1% of 99 embryos; 72 hpf: 0% of 87 embryos). The same DNA region amplified from the same patient homozygous for the rs11152221 C allele was also tested for enhancer activity in the mouse. At E14-15, the earliest age in which *MC4R* expression has been detected [40], all eight mouse embryos that carried the transgene (as determined by PCR) displayed minimal levels of reporter expression in the brain. The three embryos with detectable reporter expression did not display a consistent expression pattern.

Examination of data from the Encyclopedia of DNA Elements (ENCODE) Project [26] indicated that the 1.25 kb DNA region of interest contains a possible insulator element (Figure 4). ENCODE's assignment of a chromatin state as an insulator is based on a Hidden Markov Model applied to ChIP-seq data, including

**Table 3.** SNP associations with adiposity-related traits in Health ABC white female and white male participants.

White female participants																				
SNP <sup>a</sup>	Alleles	CAF	BMI (n = 748-786)			% Body Fat (n = 722-756)			Leptin (n = 737-775)			Leptin <sup>b</sup> (n = 711-745)			VAT (n = 723-760)			SAT (n = 707-742)		
			β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>
rs52820871	A/C	0.01	-2.48±1.16	0.03	-1.88±1.49	NS	-0.41±0.39	NS	-0.15±0.29	NS	-0.38±0.49	NS	0.26±0.47	NS						
rs2229616	G/A	0.02	-2.27±0.88	0.01	-1.38±1.09	NS	-0.42±0.30	NS	-0.16±0.21	NS	-0.19±0.38	NS	0.34±0.34	NS						
rs17782313	T/C	0.22	0.36±0.26	NS	0.29±0.33	NS	0.06±0.09	NS	-0.01±0.06	NS	-0.13±0.11	NS	-0.04±0.10	NS						
rs11152221	C/T	0.31	1.18±0.31	2×10 <sup>-4*</sup>	0.81±0.39	0.04	0.25±0.11	0.02	0.08±0.08	NS	0.18±0.13	NS	-0.09±0.12	NS						
rs1943225	T/G	0.23	0.69±0.69	NS	-0.83±0.86	NS	0.45±0.23	NS	0.61±0.17	2×10 <sup>-4*</sup>	-0.10±0.29	NS	-0.48±0.26	NS						
White male participants																				
SNP <sup>a</sup>	Alleles	CAF	BMI (n = 851-896)			% Body Fat (n = 816-860)			Leptin (n = 843-888)			Leptin <sup>b</sup> (n = 808-852)			VAT (n = 815-860)			SAT (n = 799-844)		
			β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>	β±SE	P <sup>c</sup>
rs52820871	A/C	0.01	0.13±0.97	NS	0.97±1.23	NS	0.09±0.27	NS	-0.04±0.20	NS	-0.54±0.53	NS	0.46±0.40	NS						
rs2229616	G/A	0.02	1.07±0.67	NS	-0.04±0.86	NS	0.06±0.19	NS	0.08±0.14	NS	0.33±0.37	NS	-0.37±0.29	NS						
rs17782313	T/C	0.22	0.11±0.21	NS	0.20±0.27	NS	0.04±0.06	NS	0.03±0.04	NS	-0.34±0.12	0.004	0.08±0.09	NS						
rs11152221	C/T	0.31	0.57±0.24	0.02	0.68±0.32	0.03	0.19±0.07	0.005	0.07±0.05	NS	-0.01±0.14	NS	0.08±0.11	NS						
rs1943225	T/G	0.23	0.76±0.57	NS	0.84±0.74	NS	0.15±0.16	NS	-0.01±0.12	NS	-0.38±0.32	NS	0.10±0.25	NS						

Alleles listed as reference allele/coded allele. CAF = coded allele frequency. VAT = abdominal visceral adipose tissue. SAT = abdominal subcutaneous adipose tissue.

