Lawrence Berkeley National Laboratory
Lawrence Berkeley National Laboratory

Title

Medical Sequencing at the extremes of Human Body Mass

Permalink
https://escholarship.org/uc/item/22n042p5

Authors
Ahituv, Nadav
Kavaslar, Nihan
Schackwitz, Wendy
et al.

Publication Date
2006-09-01

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Medical Sequencing at the Extremes of Human Body Mass

Nadav Ahituv[1,2], Nihan Kavaslar[3], Wendy Schackwitz[1,2], Anna Ustaszewska[2], Joel Martin[2], Sybil Hébert[3], Heather Doelle[3], Baran Ersoy[4], Gregory Kryukov[5], Steffen Schmidt[5], Nir Yosef[6], Eytan Ruppin[6,7], Roded Sharan[6], Christian Vaisse[4], Shamil Sunyaev[5], Robert Dent[3], Jonathan Cohen[8], Ruth McPherson[3] and Len A. Pennacchio[1,2,*]

[1]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720 USA

[2]U.S. Department of Energy Joint Genome Institute, Walnut Creek, California 94598 USA

[3]Division of Cardiology and Lipoprotein & Atherosclerosis Research Group, University of Ottawa Heart Institute, Ottawa, K1Y 4W7, Canada

[4] Diabetes Center, University of California, San Francisco, California, USA.

[5]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA.

[6]School of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel.

[7]School of Medicine, Tel-Aviv University, Tel-Aviv 69978, Israel.

[8]Center for Human Nutrition and McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, TX, USA.

* To whom correspondence should be addressed:

Len A. Pennacchio, Genomics Division, One Cyclotron Road, MS 84-171, Lawrence Berkeley National Laboratory, Berkeley, CA 94720. Email: LAPennacchio@lbl.gov Phone: (510) 486-7498, Fax: (510) 486-4229.

ABSTRACT

Body weight is a quantitative trait with significant heritability in humans. To identify potential genetic contributors to this phenotype, we resequenced the coding exons and splice junctions of 58 genes in 379 obese and 378 lean individuals. Our 96Mb survey included 21 genes associated with monogenic forms of obesity in humans or mice, as well as 37 genes that function in body weight-related pathways. We found that the monogenic obesity-associated gene group was enriched for rare nonsynonymous variants unique to the obese (n=46) versus lean (n=26) populations. Computational analysis further predicted a significantly greater fraction of deleterious variants within the obese cohort. Consistent with the complex inheritance of body weight, we did not observe obvious familial segregation in the majority of the 28 available kindreds. Taken together, these data suggest that multiple rare alleles with variable penetrance contribute to obesity in the population and provide a deep medical sequencing-based approach to detect them.

Obesity is reaching epidemic proportions in developed countries and represents a significant risk factor for hypertension, heart disease, diabetes, and dyslipidemia[1]. While the growing prevalence of obesity in the population is thought to be caused by increasing caloric intake and declining energy expenditure[2], individual susceptibility to obesity is strongly influenced by heredity. Twin, adoption, and family studies have indicated that 40 to 70% of inter-individual variation in BMI is heritable [3,4]. In a limited number of cases, single gene defects have been linked to obesity [5], although the majority of cases are thought to be attributed to complex genetic and/or environmental interactions. In this study, we sought to explore the relationship between sequence variation in multiple candidate genes and the extremes of human body mass.

Candidate genes for the study included: (1) 21 genes that have been strongly associated with obesity, that when disrupted lead to monogenic forms of obesity in humans and/or that cause obesity when inactivated in mice, and (2) 37 genes involved in regulation of food intake [6], adipogenesis [7], energy expenditure [8] or lipid metabolism (Table 1). The coding exons and splice junctions of each gene were sequenced in 379 extremely obese Caucasian men and women (mean BMI 49.0 kg/m^2 and >95th percentile adjusted for age and sex) ascertained through an obesity clinic at the University of Ottawa, and in 378 lean (mean BMI 19.4 kg/m^2 and <10th percentile adjusted for age and sex) apparently healthy Caucasians who participated in a study of leanness at the same institution (Supplementary Table S1). A total of 134 kb (60 kb coding and 74 kb noncoding) was sequenced in each individual, representing 96 Mb of high quality sequence data with an average coverage of 734 individuals per exon (Table 2). Cumulatively, we identified 1074 genetic variants (Supplementary table S2), averaging one variant per 125 bp of the reference human genome sequence. 252 of the variants were common polymorphisms (minor allele frequency >1%) while the remaining 822 rare variants included

400 noncoding, 150 synonymous, 272 nonsynonymous including 3 in-frame indels and 8 severe alleles (6 out of frame indels and 2 nonsense changes). In accord with previous large-scale gene centric sequence analyses[9-11], we observed a paucity of nonsynonymous variants with increasing minor allele frequency, consistent with purifying selection acting on a significant fraction of such DNA sequence changes (Figure 1). Of the 1074 variants identified in this study, 989 (92%) were not listed in dbSNP (build 124) and as expected the majority of these variants (800/989, 81%) had a rare minor allele frequency (<1%).

