Published in final edited form as:

Gene. 2017 September 05; 627: 412–419. doi:10.1016/j.gene.2017.07.002.

A global evolutionary and metabolic analysis of human obesity gene risk variants

Joseph J. Castillo, Zachary S. Hazlett, Robert A. Orlando, and William S. Garver*

Department of Biochemistry and Molecular Biology, School of Medicine, University of New Mexico Health Sciences Center, Albuquerque, New Mexico, USA

Abstract

It is generally accepted that the selection of gene variants during human evolution optimized energy metabolism that now interacts with our obesogenic environment to increase the prevalence of obesity. The purpose of this study was to perform a global evolutionary and metabolic analysis of human obesity gene risk variants (110 human obesity genes with 127 nearest gene risk variants) identified using genome-wide association studies (GWAS) to enhance our knowledge of early and late genotypes. As a result of determining the mean frequency of these obesity gene risk variants in 13 available populations from around the world our results provide evidence for the early selection of ancestral risk variants (defined as selection before migration from Africa) and late selection of derived risk variants (defined as selection after migration from Africa). Our results also provide novel information for association of these obesity genes or encoded proteins with diverse metabolic pathways and other human diseases. The overall results indicate a significant differential evolutionary pattern for the selection of obesity gene ancestral and derived risk variants proposed to optimize energy metabolism in varying global environments and complex association with metabolic pathways and other human diseases. These results are consistent with obesity genes that encode proteins possessing a fundamental role in maintaining energy metabolism and survival during the course of human evolution.

Keywords

environment; evolution; metabolism; obesity; variants

1. Introduction

The most recent report from the Centers for Disease Control and Prevention indicates that 16.9% of children and adolescents (2 to 19 years of age) and 34.9% of adults (20 years of age and older) have obesity as defined by a body mass index the age adjusted 95th

*Corresponding author. Tel.: 1-505-272-4790. Fax 1-505-272-6587. wgarver@unm.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of interest

The authors declare no conflict of interest.

percentile and 30 kg/m^2 , respectively (Ogden et al., 2014). To determine the molecular basis of common obesity that has reached epidemic proportions in the United States and other developed countries, a large number of genome-wide association studies (GWAS) have been performed to identify obesity gene risk variants that increase susceptibility to this metabolic disease (Meyre et al., 2009; Thorleifsson et al., 2009; Willer et al., 2009; Scherag et al., 2010; Speliotes et al., 2010; Kilpelainen et al., 2011; Bradfield et al., 2012; Wen et al., 2012; Berndt et al., 2013; Bian et al., 2013; Monda et al., 2013; Locke et al., 2015; Minster et al., 2016). These GWAS and subsequent transferability studies performed in different populations around the world have now validated 110 human obesity genes with 127 nearest obesity gene risk variants (Hester et al., 2012; Ahmad et al., 2015; Hagg et al., 2015; Locke et al., 2015; Abadi et al., 2016). It is generally accepted that the obesity epidemic is a recent manifestation that has occurred during the past few decades and not all individuals or populations are adversely affected, thereby suggesting individual differences based on genetic variability and interactions with an obesogenic environment (Nakamura et al., 2015; Nettleton et al., 2015; Reddon et al., 2016).

Consistent with obesity gene risk variants interacting with an obesogenic environment to increase adiposity, James V. Neel proposed the thrifty genotype hypothesis based on positive selection (adaptation) of “thrifty genes” resulting from seasonal food shortages and episodic famines during the course of human evolution (Neel, 1962, 1999). Many years later John R. Speakman proposed the drift genotype hypothesis based on neutral selection (genetic drift) of “drifty genes” resulting from predation release due to the advent of fire and development of weapons, thereby increasing the upper limit of body weight set points (Speakman, 2008, 2013). In support of these two hypotheses, commentaries indicate that both positive and neutral selection of obesity gene risk variants have occurred during human evolution to optimize efficient storage of food energy for later use when food becomes limiting (Wells, 2006; Prentice et al., 2008; Henneberg et al., 2014). However, it should be noted there is continued and intense debate concerning how to explain the evolutionary and metabolic origins of the obesity epidemic (Wells, 2011; Genne-Bacon, 2014; Waalen, 2014).

We propose that by considering the two main evolutionary environments that have occurred in human history, those experienced by early hunter-gatherers before migration from Africa and later by agriculturalists after migration from Africa, novel insight will be gained concerning the selection of obesity gene risk variants, regardless of being labeled as being either thrifty or drift genotypes. Therefore, the purpose of this study was to perform a global evolutionary and metabolic analysis of human obesity gene risk variants (110 human obesity genes with 127 nearest gene risk variants) identified using genome-wide association studies (GWAS) to enhance our knowledge of early and late genotypes.