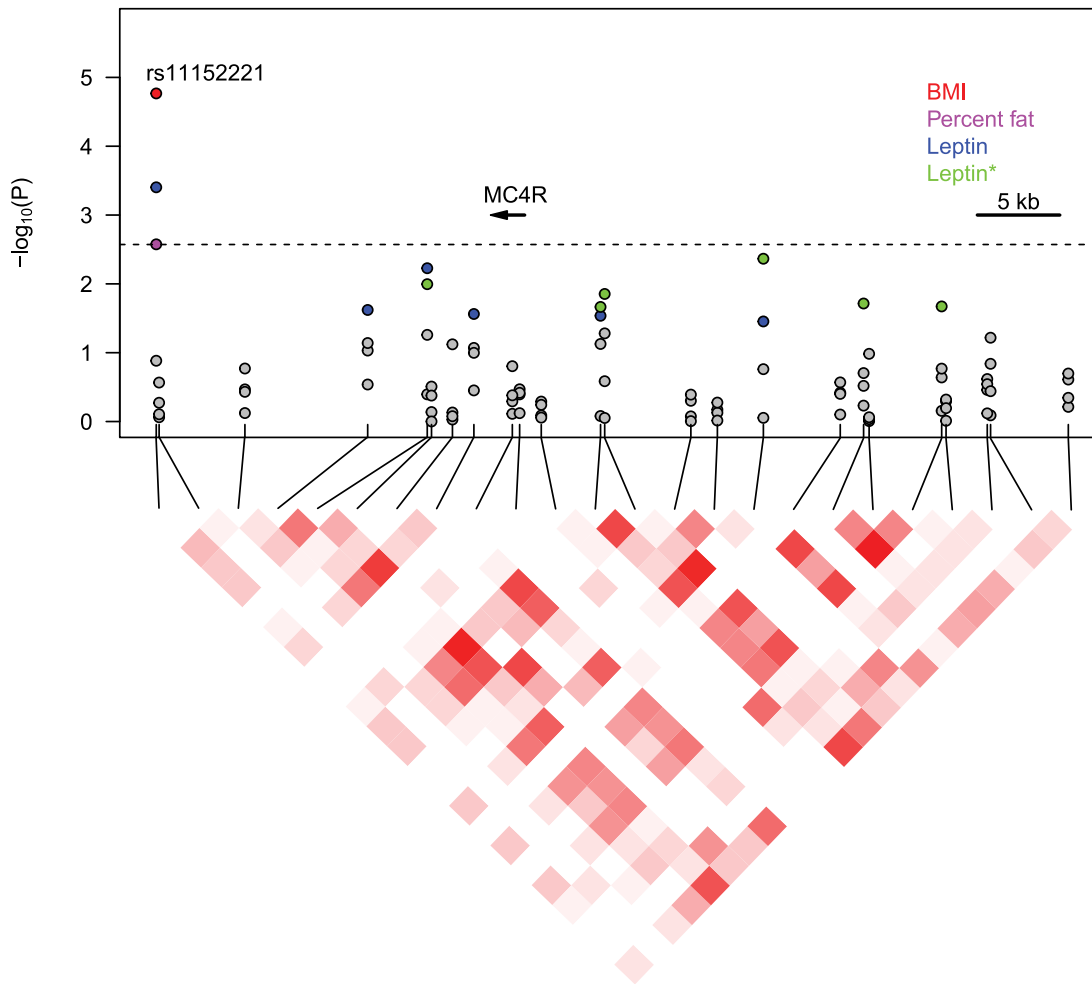
<sup>a</sup>Additive coding = rs52820871/1251L, rs2229616/V103I, and rs17782313; dominant coding = rs11152221; recessive coding = rs1943225.

<sup>b</sup>Leptin outcome adjusted for percentage of body fat.

<sup>c</sup>Unadjusted P-value. P-values >0.05 shown as NS (not significant).

\*Empirical P-value ≤0.05.

doi:10.1371/journal.pone.0096805.t003



**Figure 1. SNP associations in and near *MC4R* with adiposity in white Health ABC participants.** SNP genotypes from custom Illumina Golden Gate array. Gray points indicate association  $P$ -value  $>0.05$ . Non-gray points indicate significant ( $P$ -value  $\leq 0.05$ ) associations with an adiposity trait of the corresponding color in the legend. Leptin\* indicates association  $P$ -value for leptin adjusted for percent body fat. Dashed line indicates cut-off value for empirical  $P$ -value  $\leq 0.05$ . LD heatmap indicates higher  $r^2$  measures with darker red colors. doi:10.1371/journal.pone.0096805.g001

ChIP-seq using an antibody against the CCCTC-binding factor (CTCF), a protein that is known to associate with insulator activity [41,42,43]. ENCODE data also indicated the presence of a second possible insulator element located 200 bp upstream of the *MC4R* promoter (Figure 4). These two insulators could potentially modulate interactions between enhancers and the *MC4R* promoter.