It has been previously reported that multiple rare variants can have a strong effect on complex traits, especially in the population extremes of a given phenotype[12,13]. We therefore examined the frequencies of the nonsynonymous variants in the obese and lean cohorts. Of the 272 rare nonsynonymous changes identified, 213 were unique to one group, with a small excess in the obese (n=118) compared to the lean population (n=95) that did not reach statistical significance. A similar analysis revealed that the prevalence of unique rare synonymous variants, which approximate functionally neutral changes, was essentially identical in the obese and lean cohorts (60 in obese, 61 in lean). We next examined the distributions of nonsynonymous and synonymous variants within each gene individually and found that none of the genes had a statistically significant excess of nonsynonymous variants in the obese or lean groups. However, when the genes associated with monogenic forms of obesity were considered together, unique nonsynonymous variants were significantly more common in the obese (46 variants in 41 individuals) than in the lean group (26 variants in 27 individuals) ($P<0.05$; Fisher's exact test). In contrast, the number of unique synonymous variants in these genes was almost identical among the obese (n=18) and lean (n=16) groups. It is

worth noting that the genes that accounted for the highest difference are: *MC4R* (8 in obese versus 2 in lean), *SIM1* (6 in obese versus 0 in lean), and *UCP3* (5 in obese versus 2 in lean).

The excess of nonsynonymous variants among obese individuals may reflect chance fluctuation in allele frequencies, population stratification, or the accumulation in this group of functional sequence variants that predispose individuals to obesity. Chance fluctuation in allele frequencies seems unlikely, as the excess of nonsynonymous variants in the obese group was not due to an increased number of variants in any single gene, but rather to the cumulative contribution of variants at several unlinked loci. Population stratification also seems improbable, as both groups comprised Caucasian men and women from the same region (Ottawa, Canada). Furthermore, the number of synonymous variants (Table 1) and the allele frequencies of ~250 common sequence variants (see below) in these genes were similar in the obese and lean groups. Therefore it seems likely that the excess of rare variants in the obese group represents the accumulation of functional alleles that contribute to the phenotype in these individuals.

As a first step to assess the functional significance of the nonsynonymous sequence variants identified in the 21 genes associated with monogenic forms of obesity in humans and mice, we used the computer algorithm PolyPhen[14] to predict the effects of amino acid substitutions on protein function. We observed that variants identified in the monogenic obesity gene group were more likely to be deleterious in the obese cohort than those identified in the lean cohort (19 in the obese versus 4 in the lean; $P < 0.002$, exact binomial test; Figure 2). In comparison, the number of benign (25 in the obese and 21in the lean) and synonymous (18 in the obese versus 16 in lean) variants in these genes was similar in both cohorts. In contrast, the distribution of synonymous, benign, and deleterious alleles in

the 37 candidate genes not associated with monogenic forms of obesity was similar in the obese and lean groups. This finding is consistent with the notion that the excess of nonsynonymous sequence variants in the obese cohort among the monogenic obesity group of genes reflects the accumulation of functional variants.

Next we examined familial segregation for 28 rare variants identified in 14 genes (Table 3 & Supplementary Table S3) in obtainable obese kindreds, to determine whether nucleotide changes within these genes segregate with BMI. We used *MC4R* as a test case, since it is the most common cause of monogenic obesity, estimated to account for 1-6% of severe obesity cases[15]. In our study we identified 8 nonsynonymous variants that were unique to the obese group compared to only 2 in the lean cohort (Table 1). We found that the mutant *MC4R* alleles clearly failed to segregate with obesity in three of the five kindreds, with 6 or more family members available (Table 3), including an allele with a previously characterized frame-shift mutation (L211 4bp deletion)[16] that is almost certainly functional. To corroborate that these *MC4R* variants were indeed functional, we performed established *in vitro* functional assays for the novel *MC4R* variants[15] identified in our obese population. Of the five putative mutations analyzed for segregation, four displayed impaired MC4R function (Table 3). These findings are consistent with previous studies that also show incomplete correlation between *MC4R* mutations and obesity[17,18] illustrating the difficulties inherent in determining the correspondence between genotype and phenotype in common, complex phenotypes such as obesity. Although a number of the kindreds available for study were small, none of the other rare variants examined in 13 additional genes showed significant segregation with BMI in a total of 21 kindreds with the exception of PYY Q62P, which we have previously reported[19] (Supplementary Table S3).