2. Materials and methods

2.1 Sample data

The human obesity gene risk variants were obtained from several GWAS (Meyre et al., 2009; Thorleifsson et al., 2009; Willer et al., 2009; Scherag et al., 2010; Speliotes et al., 2010; Kilpelainen et al., 2011; Bradfield et al., 2012; Wen et al., 2012; Berndt et al., 2013; Bian et al., 2013; Monda et al., 2013; Locke et al., 2015; Minster et al., 2016). These GWAS

and subsequent transferability studies performed in diverse populations from around the world have identified gene risk variants associated with human obesity (Hester et al., 2012; Ahmad et al., 2015; Hagg et al., 2015; Locke et al., 2015; Abadi et al., 2016). The sample data consists of 110 human obesity genes (56 obesity genes with ancestral risk variants and 54 obesity genes with derived risk variants). For the 110 human obesity genes, there are a total of 127 nearest risk variants represented by 64 nearest ancestral risk variants (41 genic and 23 nongenic) and 63 nearest derived risk variants (42 genic and 21 nongenic). Therefore, the relative number of human obesity genes with ancestral risk variants and derived risk variants, in addition to genic and nongenic risk variants, are of similar number. The human obesity genes with ancestral risk variants (Supplementary material Table 1) and derived risk variants (Supplementary material Table 2) are arranged according to the chromosomal cytogenic band. Additional information provided in these two tables include the chromosome number, allele pair (ancestral or derived variant designation), risk variant, Fst (a measure of genetic variation across populations), and variant frequency in each population. The variant frequency in 13 available populations from around the world was collected from the National Center for Biotechnical Information (NCBI) online database, which was generated by the 1000 Genomes Project (Consortium, 2015). The abbreviations for these 13 populations are as follows: YRI, Yoruba in Ibadan, Nigeria (Sub-Saharan African); AFR, African American descent; MKK, Maasai in Kinyawa, Kenya; ASW, African ancestry in Southwest USA; JPT, Japanese in Tokyo, Japan; CHB, Han Chinese in Beijing; CEU, Northern and Western European ancestry in Utah; GIH, Gujarati Indians in Houston, Texas; MEX, Mexican ancestry in Los Angeles, California; EUR, European American; CHD, Chinese ancestry in Denver, Colorado; CHN, Chinese American descent; and TSI, Tuscans in Italy.

2.2 Chromosomal ideogram of obesity genes

The chromosomal loci for the 110 human obesity genes are provided using a chromosomal ideogram. The ideogram was generated using tools available on the NCBI genome decoration page (<http://www.ncbi.nlm.nih.gov/genome/tools/gdp>).

2.3 Population gene variant frequencies

The human obesity gene risk variant frequencies in 13 available populations were collected and statistical means were calculated and graphed using box and whisker plots for the ancestral and derived risk variants. The ancestral risk variants are recognized as being phylogenetically older, while the derived risk variants are recognized as being phylogenetically younger. Therefore, the mean frequency of obesity gene ancestral and derived risk variants in 13 available populations should provide evidence for the selection of these risk variants. With respect to the box and whisker plots, the middle horizontal line within boxes represents the mean frequency, the boxes represent plus and minus one standard deviation, and the whiskers represent the upper and lower range. A significant difference calculated for mean frequencies are indicated by different lower case letters among populations. The graph for obesity gene ancestral risk variants was arranged according to decreasing mean frequency. The same order of these populations was thereafter maintained for subsequent graphs to allow direct comparison between mean frequencies of risk variants. All population gene variant frequencies were tested by one-way ANOVA to

determine whether or not any of the 13 population means were significantly different, followed by each population being compared to all of the other populations using a t-test (two-sample assuming unequal variances) to determine statistically significantly different means by voiding the null hypothesis of similar means.

2.4 Gene variant composite of multiple signals

The chromosomal loci of obesity gene risk variants were compared to regions of chromosomes determined to be of high value with regard to the composite multiple signals (CMS) technique for haplotype scoring (Grossman et al., 2010). The CMS takes into account length of the haplotype region, the frequency of neighboring ancestral or derived gene variants compared to that which would be expected for genetic drift and the difference between gene variant frequencies in other populations. Any obesity gene risk variants located within a region to be of high CMS value was indicated as to which population was observed having a high CMS score at that locus because CMS hotspots vary among populations due to divergent evolutionary history. The 127 nearest gene risk variants were compared to CMS hotspots for 4 available populations indicated as Northern and Western European ancestry in Utah (CEU), Yoruba in Ibadan (YRI), Japanese in Tokyo (JPT), and Han Chinese in Beijing (CHB).

2.5 KEGG Mapper database

The Kyoto Encyclopedia of Genes and Genomes (KEGG) Mapper database (<http://genome.jp/kegg>) was developed to provide an assessment for association of genes with metabolic pathways and human diseases (Kanehisa et al., 2000; Kanehisa et al., 2016). The obesity genes were queried using the KEGG Mapper database and associations among obesity genes with ancestral risk variants or derived risk variants were organized in descending order based on the number of hits.

2.6 Genetic variation among populations

Genetic variation among populations can be quantified using the F_{st} measure. This statistic measures the difference between a variant frequency in a population versus the variant frequency in other populations (Wright, 1950). It is calculated by subtracting the heterozygosity of a sub-population (H_s ; equivalent to $2p_s q_s$) from the heterozygosity of the total population, in this case the average of the 13 available populations (H_t ; equivalent to $2p_t q_t$), then dividing the difference by H_t ($\{H_t - H_s\}/H_t$). In this way, we can determine variance of each SNP frequency in a population compared to all populations as a whole. Genetic variation among populations (F_{st}) was calculated for each variant frequency. Variant frequencies within a population displaying an F_{st} higher than 0.5 or lower than -0.5 were tallied (implying a greater than 50% degree of genetic variation compared to the average frequency in all populations combined) and then divided by the number of variants with available frequencies for that population from the 1000 genomes project.