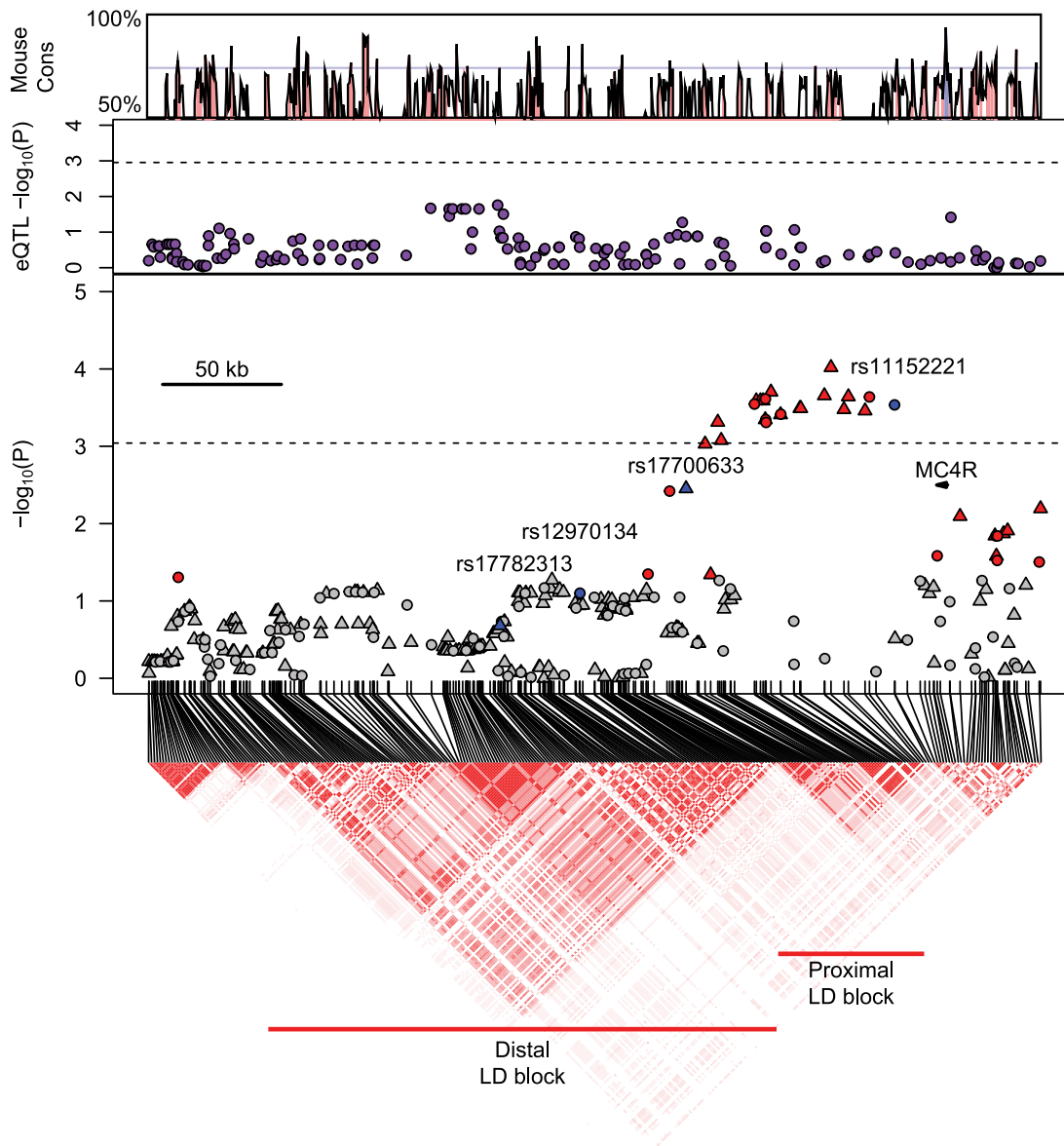
## Discussion

In this study, we examined the association between SNPs in the *MC4R* gene region with multiple measures of adiposity in a biracial study population, the Health ABC study. In addition, we used eQTL data to determine whether SNPs in the region were associated with *MC4R* expression levels, and we functionally characterized a candidate DNA region for enhancer activity *in vivo*. The associations between lower BMI and two rare non-synonymous *MC4R* SNPs replicated in white female Health ABC participants, but the association between BMI and a common SNP discovered through GWAS (rs17782313) in a distal LD block in the 3' *MC4R* non-coding region did not. We further explored SNPs in the 3' *MC4R* non-coding region, and discovered that SNPs in the proximal LD block, but not the distal LD block,

were significantly associated with BMI after correction for multiple testing in white Health ABC participants. Within the proximal LD block, we selected a DNA region to be tested for *in vivo* enhancer activity based on the fact that it contained a SNP (rs11152221) that was significantly associated with adiposity and obesity, it was the closest DNA segment to the *MC4R* gene, and it was highly conserved with the mouse genome. However, this DNA segment failed to demonstrate enhancer activity. ENCODE data suggested that the transcriptional insulator CTCF can bind this DNA segment. Not only could the presence of a potential insulator help to explain the lack of enhancer activity in our assays, but it could also explain why non-coding *MC4R* SNPs that have been consistently associated with anthropometric and adiposity traits fail to be associated with *MC4R* RNA expression in eQTL experiments.

