Hum Genet (2002) 110 : 412–417
DOI 10.1007/s00439-002-0715-5

ORIGINAL INVESTIGATION

**Susan Moffett · Jeremy Martinson · Mark D. Shriver
Ranjan Deka · Stephen T. McGarvey
Ramiro Barrantes · Robert E. Ferrell**

Genetic diversity and evolution of the human leptin locus tetranucleotide repeat

Received: 28 November 2001 / Accepted: 25 February 2002 / Published online: 4 April 2002

Abstract To better understand the evolutionary history of the gene region containing the multifunctional adipose tissue hormone leptin, we genotyped 1,957 individuals from 12 world populations for a highly variable tetranucleotide repeat polymorphism located 476 bp 3′ of exon 3 of the leptin gene. Common alleles shared among populations, alleles specific to geographically defined populations, and the homologous alleles in the common and pygmy chimpanzee, the gorilla and the orangutan, were sequenced to define the allelic variation at the nucleotide level. These data reveal a common set of alleles shared among world populations, presumed to have arisen from a great ape ancestral allele prior to the divergence of the major geographical subdivisions of the human population, a subset of alleles specific to populations of African ancestry and a second set of alleles that arose by tandem duplication of the core repeat unit following the separation of African and non-African populations. These findings emphasize the complex evolutionary history of this locus and raise cautions about the pooling of alleles at this locus in association studies.

S. Moffett · R.E. Ferrell (✉)
Department of Human Genetics,
Graduate School of Public Health, University of Pittsburgh,
Pittsburgh, PA 15261, USA
e-mail: rferrell@helix.hgen.pitt.edu,
Tel.: +1-412-6243018, Fax: +1-412-6243020

J. Martinson
Department of Genetics, University of Nottingham,
Nottingham, UK

M.D. Shriver
Department of Anthropology, Penn State University,
University Park, Pennsylvania, USA

R. Deka
Department of Environmental Health, University of Cincinnati,
Cincinnati, Ohio, USA

S.T. McGarvey
International Health Institute, Brown University Medical School,
Providence, Rhode Island, USA

R. Barrantes
Institutode Investigaciones en Salud, Universidad de Costa Rica,
San Jose, Costa Rica

Introduction

The discovery that mutations in the mouse leptin gene were responsible for the mouse obesity phenotype *ob* (Zhang et al. 1994), and that mutations in the human homolog (LEP) cause severe early-onset obesity in humans (Montague et al. 1997; Strobel et al. 1998) has stimulated enormous progress in understanding the endocrine control of adiposity. Analysis of the leptin/leptin receptor system has revealed a role for leptin in all aspects of hypothalamic function, including hypothalamic control of feeding behavior and regulation of the hypothalamic-pituitary-gonadal axis, the sympathetic nervous system and the immune system (Campfield and Smith 1998; Ahima and Flier 2000; Harris 2000; Baile et al. 2000).

While mutations in the LEP gene are a rare cause of obesity in humans, there has been much interest in determining whether polymorphic variation in or near the LEP gene influences susceptibility to obesity in the general population. Using microsatellite markers flanking the LEP locus at human chromosome 7q31.3-32.1, several groups reported evidence of linkage and/or association between variation in the LEP gene region and traits related to obesity (Borecki et al. 1994; Clement et al. 1996; Duggirala et al. 1996; Norman et al. 1996; Reed et al. 1996; Bray et al. 1996; Hasstedt et al. 1997; Oksanen et al. 1997). Although the findings were not always consistent, Allison and Heo (1998) concluded from a meta-analysis of the linkage data that the evidence of a gene influencing obesity in the region of the LEP locus was extremely strong (P=1.5×l0^{-5}). Studies examining single nucleotide polymorphisms within and flanking the LEP locus also support an association between LEP variation and obesity-related traits (Hager et al. 1998; Mammes et al. 1998; Li et al. 1999). Since these associations are likely the result of linkage disequilibrium (LD) between marker alleles and functional variation in the LEP gene,