3. Results

3.1 Chromosomal ideogram of obesity genes

The chromosomal loci for the 110 obesity genes are provided using a chromosomal ideogram (Fig. 1). The obesity genes with ancestral and derived risk variants were distributed among most of the 22 chromosomes on both p and q regions. The X and Y chromosomes had no obesity genes.

3.2 Obesity gene ancestral risk variant frequencies

The obesity gene ancestral risk variant frequencies in each of the 13 available populations were combined and statistical means, standard deviations, and ranges were provided using box and whisker plots. First, with respect to the 64 ancestral risk variants, the Yoruba in Ibadan (YRI) had significantly higher mean frequencies compared to other populations except for the population of African American descent (AFR) (Fig. 2A). The remaining populations were not significantly different from each other except for Tuscans in Italy (TSI). Second, with respect to the 41 ancestral risk variants residing within genic regions, the four populations of African or African American descent (YRI, AFR, MKK, and ASW) had significantly higher mean frequencies compared to the remaining populations (Fig 2B). Third, with respect to the 23 ancestral risk variants residing within nongenic regions, the Yoruba in Ibadan (YRI) had significantly higher mean frequencies compared to other populations except for two populations of African or African American descent (AFR and MKK) (Fig. 2C). Finally, the populations of African or African American descent (YRI, AFR, MKK, and ASW) had noticeably smaller standard deviations and ranges compared to other populations. Therefore, these results provide evidence of early (prior to human migration from Africa) selection for obesity gene ancestral obesity risk variants.

3.3 Obesity gene derived risk variant frequencies

The obesity gene derived risk variant frequencies in each of the 13 available populations were combined and statistical means, standard deviations, and ranges were provided using box and whisker plots. First, with respect to the 63 derived risk variants, the Yoruba in Ibadan (YRI) had significantly lower mean frequencies compared to other populations except for the other populations of African and African American descent (AFR, MKK, and ASW) (Fig. 3A). The remaining populations were not significantly different from each other except for Northern and Western European ancestry in Utah (CEU) and Tuscans in Italy (TSI). Second, with respect to the 42 ancestral risk variants residing within genic regions, the four populations of African or African American descent (YRI, AFR, MKK, and ASW) had significantly lower mean frequencies compared to the remaining populations (Fig 3B). The remaining populations were not significantly different from each other except for Northern and Western European ancestry in Utah (CEU). Third, with respect to the 21 derived risk variants residing within nongenic regions, the population of Yoruba in Ibadan (YRI) and African American descent (AFR) had significantly lower mean frequencies compared to other populations except for populations of African descent (MKK and ASW) (Fig. 3C). Finally, the populations of African or African American descent (YRI, AFR, MKK, and ASW) had noticeably smaller standard deviations and ranges compared to other

populations. Therefore, these results provide evidence of late (after human migration from Africa) selection for obesity gene derived obesity risk variants.

3.4 Gene variant composite of multiple signals

The chromosomal loci of obesity gene risk variants were compared to regions of chromosomes within populations determined to be of high value with regard to the composite of multiple signals (CMS) for haplotype scoring. With respect to the 127 obesity gene risk variants evaluated in this study, only a single (1) obesity gene risk variant was found to be located within a chromosomal region determined to have high a CMS score. The obesity gene risk variant is an ancestral variant (PDXDCI, rs4985155) located within a known CMS chromosomal hotspot for the population of Yoruba in Ibadan (YRI). This result provides evidence of selection for this ancestral risk variant prior to human migration from Africa.

3.5 KEGG Mapper database

The 56 obesity genes with 64 nearest ancestral risk variants queried using the KEGG Mapper database resulted in a total of 76 hits associated with metabolic pathways and human diseases (Table 1). With respect to the first category denoted as “metabolism” that provides the most pertinent information in understanding the physiological basis of obesity, 6 hits were noted for carbohydrate and lipid metabolism (starch and sucrose metabolism, adipocyte lipolysis), regulation of energy metabolism (insulin secretion and signaling) and different forms of diabetes mellitus (type II diabetes mellitus and insulin resistance) known to represent a disease of energy metabolism. In contrast, the 54 obesity genes with 63 nearest derived risk variants queried using the KEGG Mapper database resulted in a total of 146 hits associated with metabolic, cellular, and homeostatic pathways and human diseases (Table 2). Again, with respect to the first category denoted as “metabolism” that provides the most pertinent information in understanding the physiological basis of obesity, 35 hits were noted for food digestion (pancreatic secretion, gastric acid secretion, salivary secretion, bile secretion), regulation of energy metabolism (glucagon signaling, insulin secretion and signaling), lipid metabolism (aldosterone synthesis, adipocyte lipolysis, steroid hormone synthesis, fatty acid elongation, synthesis of unsaturated fatty acids, thyroid hormone synthesis), and insulin resistance. Therefore, these results indicated that obesity genes with ancestral and derived risk variants have some overlapping yet distinct differences in association with metabolic pathways and human diseases.