*Cis*-acting regulatory regions include functional elements such as enhancers and insulators [44]. Independent of their orientation and distance from the promoter, enhancers can regulate transcription and are often composed of clusters of transcription factor binding sites [44]. Neither *MC4R* transcriptional regulatory regions nor transcription factors regulating *MC4R* have been identified. Recent GWAS of adiposity-related traits have consis-





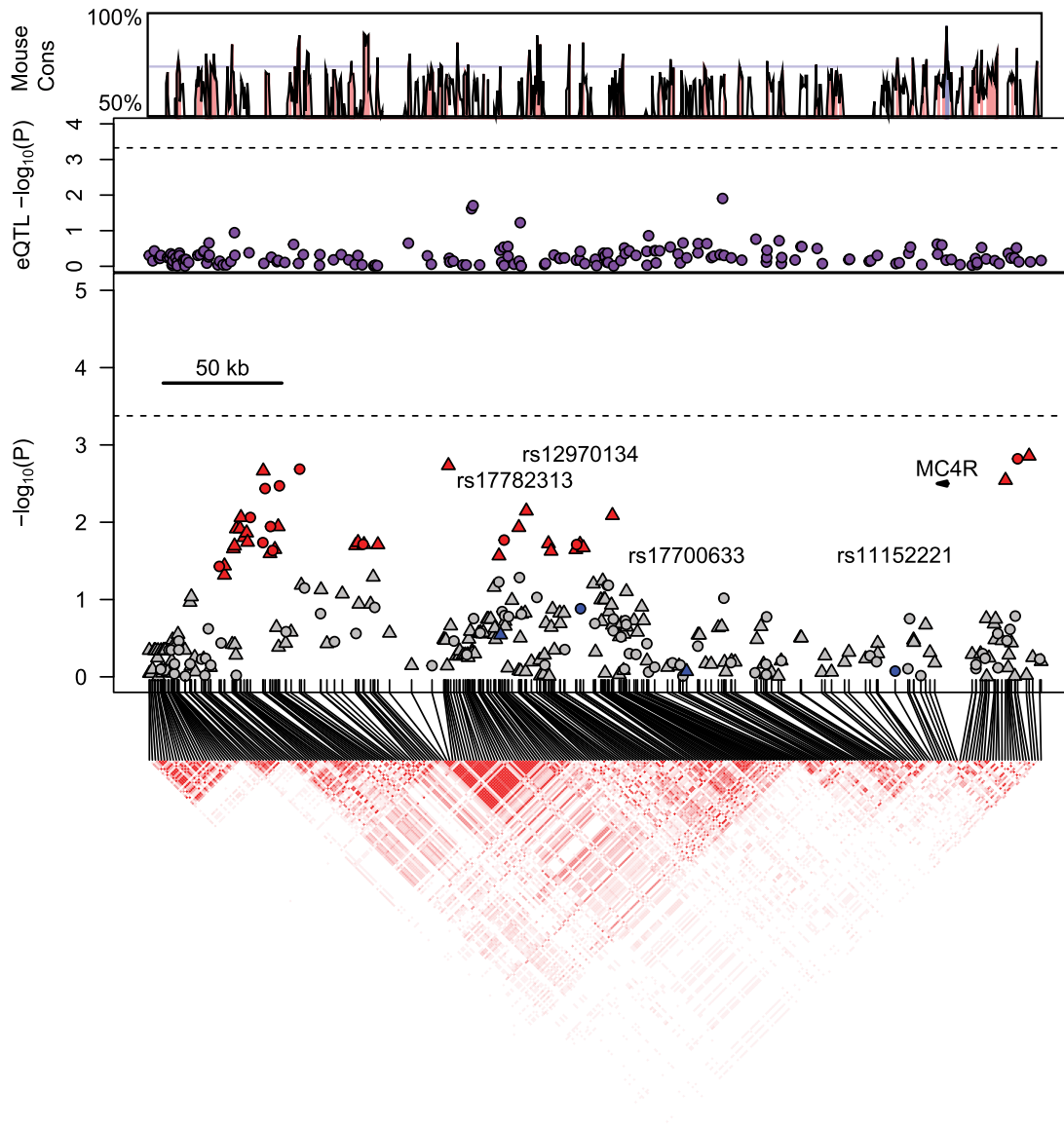
**Figure 2. Association between BMI and genotyped and imputed SNPs in white Health ABC participants.** SNP genotypes from genome-wide Illumina array. In the panel displaying BMI association  $P$ -values, circles mark directly genotyped SNPs and triangles mark imputed SNPs. Gray points indicate association  $P$ -value  $>0.05$ . Red points indicate significant ( $P$ -value  $\leq 0.05$ ) associations with BMI. Anchor SNPs colored in blue. Purple circles mark SNP association with *MC4R* expression in HapMap CEU lymphoblastoid cell lines. In the panels showing trait association and eQTL  $P$ -values, the dashed line indicates cut-off value for Bonferroni-corrected  $P$ -value  $\leq 0.05$ . LD heatmap indicates higher  $r^2$  measures with darker red colors. Nucleotide conservation between the human and mouse is indicated on the top panel of the figure and was obtained using the VISTA browser.