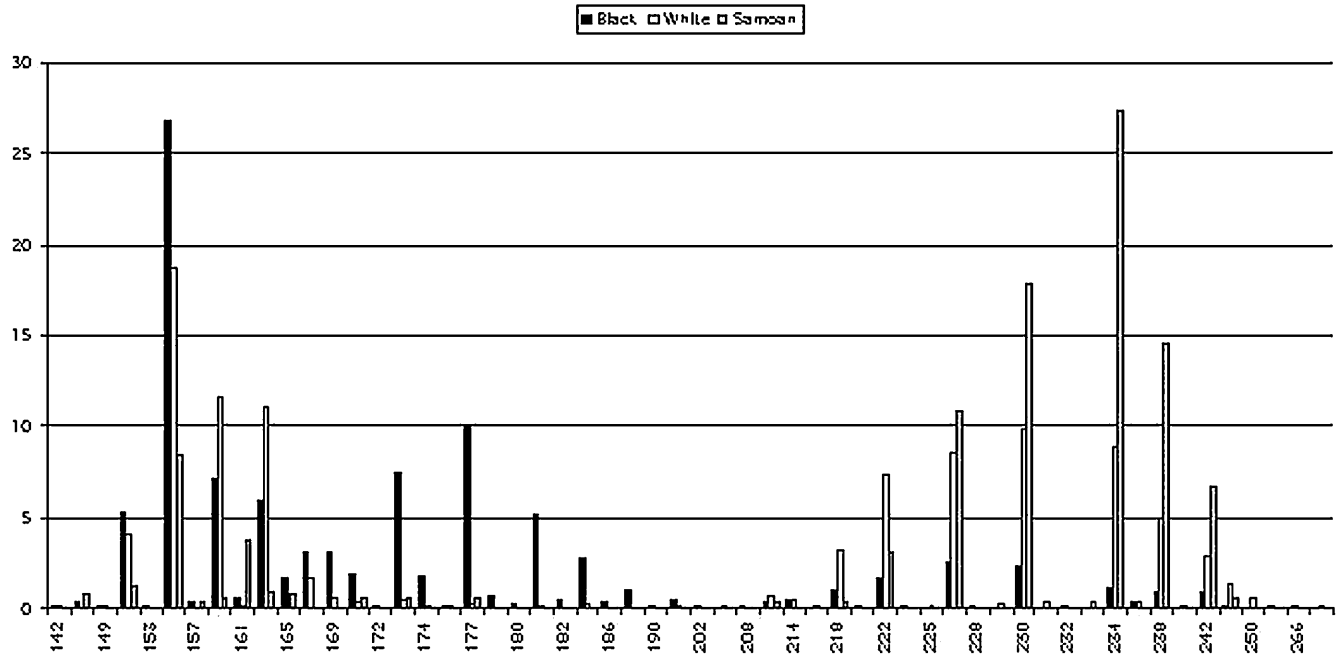

Fig. 1 Distribution of leptin alleles observed in African Americans (*n*=600), European Americans (*n*=624) and Samoans (*n*=104)

and this LD is dependent on the evolutionary history of the locus, studies of the evolution of this locus may help elucidate the nature of these associations. Shintani et al. (1996) identified a highly variable tetranucleotide repeat, $(CTTT)_n$, 3,912 bp 3′ of the LEP stop codon (476 bp 3′ of the 3′UTR), and we have used this marker to examine the population history of the LEP locus in world populations.