3.6 Genetic variation among populations

Populations with high degrees of genetic variation from the average obesity susceptibility variant frequencies are provided (Fig. 4). With respect to ancestral variants, the populations of Yoruba in Ibadan (YRI), Han Chinese in Beijing (CHB), Japanese in Tokyo, Japan (JPT), Chinese American descent (CHN), and Chinese ancestry in Denver, Colorado (CHD) display more than 25% of variants with a high degree (>50%) of genetic variation from the average (Fig. 4A). When considering the derived variants, the same populations show >25% of variants with a high degree of genetic variation, with the added inclusion of the populations of African American descent, (AFR), African ancestry in Southwest USA (ASW), and Maasai in Kinyawa, Kenya (MKK) (Fig. 4B). For those three populations

(AFR, ASW, and MKK), there is a 6–16% increase in the percentage of variants with high genetic variation when comparing ancestral to derived variants. The combined frequencies of variants of high genetic variation reflect what is seen in the ancestral and derived frequencies (Fig. 4C). Groups that have been geographically isolated from others during human evolution and global migration (YRI, CHB, JPT, CHN, and CHD) demonstrate the highest levels of genetic variation, as would be expected from limited genetic admixture. Groups with observable differences in genetic variation between ancestral and derived variant frequencies (AFR, ASW, and MKK; all of African descent but now in an environment of genetic admixture) support the hypothesis that these variants were selected in different evolutionary environments.

4. Discussion

The selection of obesity gene risk variants during the course of human evolution has shaped the efficiency of our digestion, absorption, and assimilation of nutrients to optimize energy metabolism involved in storage of energy for later use when food becomes limiting (Alves et al., 2002; Caetano-Anolles et al., 2009). However, since most developed populations no longer experience seasonal food shortages and episodic famines, these obesity gene risk variants now interact with our obesogenic environment, characterized by energy dense foods and a sedentary lifestyle, to increase the prevalence of obesity (Chakravarthy et al., 2004). The purpose of this study was to perform a global evolutionary and metabolic analysis of human obesity gene risk variants (110 human obesity genes with 127 nearest gene risk variants) identified using genome-wide association studies (GWAS) to enhance our knowledge of early and late genotypes. As a result of determining the mean frequency of these obesity gene risk variants in 13 available populations from around the world our results provide evidence for the early selection of ancestral risk variants (defined as selection before migration from Africa) and late selection of derived risk variants (defined as selection after migration from Africa). Our results also provide novel information for association of these obesity genes or encoded proteins with diverse metabolic pathways and other human diseases. The overall results indicate a significant differential evolutionary pattern for the selection of obesity gene ancestral and derived risk variants proposed to optimize energy metabolism in varying global environments and complex association with metabolic pathways and other human diseases. These results are consistent with obesity genes that encode proteins possessing a fundamental role in maintaining energy metabolism and survival during the course of human evolution.

To begin our discussion, it is important to consider what factors other than seasonal food shortages and episodic famines may have contributed to the selection of obesity gene risk variants. With respect to obesity gene ancestral risk variants selected early in human evolution, it has been reported that hunter-gatherers stalked prey in a manner similar to wolves, which is referred to as endurance hunting, in procuring food energy to survive and reproduce (Prentice, 2001; Bramble et al., 2004; Prentice et al., 2005; Lieberman, 2014). This method of hunting involved the meticulous and persistent tracking of prey that eventually became exhausted after running for hours in the heat. Once the prey becomes exhausted, the hunter could more easily sacrifice the animal to obtain a rich source of food energy. The enhanced ability of hunter-gatherers to utilize endurance hunting, which

required a highly efficient use of stored energy in the form of triacylglycerol, is proposed to have been acquired through selection among early humans and consistent with obesity gene ancestral risk variants. Moreover, after human migration from Africa into cooler climate regions, which may not have been conducive to endurance hunting, humans developed other methods for hunting such as advanced weaponry and horseback riding that reduced selective pressures acting upon obesity gene ancestral risk variants, thereby decreasing frequency of these risk variants due to genetic drift. Nonetheless, the relatively high frequency of obesity gene ancestral risk variants remaining among modern humans in our current obesogenic environment contributes to the obesity epidemic.

Seasonal food shortages and episodic famines are proposed to have occurred later in human evolution (100,000–40,000 years ago) and particularly after dawning of the agricultural era (within ~12,000 years ago) as a result of frequent catastrophic crop failures causing selection of obesity gene derived risk variants that optimized energy storage in these different environmental conditions (Prentice et al., 2005). While some have argued that the mortality occurring during these periods of food shortage and episodic famine may not have been severe enough to select for derived risk variants, others contend that decreased fertility may have been a strong enough selective force to generate pressure for the increase in population frequency of derived risk variants (Prentice et al., 2008; Speakman, 2008). Since selective pressures must have varied among different populations in the form of numerous chronologically and geologically separated events, different obesity gene derived risk variants were selected for, thus differing in frequency from one population to another. This is evidenced in modern times by a recent study performed with 9,416 individuals in 14 European countries indicating that although environmental differences masked genetic differentiation for BMI at least 8% of the captured additive genetic variance for BMI was reflected in population genetic differences (Robinson et al., 2015). Further evidence that endurance hunting may have initially selected for obesity gene ancestral risk variants is illustrated by populations that subsequently depended on agricultural methods for millennia, such as the Europeans, who now have a significantly lower frequency of these obesity gene ancestral risk variants compared to individuals from sub-Saharan Africa. As stated above, the selection for obesity gene ancestral and derived risk variants during human evolution is now proposed to be a hindrance for individuals living in our modern obesogenic environment due to an interaction between an energy rich diet and sedentary lifestyle, resulting in the heightened ability to store and maintain energy. Further evidence indicates that obesity gene risk variants may interact with other obesity gene risk variants and/or modifying gene risk variants through epistasis to influence population differences in the prevalence in obesity (Casazza et al., 2011).