While the goal of our study was not to perform an exhaustive genetic association study between common variants and BMI, we identified 252 polymorphisms with a frequency greater than 1% and examined their frequency distribution in the obese versus lean cohort (Supplementary table S2). We found only 2 variants, *DGAT1* rs6599571 and *NTS* rs1800832 (both in their gene's respective 5'UTR; Supplementary Table S4), that displayed a significant frequency difference between the two populations. In an attempt to replicate these findings, we compared their frequency in a second cohort of obese (n=382; mean BMI 38.6) and lean (n=381; mean BMI=20.8) subjects (Supplementary Table S1). In both cases, we observed no difference in the allele frequencies between the second obese and lean cohorts, supporting that the initial observation was likely a false-positive discovery or limited to very extreme BMI phenotypes (Supplementary Table S4). We should further note that none of the 38 sequenced common variants that were previously examined for their association with BMI displayed a significant frequency difference between our original obese and lean groups (Supplementary Table S5). Among them is the recently characterized rs7566605 SNP near *INSIG2* [20], that showed no difference in either of our obese/lean panels (Supplementary Table S5). These results suggest that in this population, common variants within the coding regions and their proximal exon-intron junction in this subset of 58 genes are unlikely to contribute appreciably to extreme BMI susceptibility. However, because we screened primarily the coding sequences and splice junctions of these genes, we can not exclude the possibility that common sequence variations in noncoding regions not sequenced in this study may have significant effects on BMI.

Whereas the heritability of BMI has been firmly established, the identification of genes that contribute to obesity has proved challenging. Resequencing of candidate genes selected on biological plausibility has in a few instances resulted in the identification of variants associated with obesity. For instance,

the observation that *Mc4r* mouse knockouts are obese [21] led to the subsequent finding that mutations in this gene may lead to obesity in humans [16,22]. In this study we sought to use a similar approach using a large-scale sequencing strategy of numerous obesity candidate genes in two extreme BMI populations. We did not uncover a large number of novel genes associated with obesity, an endeavor that may have been obstructed by reasons ranging from a partial candidate gene list (n=58), a large but still limited collection of strictly Caucasian individuals (n=~380 in each group), sequencing mainly coding regions, and limited power and availability of subjects pedigrees. However, we did identify several genes that warrant further investigations. For instance, we observed a noteworthy rare nonsynonymous variant difference across the obese and lean cohorts in *SIM1* (6 in obese versus 0 in lean) and *PRKAG3* (10 in obese versus 4 in lean), suggesting that nonsynonymous variants within these genes may influence susceptibility to obesity. *SIM1* is of particular interest due to its strong biological plausibility including evidence that human chromosomal aberrations within the *SIM1* region may lead to obesity[23,24], the observation that *Sim1* heterozygous null mice develop obesity[25], and absence of reported obesity associated rare nonsyndromic variants. In addition, we uncovered a significant difference between the total number of obese and lean nonsynonymous variants in previously characterized monogenic obesity genes, indicating that multiple rare variants may have an incomplete effect on this phenotype. Our familial segregation analysis demonstrated that even thoroughly characterized human monogenic obesity genes such as *MC4R* fail to show consistent linkage with BMI, further suggesting these variants exhibit variable penetrance. While our analysis encompassed only a modest fraction of candidate BMI genes, it strengthens the hypothesis that the majority of genetic etiology governing obesity is complex and likely to be influenced by a combination of multiple susceptibility alleles, the majority of which are not independently causative of extreme BMI.

MATERIALS AND METHODS

Subjects

Obese Caucasian subjects were recruited from the patient population of the University of Ottawa Weight Management Clinic and the Heart Institute Lipid Clinic using previously reported criteria [19]. Briefly, inclusion criteria included a BMI > 36 kg/m^2 and a history of obesity for at least 10 years of adult life and no history of treatment with oral glucocorticoids, anti-psychotics, lithium or medical conditions including major depression, bipolar affective disorder or psychosis. Lean subjects of the same ethnic background, with a BMI ≤ the 10th percentile for age and sex and no prior history of having had a BMI > 25th percentile for more than a 2 year consecutive period, were recruited from the Ottawa community (Supplementary Table S1). These subjects were excluded if they had any medical conditions affecting weight gain such as hypo- or hyper- thyroidism, eating disorders, major depression, or malabsorption syndromes. BMI for obese and lean subjects was categorized according to the population percentiles for age and sex using the Canadian Heart Health Survey data for subjects over the age of 18 years (data on file; Health Canada) and NHANES data for children [26]. This study was approved by the institutional review boards of the University of Ottawa Heart Institute and the Ottawa Hospital and informed written consent was obtained from all participants. Genomic DNA was extracted from white blood cells by standard methods [27].