Materials and methods

Population samples

The US population groups examined include unrelated African Americans (USAA; *n*=600) and mixed European Americans (USEA, *n*=624) from the general US population. The world populations include a sample of Samoans (SAM; *n*=104) previously described by Deka et al. (1995), the Cabecar of Costa Rica (CR; *n*=35) described by Barrantes et al. (1990), African populations from West Africa (WAF; *n*=82) and the Central African Republic (CAR; *n*=48), Asians from Hong Kong (HK; *n*=43) and Taiwan (TW; *n*=73), Europeans from Britain (BR; *n*=35), Cyprus (CR; *n*=38) and Russia (RUS; *n*=96), the Kota Kinabalu of Borneo collected from Sabah province, Malaysia (KK; *n*=129) and a sample from Madagascar (MAD; *n*=88), all previously reported by Martinson et al. (2000). The common chimpanzee (*Pan troglodytes*), pygmy chimpanzee (*Pan paniscus*), gorilla (*Gorilla gorilla*) and orangutan (*Pongo pigmaeus*) were from the Yerkes Primate Research Center, Atlanta, Ga.

Laboratory methods

The leptin 3′-tetranucleotide repeat was amplified using the polymerase chain reaction (PCR) with the flanking primers: HOBF 5′-AGTTCAAATAGAGGTCCAAATCA-3′ and HOBR 5′-TTCTGAGGTTGTGTCACTGGCA-3′ under standard conditions with 1.5 mM $MgCl_2$. The reaction was performed with an initial 5-min denaturation at 95 °C, 35 cycles of 95 °C 30 s, 54 °C 30 s and 72 °C 30 s, then a 5-min extension at 72 °C. The US European Americans, African Americans and the Samoans were genotyped using a radiolabeled forward primer (HOBF), resolution of the products on 7% polyacrylamide gels followed by autoradiography and size estimation by comparison to a radiolabeled M13 sequencing ladder. All other samples were genotyped on the ABI 377 automated sequencer (Perkin Elmer) using HOBF labeled with one of the fluorescent dyes: FAM, HEX, or TET. The PCR products labeled with FAM and TET were diluted 1:20 with dH_2O and pooled for multiplex analysis. The computer programs Genscan and Genotyper were used for genotype analysis.

Amplimers from selected homozygous individuals were sequenced using either the dRhodamine Terminator Cycle Sequencing Ready Reaction kit or the dGTP Big Dye Terminator Ready Reaction kit (Perkin Elmer). Reaction products were resolved on the ABI 377 and were analyzed using Sequencher version 3.0 (Gene Codes).

Results

Among world populations, the LEP tetranucleotide repeat alleles can be subdivided into three general classes. These classes are illustrated in Fig. 1 for African Americans. Type 1 alleles range in size from 146–178 bp, while Type 3 alleles range between 210 and 254 bp. Type 2 alleles share approximately the same size range as Type 1 alleles, except that they are one base pair smaller, with size ranges between 165 and 193 bp. African American and African populations (Fig. 1, and Table 1) typically have a preponderance of Type 1 and 2 alleles and few Type 3 alleles. Among African Americans, the occurrence of Type 3 alleles can be explained by European admixture, and the same may be true of the population of Madagascar. European and Asian derived populations have predominantly Type 1 and 3 alleles, with more than half the alleles of Asian-derived populations (Hong Kong Chinese, Taiwanese Chinese, Samoans and Kota Kinabalu) belonging

Table 1 Allele frequencies for the LEP 3′-$(TTTC)_n$ polymorphism in various world populations