doi:10.1371/journal.pone.0096805.g002

tently identified highly significant SNP associations within two large LD blocks downstream of *MC4R*, highlighting the importance of this non-coding region, but molecular mechanisms for these SNP associations have yet to be identified [3,18,19,20,21,22,23]. SNPs in these non-coding regions could be in high LD with causal variants disrupting functional *MC4R* regulatory elements, but our analysis of eQTL data from HapMap lymphoblastoid cell lines failed to support this hypothesis. It is worth noting that gene expression regulation in lymphoblastoid cell lines is unlikely to accurately reflect what occurs in hypothalamic neurons, which are the relevant cell type. Thus, we took an *in vivo* enhancer assay approach using the mouse and

zebrafish model systems to determine whether DNA surrounding SNPs significantly associated with adiposity can act as enhancers. While the DNA region that we examined did not act as an enhancer in our assays, ENCODE data indicated that the DNA region can bind CTCF. The associated SNP rs11152221 does not overlap with the ENCODE-predicted CTCF binding region and does not directly interrupt a CTCF binding site. Nevertheless, three potential CTCF binding sites are located within 250 bp of rs11152221, supporting CTCF binding to this DNA region (Table S7, Figure S2).

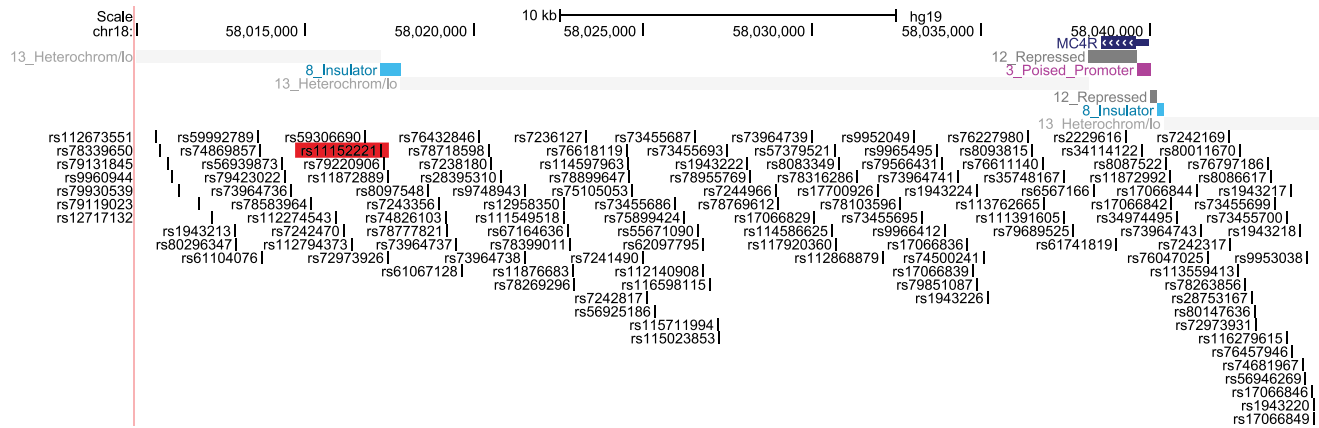
While further work will be needed to experimentally determine whether the DNA region surrounding rs11152221 does in fact



**Figure 3. Association between BMI and genotyped and imputed SNPs in black Health ABC participants.** SNP genotypes from genome-wide Illumina array. Circles mark directly genotyped SNPs and triangles mark imputed SNPs. Gray points indicate association  $P$ -value  $>0.05$ . Red points indicate significant ( $P$ -value  $\leq 0.05$ ) associations with BMI. Anchor SNPs colored in blue. Purple circles mark SNP association with *MC4R* expression in HapMap YRI lymphoblastoid cell lines. In the panels showing trait association and eQTL  $P$ -values, the dashed line indicates cut-off value for Bonferroni-corrected  $P$ -value  $\leq 0.05$ . LD heatmap indicates higher  $r^2$  measures with darker red colors. Nucleotide conservation between the human and mouse is indicated on the top panel of the figure and was obtained using the VISTA browser. doi:10.1371/journal.pone.0096805.g003

bind CTCF, the ENCODE annotation and presence of potential CTCF binding sites lead to various models and testable hypotheses. One possible model invokes CTCF's role as a transcriptional insulator. CTCF binding could create a transcriptional insulator that blocks enhancers from activating the *MC4R* promoter, and genetic variation at the *MC4R* locus could modify the efficiency of CTCF binding in the region. In addition to acting as an insulator, CTCF has also been shown to play a role in transcriptional activation by forming active chromatin hubs through intra-chromosomal interactions [41]. In addition to the downstream *MC4R* DNA region that includes rs11152221 and an ENCODE-predicted CTCF-based insulator, ENCODE also predicts a CTCF-based insulator approximately 200 bp upstream of the *MC4R* transcription start site. Intra-chromosomal interac-

tions between these two potential CTCF binding sites could bring the DNA region spanning the two large LD blocks, which contain SNPs that are significantly associated with adiposity-related traits, in close proximity to the *MC4R* promoter. CTCF has been shown to regulate interactions between promoters and distant enhancers by forming chromosomal loops. The developmental timing of the expression of genes at the  $\beta$ -globin locus ( $\epsilon$ ,  $G\gamma$ ,  $A\gamma$ ,  $\delta$ , and  $\beta$ ) is regulated by CTCF-mediated intra-chromosomal looping between the locus control region (LCR) and the promoter of the gene to be expressed [45]. At the *CFTR* gene, CTCF binds downstream of the gene and interacts with the *CFTR* promoter through a chromosomal loop, which is proposed to create an active chromatin hub [46]. Similar to CTCF's role at these loci, CTCF