Sequencing and data analysis

Primers were designed to give a maximum product size of 500bp and a minimum of 40bp flanking the splice sites using the Exon Locator & eXtractor for Resequencing program (EXLR) (http://mutation.swmed.edu/ex-lax/). An M13F tag (gttttcccagtcacgacgttgta) and M13R tag

(aggaaacagctatgaccat) was added to forward and reverse primers respectively. 10ng of DNA from each sample was amplified in a 10ul PCR reaction using AmpliTaq Gold® (Applied Biosystems) and cleaned using the PCR product pre-sequencing kit (USB Corporation). Bidirectional sequencing was carried out using both M13 primers along with BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) (http://www.jgi.doe.gov/sequencing/protocols/archive/BigDye3.1auto1.0.doc) and cleaned with Tetra-Ethylene Glycol before separation on a 3730xl DNA Analyzer (ABI). Base calling, quality assessment and assembly were carried out using the Phred, Phrap, Polyphred, Consed software suite (www.phrap.org). All sequence variants identified were verified by manual inspection of the chromatograms and 156/157 (99%) nonsynonymous variants were positive for a second reverification using another independent sequencing reaction.

PolyPhen analysis

All coding SNPs were subdivided in groups of frameshift/nonsense variants, synonymous variants and missense variants. Missense variants were further classified with respect to their potential impact on protein structure or function based on sequence conservation using a new version of the PolyPhen method[14]. PolyPhen relies on the analysis of multiple sequence alignments of homologous proteins together with functional annotation and structural information if available. The new version of PolyPhen constructs multiple sequence alignment using a pipeline of several existing programs for aligning sequences, alignment quality control, and clustering of sequences. Computational prediction methods are statistical in nature, and therefore, certain fractions of false-positive (~10%) and false-negative (20%-30%) predictions are to be expected. However, application of computational predictions increases power to detect difference in number of rare functional nonsynonymous variants in candidate genes between populations with different phenotypes.

MC4R functional analysis

Cloning and functional studies of the *MC4R* mutations were performed as described previously [17,28,29] Briefly, since *MC4R* is a single exon gene, mutated alleles were amplified and cloned directly from the genomic DNA of the patient. This also allowed for confirmation of the presence and the nature of the mutations. Human *MC4R* alleles were cloned into the pcDNA3 expression vector (Invitrogen) to express the native form and N-terminal Flag tagged and/or C-terminal V5His tagged form of the receptor. All expression vectors were sequenced to establish the presence of the mutation and the absence of PCR-induced mutations.

For alpha-MSH activation studies, receptors were transiently transfected into a HEK293 cell line stably expressing luciferase under the control of a cAMP responsive promoter [28]. Cells were split into 96 well plates 24 hour after transfection and, 36 hours post-transfection, washed and incubated in stimulation medium (MEM-alpha medium containing 0.1 mg/ml BSA and 0.25mM isobutylmethylxanthine) and stimulated with different concentrations of alpha-MSH (Sigma) for 6h at 37°C in a 5% CO_2 incubator. Luciferase activity, representing cAMP produced in response to alpha-MSH, was assessed using the Steady-Glo Luciferase Assay System (Promega) and a microplate luminescence counter (Packard Instrument). Results were normalized to the maximal stimulation by 8Br-cAMP. Basal activity of the receptors were determined by transient co-transfection with the c-AMP dependent luciferase expressing plasmid. All experiments were normalized for transfection efficiency by co-transfection of a plasmid encoding the Renilla luciferase expression plasmid PRL RSV to control for transfection efficiency. Data were analyzed using the GraphPad Prism software (Graphpad Software).

Common variant statistical analysis

Common SNPs were preprocessed to remove tri-allelic SNPs, and SNPs for which more than 50% of the data were missing. In addition we clustered together SNPs that differed in at most 3 individuals, picking one representative from each such cluster. Chi-square tests were applied for each of the remaining 252 SNPs based on a contingency table of genotype-phenotype frequencies. The obtained *p*-values were adjusted for multiple SNP testing using the false discovery rate (FDR) procedure[30].