With regard to the CMS metric used in our study only the PDXDC1 obesity gene was identified. The ancestral risk variant located in a region of high CMS score was identified in the population from Yoruba in Ibadan (YRI), thereby supporting the hypothesis that this obesity gene ancestral risk variant was selected during the hunter-gatherer era of human development. It should also be noted that not all of the obesity gene ancestral and derived risk variants, which reside in both coding and noncoding regions, may represent causal variants. This is due to the fact that these ancestral and derived risk variants were identified using GWAS which may have detected variants in linkage disequilibrium with the causal

variants nearby on the same chromosome (Hagg et al., 2015). However, the likelihood that none of these variants can be described as causal is unlikely since the case for “missing heritability” may be accounted for through imputed variants with any remaining heritability accounted for by gene interactions (Llewellyn et al., 2013; Yang et al., 2015).

With regard to genetic variation (F_{st}), it is interesting that the populations of African American descent (AFR), African ancestry in Southwest USA (ASW), and Maasai in Kinyawa, Kenya (MKK) show an increase in genetic variation when comparing the derived variants to the ancestral variants. This implies that the genetic variation of ancestral variants is limited due to their shared origin in a hunter-gatherer environment, while the circumstances that selected for derived variants differed from population to population, thus generating variation from group to group. Continued investigation of these obesity gene ancestral and derived gene risk variants identified using GWAS remains important since they are candidates for causal variants and, at least, act as a reflection of the linked causal variants. Recent advances in the emerging fields of epigenetics, whole genome sequencing, and genome editing, along with other available technologies, must be used to identify the causal variants and physiological mechanisms responsible for predisposition to obesity (Claussnitzer et al., 2015; Wu et al., 2016). It is possible that the list of obesity gene variants may grow as more genomic data from other populations becomes available (Southam et al., 2009).

In conclusion, our study provides an evolutionary and metabolic perspective of obesity genes with ancestral and derived risk variants that model early and late genotypes suspected to interact with the obesogenic environment to increase adiposity and weight gain responsible for the current obesity epidemic. As a result of determining the mean frequency of obesity gene ancestral and derived risk variants in 13 available populations, our results provide evidence for the early selection of ancestral risk variants (before migration from Africa) and late selection of derived risk variants (after migration from Africa).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project was supported in part by the National Center for Research Resources and the National Center for Advancing Translational Sciences of the National Institutes of Health through grant number 8UL1TR000041 and private funding through the University of New Mexico Foundation for the investigation of genetic and metabolic diseases.

Abbreviations

CMS	composite multiple signals
GWAS	genome-wide association study
KEGG	Kyoto Encyclopedia of Genes and Genomes
NCBI	Center for Biotechnology Information

References

- Abadi A, et al. Assessing the effects of 35 European-derived BMI-associated SNPs in Mexican Children. *Obesity*. 2016; 24:1989–1995. [PubMed: 27486100]
- Ahmad S, et al. Physical activity, smoking, and genetic predisposition to obesity in people from Pakistan: the PROMIS study. *BMC Med Genet*. 2015; 16:114. [PubMed: 26683835]
- Alves R, Ghaleil RAG, Stern JS. Evolution of enzymes in metabolism: A network perspective. *J Mol Biol*. 2002; 320:751–770. [PubMed: 12095253]
- Berndt SI, et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet*. 2013; 45:501–512. [PubMed: 23563607]
- Bian L, et al. MAP2K3 is associated with body mass index in American Indians and Caucasians and may mediate hypothalamic inflammation. *Hum Mol Genet*. 2013; 22:4438–4449. [PubMed: 23825110]
- Bradfield JP, et al. A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat Genet*. 2012; 44:526–531. [PubMed: 22484627]
- Bramble DM, Lieberman DE. Endurance running and the evolution of Homo. *Nature*. 2004; 432:345–352. [PubMed: 15549097]
- Caetano-Anolles G, et al. The origin and evolution of modern metabolism. *Int. J. Biochem. Cell Biol*. 2009; 41:285–297.
- Casazza K, Hanks LJ, Beasley M, Fernandez JR. Beyond thriftiness: Independent and interactive effects of genetic and dietary factors on variations in fat disposition and distribution across populations. *Am J Phys Anthro*. 2011; 145:181–191.
- Chakravarthy MV, Booth FW. Eating, exercise, and “thrifty” genotypes: Connecting the dots toward and evolutionary understanding of modern chronic diseases. *J Appl Physiol*. 2004; 96:3–10. [PubMed: 14660491]
- Claussnitzer M, et al. FTO obesity variant circuitry and adipocyte browning in humans. *N Eng J Med*. 2015; 373:895–907.
- Consortium, The 1000 Genomes Project. A global reference for human genetic variation. *Nature*. 2015; 68:68–74.
- Genne-Bacon EA. Thinking evolutionarily about obesity. *Yale J Biol Med*. 2014; 87:99–112. [PubMed: 24910556]
- Grossman SR, et al. A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science*. 2010; 327:883–886. [PubMed: 20056855]
- Hagg S, et al. Gene-based meta-analysis of genome-wide association studies implicates new loci involved in obesity. *Hum Mol Genet*. 2015; 24:6849–6860. [PubMed: 26376864]
- Henneberg M, Grantham J. Obesity-A natural consequence of human evolution. *Anthropol Rev*. 2014; 77:1–10.
- Hester JM, et al. Implication of European-derived adiposity loci in African Americans. *Int J Obes*. 2012; 36:465–473.
- Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucl Acid Res*. 2000; 28:27–30.
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucl Acid Res*. 2016; 44:457–462.
- Kilpelainen TO, et al. Genetic variation near IRS1 associates with reduced adiposity and an impaired metabolic profile. *Nat Genet*. 2011; 43:753–760. [PubMed: 21706003]
- Lieberman DE. Human locomotion and heat loss: An evolutionary perspective. *Comp Physiol*. 2014; 5:99–117.
- Llewellyn CH, Trzaskoski M, Plomin R, Wardle J. Finding the missing heritability in pediatric obesity: the contribution of genome-wide complex trait analysis. *Int J Obes*. 2013; 37:1506–1509.
- Locke AE, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015; 518:197–206. [PubMed: 25673413]
- Meyre D, et al. Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat Genet*. 2009; 41:157–159. [PubMed: 19151714]