**Figure 4. ENCODE-based transcriptional insulators near *MC4R*.** Schematic depicting the genomic region (hg19 assembly) surrounding *MC4R* that contains rs1152221 (highlighted in red) and ENCODE-annotated insulators. Adapted from the UCSC Genome Browser, <http://genome.ucsc.edu/> [51].

doi:10.1371/journal.pone.0096805.g004

could potentially facilitate *MC4R* expression through chromosomal loop formation.

A previous study conducted using participants in the Health ABC study and the Age Gene/Environment Susceptibility-Reykjavik (AGES-Reykjavik) study examined whether reported BMI-associated SNPs were associated with anthropometric and adiposity-related traits in the elderly [47]. The single SNP they examined in the *MC4R* region, rs571312, is located in the distal LD block, and not surprisingly, no association with BMI was identified. We also found no evidence for association between SNPs in the distal LD block and BMI, but by examining the entire genomic region, we identified highly significant SNP associations with BMI in the proximal LD block, thus highlighting the importance of the examination of the entire *MC4R* gene region.

Despite the significant SNP associations in non-coding DNA downstream of *MC4R* that we observed in Health ABC white participants, we did not observe significant SNP associations in these DNA regions in Health ABC black participants. There were fewer black participants in the Health ABC study than white participants, resulting in a loss of power in the analysis of SNP associations in black participants. A previously reported GWAS of BMI performed in individuals of African ancestry failed to identify SNP associations reaching genome-wide significance levels, but nominally significant SNP associations were identified near the 3' distal LD block of *MC4R* [48]. Previously reported BMI-associated SNPs from populations of European descent were evaluated in a meta-analysis of SNP associations with BMI in six cohorts composed of individuals of African ancestry ( $n = 4992$ ), and 2 of the 7 SNPs examined at the *MC4R* locus were nominally significant ( $P$ -value  $\leq 0.05$ ) [49]. A GWAS meta-analysis of BMI performed in a total of 71,412 individuals of African ancestry, in which Health ABC black participants contributed to 1.6% of the sample size, identified a genome-wide significant SNP association (rs6567160) near the distal LD block downstream of *MC4R* [50]. At the *MC4R* locus, the most significant SNP in African Americans (rs6567160) was not in LD (AFR  $r^2 = 0.03$ ) with the most significant SNP reported in individuals of European ancestry (rs571312), and rs571312 was not nominally associated with BMI in African Americans [50]. Taken together, these results indicate that SNPs downstream of *MC4R* are significantly associated with

BMI in African Americans, but allelic heterogeneity is likely to exist.

In addition to the low power for our analysis of SNP associations in black participants from the Health ABC study, there were also limitations to our case-control study using cases from the UCSF study of severe obesity. Specifically, the cases were younger and the percentage of females was higher compared to controls. Regression models adjusted for the effect of sex. However, the nearly perfect case-control separation by age prevented the assessment of the confounding effect of age. Only a single case (aged 70 years) overlapped with the age range of controls (minimum age of controls 69 years).

In summary, the DNA region downstream of *MC4R* containing our most significantly associated SNP did not act as an enhancer, but genomic annotation by ENCODE led us to a proposed model where intra-chromosomal interactions mediated by CTCF could bring a region containing SNPs significantly associated with adiposity in close proximity to the *MC4R* promoter. Our study draws attention to the region of the proximal LD block containing this putative insulator. This information could help to guide studies aimed at identifying the molecular mechanisms of genetic associations with adiposity in the *MC4R* region.

## Supporting Information

**Figure S1 SNP associations in and near *MC4R* with adiposity in white female Health ABC participants.** SNP genotypes from custom Illumina Golden Gate array. Gray points indicate association  $P$ -value  $> 0.05$ . Non-gray points indicate significant ( $P$ -value  $\leq 0.05$ ) associations with an adiposity trait of the corresponding color in the legend. Dashed line indicates cut-off value for empirical  $P$ -value  $\leq 0.05$ . LD heatmap indicates higher  $r^2$  measures with darker red colors. (EPS)

**Figure S2 CTCF position weight matrix from JASPAR core.** (EPS)

**Table S1 Adiposity-related traits in the Health ABC cohort by SNP genotype and race.** (DOCX)

**Table S2 SNP associations with adiposity-related traits in Health ABC black participants and stratified by sex.** (DOCX)

**Table S3 Obesity association with rs11152221 using cases and controls from Health ABC white participants.** (DOCX)

**Table S4 Characteristics of participants in case-control obesity study with cases from UCSF study.** (DOCX)

**Table S5 Obesity association with rs11152221 using cases from UCSF study.** (DOCX)

**Table S6 Characteristics of sequenced patients from UCSF study.** (DOCX)

**Table S7 CTCF binding sites within 250 bp of rs11152221.** (DOCX)

**Text S1 ARRIVE guidelines checklist.** (DOC)

## Acknowledgments

The authors are indebted to the subjects who participated in this study.