ACKNOWLEDGEMENTS

We wish to thank the Joint Genome Institute's production sequencing group and Thet Naing for technical assistance, members of the Rubin lab for helpful comments on the manuscript, and the many subjects and their families who participated in these studies. Research was conducted at the E.O. Lawrence Berkeley National Laboratory and the Joint Genome Institute performed under Department of Energy Contract DE-AC0378SF00098, University of California (LAP). Research performed at the Ottawa Heart Institute was supported by a grant from the Heart & Stroke Foundation of Ontario #NA5413 (to RM). Subject recruitment was supported, in part, by a grant from GlaxoSmithKline (to RM & RD). RS was supported by an Alon Fellowship.

REFERENCES

1. Bell, C.G., Walley, A.J. & Froguel, P. The genetics of human obesity. *Nat Rev Genet* **6**, 221-34. (2005).
2. French, S.A., Story, M. & Jeffery, R.W. Environmental influences on eating and physical activity. *Annu Rev Public Health* **22**, 309-35. (2001).
3. Comuzzie, A.G. & Allison, D.B. The search for human obesity genes. *Science* **280**, 1374-7. (1998).
4. Friedman, J.M. A war on obesity, not the obese. *Science* **299**, 856-8. (2003).
5. Perusse, L. et al. The human obesity gene map: the 2004 update. *Obes Res* **13**, 381-490. (2005).
6. Flier, J.S. Obesity wars: molecular progress confronts an expanding epidemic. *Cell* **116**, 337-50. (2004).
7. Harp, J.B. New insights into inhibitors of adipogenesis. *Curr Opin Lipidol* **15**, 303-7. (2004).
8. Lowell, B.B. & Spiegelman, B.M. Towards a molecular understanding of adaptive thermogenesis. *Nature* **404**, 652-60. (2000).
9. Halushka, M.K. et al. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet.* **22**, 239-47. (1999).
10. Leabman, M.K. et al. Natural variation in human membrane transporter genes reveals evolutionary and functional constraints. *Proc Natl Acad Sci U S A.* **100**, 5896-901. (2003).
11. Cargill, M. et al. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet.* **22**, 231-8. (1999).
12. Cohen, J.C. et al. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* **305**, 869-72. (2004).
13. Cohen, J.C. et al. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc Natl Acad Sci U S A.* **103**, 1810-5. (2006).
14. Ramensky, V., Bork, P. & Sunyaev, S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* **30**, 3894-900. (2002).
15. Lubrano-Berthelier, C. et al. Molecular genetics of human obesity-associated MC4R mutations. *Ann N Y Acad Sci* **994**, 49-57. (2003).
16. Yeo, G.S. et al. A frameshift mutation in MC4R associated with dominantly inherited human obesity. *Nat Genet.* **20**, 111-2. (1998).
17. Vaisse, C. et al. Melanocortin-4 receptor mutations are a frequent and heterogeneous cause of morbid obesity. *J Clin Invest.* **106**, 253-62. (2000).
18. Hinney, A. et al. Prevalence, spectrum, and functional characterization of melanocortin-4 receptor gene mutations in a representative population-based sample and obese adults from Germany. *J Clin Endocrinol Metab.* **91**, 1761-9. (2006).
19. Ahituv, N. et al. A PYY Q62P variant linked to human obesity. *Hum Mol Genet.* **15**, 387-91. (2006).
20. Herbert, A. et al. A common genetic variant is associated with adult and childhood obesity. *Science.* **312**, 279-83. (2006).
21. Huszar, D. et al. Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell.* **88**, 131-41. (1997).
22. Vaisse, C., Clement, K., Guy-Grand, B. & Froguel, P. A frameshift mutation in human MC4R is associated with a dominant form of obesity. *Nat Genet.* **20**, 113-4. (1998).

23. Holder, J.L., Jr., Butte, N.F. & Zinn, A.R. Profound obesity associated with a balanced translocation that disrupts the SIM1 gene. *Hum Mol Genet.* **9**, 101-8. (2000).
24. Faivre, L. et al. Deletion of the SIM1 gene (6q16.2) in a patient with a Prader-Willi-like phenotype. *J Med Genet.* **39**, 594-6. (2002).
25. Michaud, J.L. et al. Sim1 haploinsufficiency causes hyperphagia, obesity and reduction of the paraventricular nucleus of the hypothalamus. *Hum Mol Genet.* **10**, 1465-73. (2001).
26. Kuczmarski, R.J. et al. 2000 CDC Growth Charts for the United States: methods and development. *Vital Health Stat 11*, 1-190. (2002).
27. Miller, S.A., Dykes, D.D. & Polesky, H.F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**, 1215. (1988).
28. Lubrano-Berthelier, C. et al. Intracellular retention is a common characteristic of childhood obesity-associated MC4R mutations. *Hum Mol Genet.* **12**, 145-53. (2003).
29. Lubrano-Berthelier, C. et al. Melanocortin 4 receptor mutations in a large cohort of severely obese adults: prevalence, functional classification, genotype-phenotype relationship, and lack of association with binge eating. *J Clin Endocrinol Metab.* **91**, 1811-8. (2006).
30. Benjamini, Y. & Hochberg, T. Controlling the False Discovery Rate: a practical and powerful approach to multiple testing. *J. Royal Stat. Soc.* **B85**, 289-300 (1995).