Population												
Allele size (bp)	USEA *n*=624	USAA *n*=600	CAR *n*=48	WAF *n*=82	MAD *n*=88	BR *n*=35	RUS *n*=96	HK *n*=43	TW *n*=73	SAM *n*=104	CR *n*=35	KK *n*=129
150	0.04	0.06	0.11	0.04	0.02	0.07	0.07	0.01	–	0.01	–	0.01
154	0.19	0.29	0.29	0.35	0.31	0.36	0.36	0.10	0.06	0.08	0.22	0.14
155	–	–	–	0.01	0.01	–	–	0.03	–	–	–	–
157	–	–	–	–	0.05	–	–	0.05	0.11	–	0.01	0.16
158	0.12	0.08	0.10	0.01	0.14	0.20	0.18	0.05	0.05	–	0.20	0.05
161	–	–	–	–	0.06	–	–	0.01	0.01	0.04	–	0.02
162	0.11	0.06	0.06	0.04	0.04	0.15	0.19	0.02	–	–	0.23	0.01
165	–	0.02	0.03	0.01	0.01	–	–	0.01	–	–	–	–
166	0.02	0.03	0.02	0.01	0.03	0.01	–	0.01	–	–	0.03	–
169	0.01	0.03	0.02	0.10	0.04	–	–	0.01	–	–	–	–
170	–	0.02	0.02	0.01	0.01	–	–	–	–	–	–	–
173	–	0.08	0.08	0.10	0.07	0.01	0.01	–	–	–	–	
174	–	0.015	–	0.04	0.02	–	0.01	–	–	–	–	–
176	–	–	0.01	0.01	–	–	–	–	–	–	–	–
177	–	0.11	0.13	0.10	0.05	–	–	–	–	–	–	–
178	–	0.01	–	0.03	0.01	–	–	–	–	–	–	–
181	–	0.06	0.05	0.08	0.01	–	–	–	–	–	–	–
185	–	0.03	–	0.01	0.01	–	–	–	–	–	–	–
189	–	0.01	0.01	–	–	–	–	–	–	–	–	–
218	0.03	–	–	0.01	0.04	0.01	0.01	0.01	0.05	–	0.01	–
222	0.08	0.01	–	–	0.01	0.04	0.01	0.06	0.10	0.03	0.09	0.11
226	0.09	0.02	0.02	–	0.01	0.05		0.15	0.12	0.16	0.03	0.09
230	0.10	0.03	–	–	0.02	–	0.05	0.15	0.23	0.20	0.04	0.14
234	0.10	0.01	–	–	0.03	0.04	0.04	0.15	0.13	0.27	0.04	0.15
238	0.05	–	–	–	–	0.01	–	0.13	0.06	0.15	–	0.05
242	0.03	–	–	–	–	–	0.04	0.02	0.04	0.06	0.03	0.01
246	0.02	–	–	–	–	–	–	0.01	0.01	–	–	–

to Type 3 (see Table 1 and Fig. 2). Discontinuities in the allele size distribution suggest that the LEP-(TTTC) polymorphisms did not evolve by simple expansion of the (TTTC)-track. In order to better characterize the molecular basis of the LEP-(TTTC) repeat variation, we sequenced the common alleles at this locus from mixed Europeans, African Americans and Samoans.

Direct sequence analysis in homozygous individuals showed that the Type 1 alleles were of the general form xxx$(TTTC)_n$Tyyy, where *n*=11 for allele 150, *n*=12 for allele 154, etc. The smaller Type 2 (Type 2a) alleles (165 and 173) were of the form xxx$(TTTC)_n$ –yyy, where *n*=15 for the 165 allele and *n*=17 for the 173 allele. The larger Type 2 (Type 2b) alleles (sizes 173–181) showed a deletion of a T in the fifth repeat giving the form xxx$(TTTC)_4$TTC$(TTTC)_n$Tyyy, where *n*=12 for allele 173, *n*=13 for allele 177, etc. The Type 3 alleles showed greater sequence variation with a generalized form xxx$(TTTC)_m$Y$(TTTC)_n$TT$(TTTC)_3$TTyyy, where *m*+*n*=25 for allele 222, *m*+*n*=26 for allele 226, etc. (Table 2).