- Minster RL, et al. A thrifty variant in CREBRF strongly influences body mass index in Samoans. *Nat Genet.* 2016; 48:1049–1054. [PubMed: 27455349]
- Monda KL, et al. A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. *Nat Genet.* 2013; 45:690–696. [PubMed: 23583978]
- Nakamura S, et al. Gene-environment interactions in obesity: Implication for future applications in preventive medicine. *J Hum Genet.* 2015; 61:317–322. [PubMed: 26657934]
- Neel JV. Diabetes mellitus: a “thrifty” genotype rendered detrimental by “progress”? *Am J Hum Genet.* 1962; 14:353–362. [PubMed: 13937884]
- Neel JV. The “Thrifty Genotype” in 1998. *Nutr Rev.* 1999; 57:S2–S9. [PubMed: 10391020]
- Nettleton JA, et al. Gene × dietary pattern interactions in obesity: Analysis of up to 68,317 adults of European ancestry. *Hum Mol Genet.* 2015; 24:4728–4738. [PubMed: 25994509]
- Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011–2012. *J Am Med Asso.* 2014; 311:806–814.
- Prentice AM. Fires of life: The struggles of an ancient metabolism in a modern world. *Nutr Bull.* 2001; 26:13–27.
- Prentice AM, Rayco-Solon P, Moore SE. Insights from the developing world: Thrifty genotypes and thrifty phenotypes. *Proc Nutr Soc.* 2005; 64:153–161. [PubMed: 15960860]
- Prentice AM, Hennig BJ, Fulford AJ. Evolutionary origins of the obesity epidemic: Natural selection of thrifty genes or genetic drift following predation release? *Int J Obes.* 2008; 32:1607–1610.
- Reddon H, et al. Physical activity and genetic predisposition to obesity in a multiethnic longitudinal study. *Sci Rep.* 2016; 6:18672. [PubMed: 26727462]
- Robinson MR, et al. Population genetic differentiation of height and body mass index across Europe. *Nat Genet.* 2015; 47:1357–1362. [PubMed: 26366552]
- Scherag A, et al. 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.* 2010; 6:e10000916.
- Southam L, et al. Is the thrifty genotype hypothesis supported by evidence based on confirmed type 2 diabetes and obesity susceptibility variants. *Diabetologia.* 2009; 52:1846–1851. [PubMed: 19526209]
- Speakman JR. Thrifty genes for obesity, an attractive but flawed idea, and an alternative perspective: the ‘drifty gene’ hypothesis. *Int J Obes.* 2008; 64:1611–1617.
- Speakman JR. Evolutionary perspectives on the obesity epidemic: Adaptive, maladaptive, and neutral viewpoints. *Annu Rev Nutr.* 2013; 33:289–317. [PubMed: 23862645]
- Speliotes EK, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet.* 2010; 42:937–948. [PubMed: 20935630]
- Thorleifsson G, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet.* 2009; 41:18–24. [PubMed: 19079260]
- Wahlen J. The genetics of human obesity. *Trans Res.* 2014; 164:293–301.
- Wells JCK. The evolution of human fatness and susceptibility to obesity: An ethological approach. *Biol Rev.* 2006; 81:183–205. [PubMed: 16677431]
- Wells JCK. The thrifty phenotype: An adaptation in growth or metabolism? *Am J Hum Biol.* 2011; 23:65–75. [PubMed: 21082685]
- Wen W, et al. Meta-analysis identifies common variants associated with body mass index in east Asians. *Nat Genet.* 2012; 44:307–311. [PubMed: 22344219]
- Willer CJ, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet.* 2009; 41
- Wright S. Genetical structure of populations. *Nature.* 1950; 166:247–249. [PubMed: 15439261]
- Wu C, Arora P. Noncoding genome-wide association studies variant for obesity: Inroads into mechanism. *J Am Heart Assoc.* 2016; 5:e003060. [PubMed: 27436305]
- Yang J, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body weight. *Nat Genet.* 2015; 47:1114–1120. [PubMed: 26323059]

Fig. 1. Chromosomal ideogram of human obesity genes. The chromosomal loci for the 110 obesity genes are provided using a chromosomal ideogram. The obesity genes with ancestral risk variants are localized adjacent to chromosomal loci denoted by gray arrowheads, whereas the obesity genes with derived risk variants are localized adjacent to chromosomal loci denoted with black arrowheads.