FIGURE LEGENDS

Figure 1. The percentage of nonsynonymous, synonymous, and intronic variants for different minor allele frequencies.

Figure 2. PolyPhen distribution analysis of variants unique to the obese and lean cohorts. **(a)** Data presented for genes with evidence of monogenic involvement in obesity and **(b)** for genes with biological plausibility for a role in obesity. ** $P < 0.002$.

Table 1. A summary of rare coding variants unique to the obese or lean population. Genes were divided into (1) genes that have been strongly associated with obesity (Monogenic Obesity Genes) and (2) obesity candidate genes based on biological plausibility (Candidate Obesity Genes). References can be found in the supplementary material. NS=Nonsynonymous, S=Synonymous.

Monogenic Obesity Genes								
Gene	**Mouse Knockouts**	**Mouse Transgenics**	**Human Mutations**	**Associations (ref.)**	**OBESE**		**LEAN**	
					NS	**S**	**NS**	**S**
ATGL	Obese[31]	None	None	None	3	1	2	2
BRS3	Obese[32]	None	None	None	1	0	3	0
CART	Obese[33]	None	Obese[34]	[35-37]	1	1	0	1
FABP4	Obese[38]	None	None	[39]	1	0	2	0
GPR7	Obese[40]	None	None	None	3	0	1	0
HTR2C	Obese[41]	None	None	[42,43]	1	0	0	0
IL6	Obese[44]	None	None	[45-51]	0	1	0	0
LEPTIN	Obese[52]	Lean[53]	Obese[54]	[55-62]	0	3	0	1
MC3R	Obese[63]	None	Obese[64]	[65,66]	2	0	1	1
MC4R	Obese[22]	None	Obese[15]	[67-71]	8	1	2	1
NHLH2	Obese[72]	None	None	None	2	0	1	1
NMU	Obese[73]	None	None	None	1	0	1	0
NPB	Obese[74]	None	None	None	1	0	2	0
NPY1R	Obese[75]	None	None	None	1	1	2	1
NPY2R	Obese[76]	None	None	[77-79]	2	3	2	0
NPY5R	Obese[80]	None	None	[81]	1	1	1	0
NR0B2	No apparent phenotype[82]	None	Obese[83,84]	[84-86]	3	0	2	0
POMC	Obese[87]	None	Obese[88]	[89-97]	2	3	1	3
PYY	Obese[98]	None	Obese[20]	[77-79,99]	2	0	1	0
SIM1	Obese[26]	Lean[100]	Obese[24,25]	None	6	2	0	2
UCP3	No apparent phenotype[101]	Lean[102]	Obese[103]	[104-115]	5	1	2	3
TOTAL					**46**	**18**	**26**	**16**
Obesity Candidate Genes								
Gene	**Mouse KO**	**Mouse Tg**	**Human Mutations**	**Associations (ref.)**	**OBESE**		**LEAN**	
					NS	**S**	**NS**	**S**
ACDC	Insulin resistance[116]	None	None	[117-123]	2	0	2	0
AGRP	No apparent phenotype[124]	Obese[125]	None	[126,127]	1	1	0	2
APOA5	Hyperlipidemia[128]	Lipid[128]	None	[129,130]	1	0	2	1
ARNT2	Lethal[131]	None	None	None	2	2	3	0
ASIP	No apparent phenotype[132]	Obese[133]	None	None	0	0	0	0
C3AR1	Hypoallergic[134]	None	None	None	4	0	4	3
CCK	No apparent phenotype[135]	None	None	None	0	0	1	0
CPT1B	None	None	None	None	5	2	7	2
CSF2	Pulmonary anomalies[136]	None	None	None	0	0	0	1
CTRP2	None	None	None	None	1	2	0	2
DGAT1	Lean[137]	None	None	[138,139]	5	3	2	2
DGAT2	Lean[140]	None	None	None	5	0	3	2
GHRL	No apparent phenotype[141]	None	None	[142-145]	1	0	0	1