## Author Contributions

Conceived and designed the experiments: W-CH CV. Performed the experiments: DSE MAC MJK. Analyzed the data: DSE MAC MJK. Contributed reagents/materials/analysis tools: P-YK TH YL NA W-CH CV. Wrote the paper: DSE MAC CV. Critical revision of article: P-YK IM TH AK YL GJT NA W-CH CV.

## References

- Kelly T, Yang W, Chen CS, Reynolds K, He J (2008) Global burden of obesity in 2005 and projections to 2030. *Int J Obes (Lond)* 32: 1431–1437.
- Haslam DW, James WP (2005) Obesity. *Lancet* 366: 1197–1209.
- Hebebrand J, Volkmar AL, Knoll N, Hinney A (2010) Chipping away the 'missing heritability': GIANT steps forward in the molecular elucidation of obesity - but still lots to go. *Obes Facts* 3: 294–303.
- Mountjoy KG, Mortrud MT, Low MJ, Simerly RB, Cone RD (1994) Localization of the melanocortin-4 receptor (MC4R) in neuroendocrine and autonomic control circuits in the brain. *Mol Endocrinol* 8: 1298–1308.
- Lu D, Willard D, Patel IR, Kadwell S, Overton L, et al. (1994) Agouti protein is an antagonist of the melanocyte-stimulating-hormone receptor. *Nature* 371: 799–802.
- Huszar D, Lynch CA, Fairchild-Huntress V, Dunmore JH, Fang Q, et al. (1997) Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* 88: 131–141.
- Schwartz MW, Woods SC, Porte D Jr., Seeley RJ, Baskin DG (2000) Central nervous system control of food intake. *Nature* 404: 661–671.
- Calton MA, Ersoy BA, Zhang S, Kane JP, Malloy MJ, et al. (2009) Association of functionally significant Melanocortin-4 but not Melanocortin-3 receptor mutations with severe adult obesity in a large North American case-control study. *Hum Mol Genet* 18: 1140–1147.
- Beckers S, Zegers D, de Freitas F, Peeters AV, Verhulst SL, et al. (2010) Identification and functional characterization of novel mutations in the melanocortin-4 receptor. *Obes Facts* 3: 304–311.
- Hinney A, Vogel CI, Hebebrand J (2010) From monogenic to polygenic obesity: recent advances. *Eur Child Adolesc Psychiatry* 19: 297–310.
- Geller F, Reichwald K, Dempfle A, Illig T, Vollmert C, et al. (2004) Melanocortin-4 receptor gene variant 1103 is negatively associated with obesity. *Am J Hum Genet* 74: 572–581.
- Heid IM, Vollmert C, Hinney A, Doring A, Geller F, et al. (2005) Association of the 1031 MC4R allele with decreased body mass in 7937 participants of two population based surveys. *J Med Genet* 42: e21.
- Stutzmann F, Vatin V, Cauchi S, Morandi A, Jouret B, et al. (2007) Non-synonymous polymorphisms in melanocortin-4 receptor protect against obesity: the two facets of a Janus obesity gene. *Hum Mol Genet* 16: 1837–1844.
- Young EH, Wareham NJ, Farooqi S, Hinney A, Hebebrand J, et al. (2007) The V103I polymorphism of the MC4R gene and obesity: population based studies and meta-analysis of 29 563 individuals. *Int J Obes (Lond)* 31: 1437–1441.
- Wang D, Ma J, Zhang S, Hinney A, Hebebrand J, et al. (2010) Association of the MC4R V103I polymorphism with obesity: a Chinese case-control study and meta-analysis in 55,195 individuals. *Obesity (Silver Spring)* 18: 573–579.
- Meyre D, Delplanque J, Chevre JC, Lecoer C, Lobbens S, et al. (2009) Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat Genet* 41: 157–159.
- Xiang Z, Litherland SA, Sorensen NB, Proneth B, Wood MS, et al. (2006) Pharmacological characterization of 40 human melanocortin-4 receptor polymorphisms with the endogenous proopiomelanocortin-derived agonists and the agouti-related protein (AGRP) antagonist. *Biochemistry* 45: 7277–7288.
- Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, et al. (2009) Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* 41: 18–24.
- Loos RJ, Lindgren CM, Li S, Wheeler E, Zhao JH, et al. (2008) Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet* 40: 768–775.
- Willer CJ, Speliotes EK, Loos RJ, Li S, Lindgren CM, et al. (2009) Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 41: 25–34.
- Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, et al. (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 42: 937–948.
- Chambers JC, Elliott P, Zabaneh D, Zhang W, Li Y, et al. (2008) Common genetic variation near MC4R is associated with waist circumference and insulin resistance. *Nat Genet* 40: 716–718.
- Lindgren CM, Heid IM, Randall JC, Lamina C, Steinthorsdottir V, et al. (2009) Genome-wide association scan meta-analysis identifies three Loci influencing adiposity and fat distribution. *PLoS Genet* 5: e1000508.
- Scherag A, Dina C, Hinney A, Vatin V, Scherag S, et al. (2010) Two new Loci for body-weight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and german study groups. *PLoS Genet* 6: e1000916.
- Scherag A, Jarick I, Grothe J, Biebermann H, Scherag S, et al. (2010) Investigation of a genome wide association signal for obesity: synthetic association and haplotype analyses at the melanocortin 4 receptor gene locus. *PLoS ONE* 5: e13967.
- Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, et al. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489: 57–74.
- Snijder MB, Visser M, Dekker JM, Seidell JC, Fuerst T, et al. (2002) The prediction of visceral fat by dual-energy X-ray absorptiometry in the elderly: a comparison with computed tomography and anthropometry. *Int J Obes Relat Metab Disord* 26: 984–993.
- Swarbrick MM, Waldenmaier B, Pennacchio LA, Lind DL, Cavazos MM, et al. (2005) Lack of support for the association between GAD2 polymorphisms and severe human obesity. *PLoS Biol* 3: e315.
- de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, et al. (2005) Efficiency and power in genetic association studies. *Nat Genet* 37: 1217–1223.
- Li Q, Ritter D, Yang N, Dong Z, Li H, et al. (2010) A systematic approach to identify functional motifs within vertebrate developmental enhancers. *Dev Biol* 337: 484–495.
- Fisher S, Grice EA, Vinton RM, Bessling SL, Urasaki A, et al. (2006) Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nat Protoc* 1: 1297–1305.
- Kothary R, Clapoff S, Brown A, Campbell R, Peterson A, et al. (1988) A transgene containing lacZ inserted into the dystonia locus is expressed in neural tube. *Nature* 335: 435–437.
- Nagy A, Gertsenstein M, Vintersten K, Behringer R (2002) *Manipulating the Mouse Embryo: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory Press.
- Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, et al. (2006) In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 444: 499–502.
- Westfall PH, Young SS (1993) *Resampling-Based Multiple Testing: Examples and Methods for P-value Adjustment*. New York: Wiley-Interscience.
- Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32: W273–279.
- Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, et al. (2010) Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* 26: 2474–2476.
- Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, et al. (2012) Patterns of cis regulatory variation in diverse human populations. *PLoS Genet* 8: e1002639.
- Dempfle A, Hinney A, Heinzel-Gutenbrunner M, Raab M, Geller F, et al. (2004) Large quantitative effect of melanocortin-4 receptor gene mutations on body mass index. *J Med Genet* 41: 795–800.