The sequence of the Type 3 alleles revealed several interesting characteristics specific to the Type 3 alleles. First, all of the Type 3 alleles sequenced showed an insertion of a second T at the 3′ end of the repeat. All of the other alleles had either a single T (Type 1 and large Type 2b) or no T (Type 2a). Also, every Type 3 allele had an insertion of two Ts three units from the end of the repeat and an insertion of either a C or a T approximately 12–15 units into the repeat. Most of the alleles showed a C as the inserted base, but several of the smaller group 3 alleles clearly had a T in the same position. These additions to the sequence break the repeat region into three segments of normal TTTC units with the most 3′ of these segments being of a constant three-repeat unit length. The two more 5′ segments vary in length independently of one another, although they both remained in the range of 11–16 repeat units. Alleles of a given size could have different numbers of repeat units in (TTTC) segment one or two to give alleles of identical size, but different composition. For example, the 230 allele showed three different sequences: xxx$(TTTC)_{13}$C$(TTTC)_{14}$TT$(TTTC)_3$TTyyy, xxx$(TTTC)_{14}$ C$(TTTC)_{13}$TT $(TTTC)_3$TTyyy, and xxx$(TTTC)_{15}$ C$(TTTC)_{12}$TT$(TTTC)_3$TTyyy, but the first two segments add up to 27 in all cases. This heterogeneity in the sequence of same size alleles was confirmed by sequencing of several compound heterozygotes. From the sequencing of human LEP-(TTTC) alleles it is clear that the Type 3 alleles did not arise by a simple expansion of Type 1 alle-