Fig. 2. Obesity gene ancestral risk variant frequencies. Box and whisker plots represent the mean frequency, standard deviation, and range of obesity gene ancestral risk variant frequencies in 13 available populations. The middle horizontal lines within boxes represent the mean frequency, boxes represent plus and minus one standard deviation, and the whiskers represent the range. A significant difference for the mean frequency of a population is indicated by different lower case letters. (A) Total obesity gene ancestral risk variant frequencies (64 variants), (B) genic obesity gene ancestral risk variant frequencies (41

variants), and (C) nongenetic obesity ancestral risk variant frequencies (23 variants) are provided. The abbreviations for 13 populations are as follows: YRI, Yoruba in Ibadan, Nigeria (Sub-Saharan African); AFR, African American descent from the Coriell Cell Repository; MKK, Maasai in Kinyawa, Kenya; ASW, African ancestry in Southwest USA; JPT, Japanese in Tokyo, Japan; CHB, Han Chinese in Beijing; CEU, Northern and Western European ancestry in Utah; GIH, Gujarati Indians in Houston, Texas; MEX, Mexican ancestry in Los Angeles, California; EUR, European American; CHD, Chinese ancestry in Denver, Colorado; CHN, Chinese American descent from the Coriell Cell Repository; and TSI, Tuscans in Italy.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 3. Obesity gene derived risk variant frequencies. Box and whisker plots represent the mean frequency, standard deviation, and range of obesity gene ancestral risk variant frequencies in 13 available populations. The middle horizontal lines within boxes represent the mean frequency, boxes represent plus and minus one standard deviation, and the whiskers represent the range. A significant difference for the mean frequency of a population is indicated by different lower case letters. (A) Total obesity gene derived risk variant frequencies (63 variants), (B) genic obesity gene derived risk variant frequencies (42

variants), and (C) nongenetic obesity derived risk variant frequencies (21 variants) are provided. The abbreviations for 13 populations are as follows: YRI, Yoruba in Ibadan, Nigeria (Sub-Saharan African); AFR, African American descent from the Coriell Cell Repository; MKK, Maasai in Kinyawa, Kenya; ASW, African ancestry in Southwest USA; JPT, Japanese in Tokyo, Japan; CHB, Han Chinese in Beijing; CEU, Northern and Western European ancestry in Utah; GIH, Gujarati Indians in Houston, Texas; MEX, Mexican ancestry in Los Angeles, California; EUR, European American; CHD, Chinese ancestry in Denver, Colorado; CHN, Chinese American descent from the Coriell Cell Repository; and TSI, Tuscans in Italy.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 4.

Genetic variation among populations. Bar graph representing the frequency in percent of variants with a greater than 50% difference from the average obesity gene variants' frequency: (A) Ancestral obesity gene variants, (B) derived obesity gene variants, and (C) combined obesity gene variants. The abbreviations for 13 populations are as follows: YRI, Yoruba in Ibadan, Nigeria (Sub-Saharan African); AFR, African American descent; MKK, Maasai in Kinyawa, Kenya; ASW, African ancestry in Southwest USA; JPT, Japanese in Tokyo, Japan; CHB, Han Chinese in Beijing; CEU, Northern and Western European

ancestry in Utah; GIH, Gujarati Indians in Houston, Texas; MEX, Mexican ancestry in Los Angeles, California; EUR, European American; CHD, Chinese ancestry in Denver, Colorado; CHN, Chinese American descent; and TSI, Tuscans in Italy.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Human obesity genes with ancestral risk variants

Pathway and/or disease phenotype	Hits
<u>Metabolism</u>	
Adipocytokine signaling pathway	2
Metabolic pathways	1
Insulin signaling pathway	1
Regulation of lipolysis in adipocytes	1
Insulin resistance	1
Type II diabetes mellitus	1
Non-alcoholic fatty liver disease	1
Starch and sucrose metabolism	1
Aldosterone-regulated NA reabsorption	1
<u>Pathology/Infection</u>	
Pathways in cancer	2
Transcriptional misregulation in cancer	2
Huntington's disease	2
Basal cell carcinoma	2
Proteoglycans in cancer	2
Inflammatory bowel disease	2
Rheumatoid arthritis	1
Salmonella infection	1
Thyroid cancer	1
Prostate cancer	1
Cocaine addiction	1
Hepatitis B	1
Legionellosis	1
Toxoplasmosis	1
Breast cancer	1
Alcoholism	1
Parkinson's disease	1
Malaria	1
Influenza A	1
Pertussis	1
Leishmaniasis	1
MicroRNAs in cancer	1
Arrhythmogenic right ventricular cardiomyopathy	1
Chagas disease	1
Acute myeloid leukemia	1
Measles	1
Pathogenic Escherichia coli infection	1
Tuberculosis	1