40. Kistler-Heer V, Lauber ME, Lichtensteiger W (1998) Different developmental patterns of melanocortin MC3 and MC4 receptor mRNA: predominance of Mc4 in fetal rat nervous system. *J Neuroendocrinol* 10: 133–146.
41. Phillips JE, Corces VG (2009) CTCF: master weaver of the genome. *Cell* 137: 1194–1211.
42. Ernst J, Kellis M (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* 28: 817–825.
43. Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, et al. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473: 43–49.
44. Maston GA, Evans SK, Green MR (2006) Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* 7: 29–59.
45. Hou C, Dale R, Dean A (2010) Cell type specificity of chromatin organization mediated by CTCF and cohesin. *Proc Natl Acad Sci U S A* 107: 3651–3656.
46. Blackledge NP, Ott CJ, Gillen AE, Harris A (2009) An insulator element 3' to the CFTR gene binds CTCF and reveals an active chromatin hub in primary cells. *Nucleic Acids Res* 37: 1086–1094.
47. Murphy RA, Nalls MA, Keller M, Garcia M, Kritchevsky SB, et al. (2013) Candidate gene association study of BMI-related loci, weight, and adiposity in old age. *J Gerontol A Biol Sci Med Sci* 68: 661–666.
48. Kang SJ, Chiang CW, Palmer CD, Tayo BO, Lettre G, et al. (2010) Genome-wide association of anthropometric traits in African- and African-derived populations. *Hum Mol Genet* 19: 2725–2738.
49. Hester JM, Wing MR, Li J, Palmer ND, Xu J, et al. (2012) Implication of European-derived adiposity loci in African Americans. *Int J Obes (Lond)* 36: 465–473.
50. Monda KL, Chen GK, Taylor KC, Palmer C, Edwards TL, et al. (2013) A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. *Nat Genet* 45: 690–696.
51. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, et al. (2002) The human genome browser at UCSC. *Genome Res* 12: 996–1006.