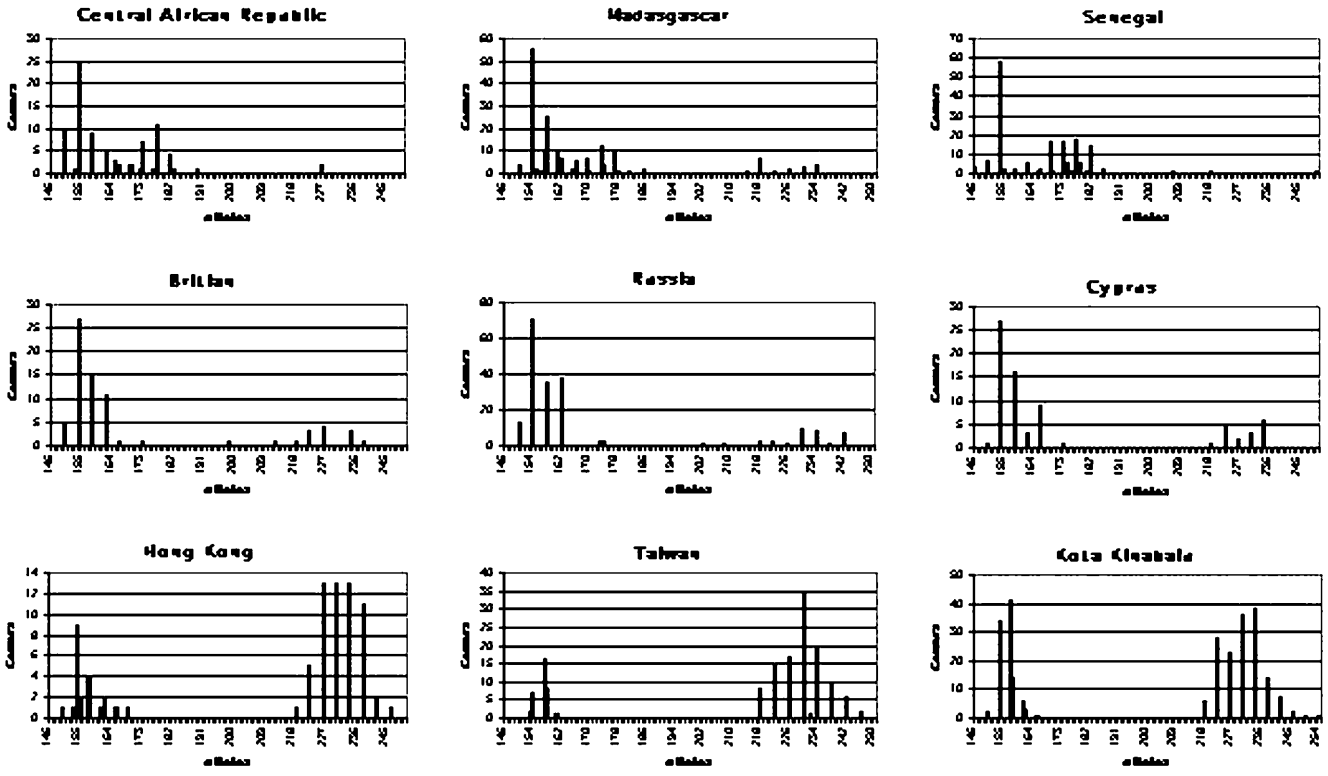

Fig. 2 Distribution of leptin alleles observed in world populations

Table 2 Generalized forms for the sequence of LEP-(TTTC) alleles in humans and great apes; *xxx* denotes the common upstream sequence and *yyy* denotes the common downstream sequence of the repeat region

Human	
Type 1:	xxx$(TTTC)_n$Tyyy
Type 2a:	xxx$(TTTC)_n$-yyy
Type 2b:	xxx$(TTTC)_4$TTC$(TTTC)_n$Tyyy
Type 3:	xxx$(TTTC)_m$Y$(TTTC)_n$TT$(TTTC)_3$TTyyy
Primate	
Orangutan:	xxx$(TTTC)_4$TTTTyyy
Gorilla:	xxxTTTCCTTC$(TTTC)_4$yyy
Chimpanzee:	xxx$(TTTC)_4$TTC$(TTTC)_4$TTTTyyy
Pygmy chimp:	xxx$(TTTC)_4$TTC$(TTTC)_6$TTTTyyy

les. The Type 2a alleles, which lack the 3′ terminal T are restricted to populations of African decent.

To further explore the organization of the LEP-(TTTC) locus, we sequenced multiple alleles from other great ape species, the orangutan (n=7), the gorilla (n=7), the common chimpanzee (n=26) and the pygmy chimpanzee (n=3). These results are summarized in Table 2.

Discussion

DNA sequence analysis and direct comparison of allele sizes in populations to alleles of known sequence reveals enormous variability at the LEP-(TTTC) locus and a complex population history. Alleles can be grouped roughly by size with the largest alleles being more complex. Type 1 alleles are shared by all study populations and alleles in this group differ by simple unit differences in the number of tandemly repeated (TTTC) units. Type 2 alleles fall into two classes, smaller Type 2a alleles overlapping with Type 1 alleles in size, but differing from Type 1 alleles by the absence of the T immediately 3′ of the final (TTTC) repeat unit.

Type 2a alleles are restricted to populations of African descent and may have arisen in Africa after the migration of the lineage ancestral to present day European and Asian populations out of Africa. Alternatively, Type 2a alleles may have existed prior to this divergence, but were not represented in the populations that migrated out of Africa or were lost from these populations by genetic drift. Type 1 alleles and Type 2a alleles all have one uninterrupted array of (TTTC) repeats, a pattern that was observed in all of the apes except *Pan paniscus* and *Pan troglodytes*.

The larger Type 2b and the Type 3 alleles appear to have arisen from a tandem duplication of a Type 1 allele and the similarity in structure and sequence of the Type 2b alleles to those of the *Pan* species suggests that this duplication occurred after the separation of the human/chimp ancestral line from that of the gorilla. A further partial duplication of a larger Type 2b allele could yield the Type 3 class of alleles. The Type 3 alleles appear to have arisen after the separation of the human and chimp lineages. The preponderance of Type 3 alleles in European and Asian populations and the paucity of these alleles in African- de-

rived populations, particularly those of continental Africans suggests that this group of alleles may have arisen after the separation of the European/Asian lineage from the larger African population. The presence and frequency of Type 3 alleles in African Americans is easily explained by known European admixture in this population group. The Type 3 alleles observed in Madagascar are probably of Asian origin. Prior to about AD 800, Madagascar was unoccupied. The first people to settle there were Austronesian-speakers originating from island Southeast Asia. They established the island as a slave-trading base and brought over Africans from the east coast of southern Africa as slaves. The present-day Madagascar population are a mixture of sub-Saharan African and island southeast Asian gene pools. This mix of African alleles and alleles or haplotypes observed primarily in southeastern Asia and the islands of the South Pacific has been observed at other loci (Soodyall et al. 1995; Hewitt et al. 1996; Martinson et al. 1996).