Pathway and/or disease phenotype	Hits
Amoebiasis	1
Endometrial cancer	1
Colorectal cancer	1
Tuberculosis	1
Amoebiasis	1
Endometrial cancer	1
Colorectal cancer	1
<u>Development/Homeostasis</u>	
Melanogenesis	2
Th1 and Th2 cell differentiation	1
Aldosterone synthesis and secretion	1
Spliceosome	1
<u>Cell signaling/maintenance</u>	
PI3K-Akt signaling pathway	3
cAMP signaling pathway	3
Ubiquitin mediated proteolysis	2
mTOR signaling pathway	2
MAPK signaling pathway	2
Cell adhesion molecules	2
mRNA surveillance pathway	1
Hippo signaling pathway	1
Lysosome	1
Protein processing in ER	1
ErbB signaling pathway	1
NOD-like receptor signaling pathway	1
Calcium signaling pathway	1
cGMP-PKG signaling pathway	1
Prolactin signaling pathway	1
Wnt signaling pathway	1
NF-kappa B signaling pathway	1
Cytokine-cytokine receptor interaction	1
Phagosome	1
FoxO signaling pathway	1
Adherens junction	1
Hedgehog signaling pathway	1
Rap1 signaling pathway	1
Toll-like receptor signaling pathway	1
Gap junction	1
Longevity regulating pathway	1
AMPK signaling pathway	1
Oxytocin signaling pathway	1
Endocytosis	1

Pathway and/or disease phenotype	Hits
Jak-STAT signaling pathway	1
HIF-1 signaling pathway	1

The KEGG Mapper database was developed to provide an assessment for association of genes with metabolic pathways and human diseases.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Human obesity genes with derived risk variant associated pathways

Pathway and/or disease phenotype	Hits
<u>Metabolism</u>	
Metabolic pathways	4
Aldosterone synthesis and secretion	4
Purine metabolism	3
Insulin secretion	2
Vasopressin-regulated water reabsorption	2
Thyroid hormone synthesis	2
Regulation of lipolysis in adipocytes	2
Insulin resistance	1
Glucagon signaling pathway	1
Pyrimidine metabolism	1
Endocrine resistance	1
Pancreatic secretion	1
Fatty acid metabolism	1
Bile secretion	1
Fatty acid elongation	1
Amino and nucleotide sugar metabolism	1
Calcium reabsorption	1
Nicotinate and nicotinamide metabolism	1
Salivary secretion	1
Mucin type O-glycan biosynthesis	1
Biosynthesis of unsaturated fatty acids	1
Gastric acid secretion	1
Steroid hormone biosynthesis	1
<u>Pathology/Infection</u>	
HTLV-I infection	3
Non-small cell lung cancer	3
Huntington's disease	2
Alcoholism	2
Pathways in cancer	2
Cocaine addiction	2
Small cell lung cancer	2
Viral carcinogenesis	2
Tuberculosis	2
Herpes simplex infection	1
Measles	1
Hepatitis B	1
Fanconi anemia pathway	1
Inflammatory bowel disease	1

Pathway and/or disease phenotype	Hits
Endometrial cancer	1
Dilated cardiomyopathy	1
Salmonella infection	1
Influenza A	1
Morphine addiction	1
Chagas disease	1
Amphetamine addiction	1
Leishmaniasis	1
Prostate cancer	1
MicroRNAs in cancer	1
Toxoplasmosis	1
Vibrio cholerae infection	1
<u>Development/Homeostasis</u>	
Osteoclast differentiation	3
Circadian entrainment	3
Melanogenesis	2
Oocyte meiosis	1
Progesterone-mediated oocyte maturation	1
Th17 cell differentiation	1
Vascular smooth muscle contraction	1
Renin secretion	1
Platelet activation	1
Ribosome	1
Ovarian steroidogenesis	1
GnRH signaling pathway	1
Natural killer cell mediated cytotoxicity	1
Dorso-ventral axis formation	1
Th1 and Th2 cell differentiation	1
<u>Cell signaling/maintenance</u>	
cAMP signaling pathway	3
Neurotrophin signaling pathway	3
Longevity regulating pathway	3
Neuroactive ligand-receptor interaction	3
Estrogen signaling pathway	2
Rap1 signaling pathway	2
Cholinergic synapse	2
MAPK signaling pathway	2
Ras signaling pathway	2
Adrenergic signaling	2
cGMP-PKG signaling pathway	2
Chemokine signaling pathway	2
AMPK signaling pathway	2

Pathway and/or disease phenotype	Hits
PI3K-Akt signaling pathway	2
EGFR tyrosine kinase inhibitor resistance	1
FoxO signaling pathway	1
Tight junction	1
Regulation of TRP channels	1
HIF-1 signaling pathway	1
Gap junction	1
TNF signaling pathway	1
Jak-STAT signaling pathway	1
Endocannabinoid signaling	1
Cell adhesion molecules	1
Glutamatergic synapse	1
Calcium signaling pathway	1
Prolactin signaling pathway	1
Oxytocin signaling pathway	1
Ubiquitin mediated proteolysis	1
Antigen processing and presentation	1
Endocytosis	1
Phospholipase D signaling pathway	1
Dopaminergic synapse	1
Lysosome	1
GABAergic synapse	1
Cytokine-cytokine receptor interaction	1

The KEGG Mapper database was developed to provide an assessment for association of genes with metabolic pathways and human diseases.