GHSR	No apparent phenotype[146]	None	None	[147-149]	1	2	2	1
GPR8	None	None	None	None	1	2	2	5
HSD11B1	Obesity resistance[150]	Obese[151]	None	[152,153]	0	1	1	0
HTR7	Hyperthermia[154]	None	None	None	1	2	1	3
INSIG1	None	None	None	None	0	2	3	0
INSIG2	None	None	None	None	1	2	2	1
LIPC	Hyperlipidemia[155]	None	None	[156]	4	5	2	7
NMUR1	None	None	None	None	4	4	2	1
NMUR2	None	None	None	None	4	0	3	0
NPY	No apparent phenotype[157]	None	None	[158-161]	0	0	0	0
NTS	No apparent phenotype[162]	None	None	None	0	0	4	0
PPARGC1A	Lean[163]	None	None	[164-169]	3	1	4	1
PPY	None	Lean[170]	None	None	0	0	1	0
PRKAA1	None	None	None	None	3	1	4	0
PRKAA2	Reduced glucose tolerance [171]	None	None	[172]	4	2	3	1
PRKAB1	None	None	None	[172]	0	1	0	0
PRKAB2	None	None	None	[172]	2	0	1	0
PRKAG1	None	None	None	None	0	1	0	1
PRKAG2	None	None	Heart[173]	None	2	0	1	2
PRKAG3	Impaired glycogen metabolism[174]	Glycogen[174]	None	None	10	3	4	1
RETN	Impaired gluconeogenesis[175]	Obese[176]	None	[177-182]	0	0	1	0
SIRT1	Increased insulin sensitivity[183]	None	None	None	3	2	1	4
TGFBR2	Impaired embryogenesis[184]	None	None	None	1	1	1	1
WDTC1	None	None	None	None	1	0	2	0
TOTAL					**72**	**42**	**69**	**45**

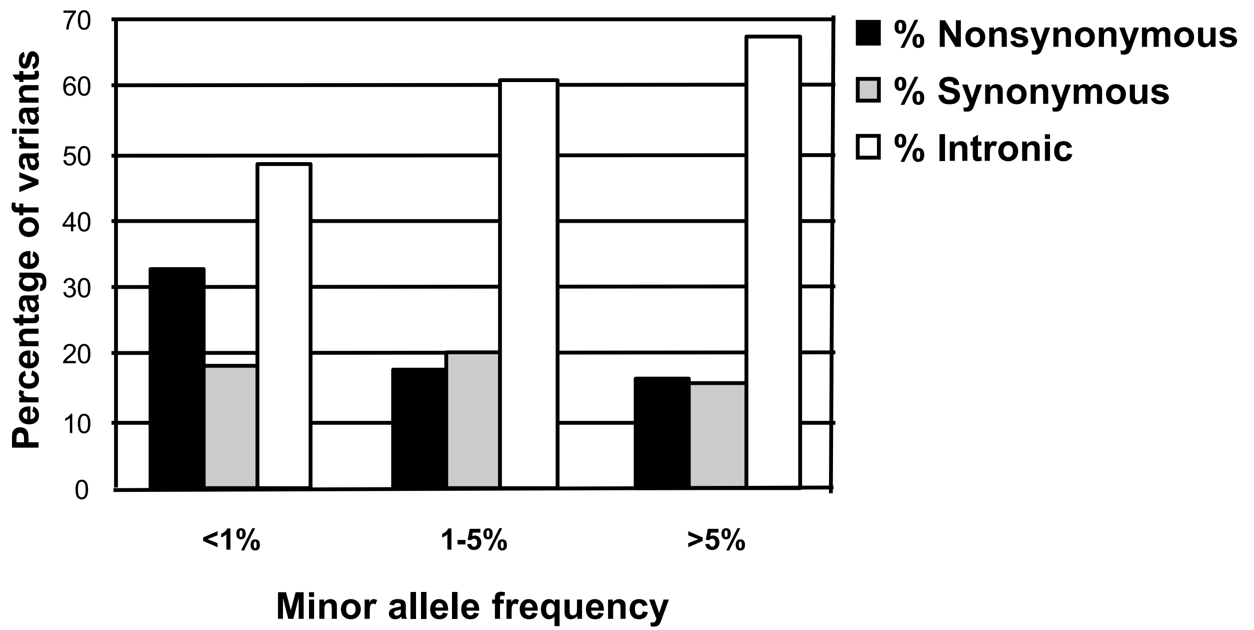
Table 2. Sequencing summary.