These analyses suggest a complex history at the LEP-(TTTC) locus that may preclude using this locus as a surrogate for functional variation in the leptin structural or other linked loci in association studies. Type 1, Type 2a, Type 2b and Type 3 alleles appear to have very different histories and apparently similar alleles at this locus may have a very different mutation history. Type 1 and Type 2a alleles overlap in size and in the absence of careful allele size estimates (i.e., within the binning error of programs typically used for estimating allele sizes on automatic sequencers) these two groups of alleles are easily misclassified. This misclassification could also influence identity by descent analyses. Within Type 3 alleles, limited sequence analysis revealed several alleles of equal size, which differed in the number of repeats at the 5′-most or central (TTTC) block. This heterogeneity may explain the failure of population-based studies to detect association between measures of obesity and the LEP-(TTTC) locus. Clearly, identity by state does not equal identity by descent for alleles in Type 3. Certainly, one could not expect association studies based on this locus to be replicated among geographically diverse populations, and association studies in admixed populations would be confounded by the complex history of this locus.

Acknowledgements This work was supported in part by N.I.H. grants GM45861(R.D.), HL52611(S.T.M.) and HL45778(S.M., R.E.F.).

References

Ahima RS, Flier JS (2000) Leptin. Ann Rev Physiol 62:413–437

Allison DB, Heo M (1998) Meta-analysis of linkage data under worst-case conditions: a demonstration using the human OB region. Genetics 148:859–865

Baile CA, Della-Fera MA, Martin RJ (2000) Regulation of metabolism and body fat mass by leptin. Ann Rev Nutr 20:105–127

Barrantes R, Smouse PE, Mohrenweiser HW, Gershowitz H, Azofeifa J, Arias TD, Neel JV (1990) Microevolution in lower Central America: genetic characterization of the Chibcha-speaking groups of Costa Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity. Am J Hum Genet 46:63–84

Borecki IB, Rice T, Perusse L, Bouchard C, Rao DC (1994) An exploratory investigation of genetic linkage with body composition and fatness phenotypes: the Quebec Family Study. Obes Res 2:213–219

Bray MS, Boerwinkle E, Hanis CL (1996) OB gene not linked to human obesity in Mexican American affected sib pairs from Starr County, Texas. Hum Genet 98:590–595

Campfield LA, Smith FJ (1998) Overview: neurobiology of OB protein (leptin). Proc Nutr Soc 57:429–440

Clement K, Garner C, Hager J, Philippi A, LeDuc C, Carey A, Harris TJ, Jury C, Cardon LR, Basdevant A, Demenais F, Guy-Grand B, North M, Froguel P (1996) Indication for linkage of the human OB gene region with extreme obesity. Diabetes 45:687–690

Deka R, Jin L, Shriver MD, Yu LM, DeCroo S, Hundrieser J, Bunker CH, Ferrell RE, Chakraborty R (1995) Population genetics of dinucleotide $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ polymorphisms in world populations. Am J Hum Genet 56:461–474

Duggirala R, Stern MP, Mitchell BD, Reinhart LJ, Shipman PA, Uresandi OC, Chung WK, Leibel RL, Hales CN, O'Connell P, Blangero J (1996) Quantitative variation in obesity-related traits and insulin precursors linked to the OB gene region on human chromosome 7. Am J Hum Genet 59:694–703

Hager J, Clement K, Francke S, Dina C, Raison J, Lahlou N, Rich N, Pelloux V, Basdevant A, Guy-Grand B, North M, Froguel P (1998) A polymorphism in the 5′ untranslated region of the human ob gene is associated with low leptin levels. Int J Obesity 22:200–205

Harris RB (2000) Leptin-much more than a satiety signal. Ann Rev Nutr 20:45–75

Hasstedt SJ, Hoffman M, Leppert MF, Elbein SC (1997) Recessive inheritance of obesity in familial non-insulin dependent diabetes mellitus, and lack of linkage to nine candidate genes. Am J Hum Genet 61:668–677

Hewitt R, Krause A, Goldman A, Campbell G, Jenkins T (1996) Beta-globin haplotype analysis suggests that a major source of Malagasy ancestry is derived from Bantu-speaking Negroids. Am J Hum Genet 58:1303–1308

Li WD, Reed DR, Lee JH, Xu W, Kilker RL, Sodam BR, Price RA (1999) Sequence variants in the 5′ flanking region of the leptin gene are associated with obesity in women. Ann Hum Genet 63:227–234

Mammes O, Betoulle D, Aubert R, Giraud V, Tuzet S, Petiet A, Colas-Linhart N, Fumeron F (1998) Novel polymorphisms in the 5′ region of the LEP gene: association with leptin levels and responses to low-calorie diet in human obesity. Diabetes 47: 487–489

Martinson JJ, Boyce AJ, Rees DC, Rugless MJ, Clegg JB (1996) α-globin genotype and haplotype distributions in the indigenous populations of island South East Asia and Madagascar: relationships with other populations. Am J Hum Genet 58 (suppl.): 1017

Martinson JJ, Hong L, Karanicolas R, Moore JP, Kostrikis LG (2000) Global distribution of the CCR2–64I/CCR5–59653T HIV-1 disease-protective haplotype. AIDS 14:483–489

Montague CT, Farooqi IS, Whitehead JP, Soos MA, Rau H, Wareham NJ, Sewter CP, Digby JE, Muhammed SN, Hurst JA, Cheetham CH, Earley AR, Barnett AH, Prins JB, O'Rahilly S (1997) Congenital leptin deficiency is associated with severe early-onset obesity in humans. Nature 387:903–908

Norman RA, Leibel RL, Chung WK, Powell-Kehoe L, Clua S, Knowler WC, Thompson DB, Bogardus C, Ravussin E (1996) Absence of linkage of obesity and energy metabolism to markers flanking homologues of rodent obesity genes in Pima Indians. Diabetes 45:1229–1232

Oksanen L, Ohman M, Heiman M, Kainulainen K, Kaprio J, Mustajoki P, Koivisto V, Koskenvuo M, Janne OA, Peltonen L, Kontula K (1997) Markers for the gene ob and serum leptin levels in human morbid obesity. Hum Genet 99:559–564

Reed DR, Ding Y, Xu W, Cather C, Green ED, Price RA (1996) Extreme obesity may be linked to markers flanking the human OB gene. Diabetes 45:691–694

Shintani M, Ikegami H, Yamato E, Kawaguchi Y, Fujisawa T, Nakagawa Y, Hamada Y, Ueda H, Miki T, Ogihara T (1996) A novel microsatellite polymorphism in the human OB gene: a highly polymorphic marker for linkage analysis. Diabetologia 39:1398–1401

Soodyall H, Jenkins T, Stoneking M (1995) 'Polynesian' mtDNA in the Malagasy. Nat Genet 10:377–378

Strobel A, Issad T, Camoin L, Ozata M, Strosberg AD (1998) A leptin missense mutation associated with hypogonadism and morbid obesity. Nat Genet 18:213–215

Zhang Y, Proenca R, Maffei M, Barone M, Leopold L, Friedman JM (1994) Positional cloning of the mouse *obese* gene and its human homologue. Nature 372:425–432