Genes	58
Exons	324
TOTAL genomic sequence covered (bp)	134,449
Coding sequence	60,372
Non-coding sequence	74,077
TOTAL sequence scanned (bp)	96,059,368
Coding sequence	44,254,489
Non-coding sequence	51,804,879
TOTAL variants	1,074
Rare nonsynonymous variants	272
Rare synonymous variants	150
Rare non-coding variants	400
Total rare variants	822
Common nonsynonymous variants	43
Common synonymous variants	44
Common non-coding variants	165
Total common variants	252
Novel variants identified	989
Known dbSNPs validated	85
dbSNPs not validated	366

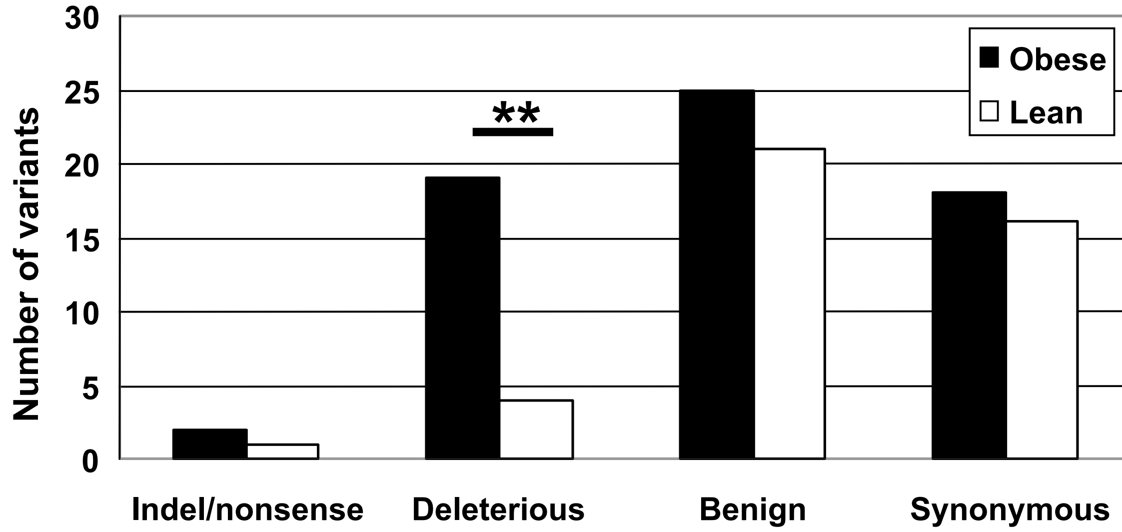
Table 3. Characterization of *MC4R* nonsynonymous variants found in obese subjects.

Variant	Sequence	N	Known/ Novel	Functional Studies			Carriers		Non-carriers		
				alpha-MSH activation (EC50)	Basal activity	Summary	genotyped (n)	BMI kg/m^2 mean (SD)	genotyped (n)	BMI kg/m^2 mean (SD)	P value
S30F	tgagt[c/t]ccttg	1	Known[185]	Not tested alone[28]			Not tested		Not tested		
G32E	ccttg[g/a]aaaag	1	Novel	0.3nM	70%	Minor	5 G/A	30.3 (4.9)	5 G/G	29.5 (4.6)	ns[a]
E61K	tgttg[g/a]agaat	1	Novel	low	≤10%	Severe	3G/A	40.6 (6.7)	1G/G	31.2	-
S127L	tgact[c/t]ggtga	1	Known[28]	29nM	80%	Intermediate	3C/T	37.1 (4.3)	3C/C	36.7 (14.0)	ns
L211Del[b]	ttct[ctct/-]atgt	2	Known[16]	Truncated receptor		Severe	F1.2C/A & +/-	43.2 (9.4)	F1.4C/C & +/+	28.0 (4.1)	-
P299H[b]	cgatc[c/a]tctga	2	Known[28]	negative	≤10%	Severe	F2.2C/A & +/-	35.9 (12.4)	F2.5C/C & +/+	39.7 (6.2)	ns
A303T	tttat[g/a]cactc	1	Novel	low	≤10%	Severe	2G/A	35.8 (mom) 20.1[c] (child)	4G/G	26.6 (1.7)	-
C326R	gcctt[t/c]gtgac	1	Novel	0.4nM	150%	Minor	2T/C	30.2 (1.7)	1T/T	25.3	-
WT[d]				0.3nM	100%						

[185]Reference appears in the supplementary references.
[a] Not significant.
[b]Individuals who had the L211Del also had the P299H variant.
[c]1.5 year old child, BMI >99% percentile.
[d]Wild-type.


Percentage of variants
70
60
50
40
30
20
10
0
<1%
1-5%
>5%
Minor allele frequency
% Nonsynonymous
% Synonymous
% Intronic

a



b