
Polygenic Contribution to Obesity: Genome-Wide Strategies Reveal New Targets

Antje Körner^a · Wieland Kiess^a · Michael Stumvoll^b ·
Peter Kovacs^c

^aUniversity Hospital for Children and Adolescents, ^bDepartment of Internal Medicine III,
and ^cInterdisciplinary Center for Clinical Research, University of Leipzig, Leipzig, Germany

Abstract

Obesity results from the complex interaction of environmental factors that act on a genetic background that determines the susceptibility to obesity. The identification of such obesity susceptibility genes can provide important insights into the mechanism underlying this condition. While candidate gene approaches have not been tremendously successful in identifying relevant genetic contributors to obesity, except *PPAR γ* , the advent of genome-wide strategies has recently revealed novel and unexpected genetic factors with strong associations with obesity and/or diabetes, i.e. *FTO*, *TCF7L2*, *INSIG2*, *ENPP1*, or *FASN* (reviewed herein), although some of them are not undebated. Considering the function of the encoded proteins, it will now be of interest to investigate the cellular and molecular mechanisms, how these genetic variations affect body weight, energy metabolism and/or obesity-associated morbidity.

Copyright © 2008 S. Karger AG, Basel

The 'Thrifty' Heritability of Obesity?

Obesity results from an imbalance of energy expenditure and energy intake. There is a great variety of factors affecting this fragile balance. It is obvious that major environmental factors such as the accessibility to food and the degree of physical activity, but also the psychosocial environment, the perinatal environment and many other factors affect an individual's body weight. The pandemic increase in obesity prevalence had been attributed to this increasingly urbanized and sedentary lifestyle with convenient access to food, increased calorie intake and a reduction of energy expenditure in the industrialized world.

Nevertheless, over the last two decades, it has become clear that genetic factors play an important role in the determination of body weight. First evidence for the

heritability of obesity came from early twin studies that observed a heritability for body weight of 0.78–0.81% in monozygous twins [1–3] and similar values have been obtained in subsequent studies analyzing the impact of the genetic background [4–7]. However, this high degree of heritability is rarely attributed to monogenic forms of obesity [see the chapter by Farooqi, this vol., pp. 1–11], which usually result in extreme and early-onset obesity and are usually accompanied by additional phenotypic and endocrine abnormalities. Even though the discovery of monogenic forms of obesity has allowed important insights into some of these mechanisms by revealing a highly conserved pathway regulating mammalian body weight, it is obvious that the pathology of obesity is far more complex.

Considering that the genetic pool has not dramatically changed over the last decades, the secular trend of increased obesity prevalence is now regarded as the interaction of the modern lifestyle factors that act on a genetic background that determines an individual's susceptibility to weight gain and obesity. According to this 'thrifty gene' hypothesis [8] individuals with a genetic disposition to accumulate ample energy stores in times of good food availability were evolutionary more likely to survive times of nutrient scarcity and to pass these genotypes to successive generations. For example, if feast and famine cycles characterized early human life, the 'thrifty genotype' was more likely to survive periods of food scarcity. This ancient genetic selection to deposit fat efficiently is maladaptive in our modern obesogenic environment with excess calorie intake and sedentary life style, and hence the same genes now contribute to obesity.

It is now well acknowledged that a multitude of genetic polymorphisms and candidate regions scattered all over the genome regulate an individual's susceptibility to weight gain. Evolutionary concepts together with extensive population genetics to characterize geographical and haplotypic structures of newly emerging biological as well positional candidate genes will be inevitable to reveal whether these new genes could indeed represent the 'thrifty' genes. These genetic studies provide valuable insights as well as promoting new concepts into the mechanisms from the identification of previously unsuspected genetic factors.

Overview of Tools for Identifying Genes Relevant to Human Obesity

Genetic Dissection of Complex Diseases

Complex diseases such as diabetes or obesity have genetic components, which due to their polygenic nature can not easily be identified. Two basic approaches have been used to identify susceptibility genes for complex diseases: candidate gene approach and genomic approach (fig. 1). However, only limited success has been seen so far.

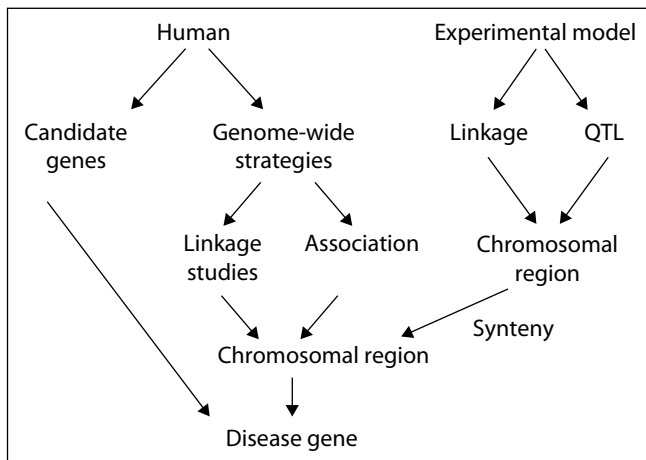


Fig. 1. Identifying genetic factors that underlie complex diseases. QTL = Quantitative trait locus.

Candidate Gene Approach

Selection of candidate genes for obesity is usually based on their known physiological role in pathways related to energy expenditure, food intake but also glucose and/or lipid metabolism and hence requires some a priori knowledge of the pathophysiology of a disease. In addition, candidate genes are selected on the basis of previous evidence on association with obesity and/or diabetes in other populations or experimental animal models. These genes are then analyzed for sequence variation that is associated or linked with the disease. Even though advances in genotyping technology allied with our knowledge of the human genome's structure will lead to novel common gene variants involved in susceptibility to human obesity, the candidate gene approach may still be very powerful when it comes to identifying rare variants predisposing to obesity. This is given by the fact that whereas coverage of common variation in genes by commercially available single nucleotide polymorphism (SNP) panels (provided by Affymetrix and Illumina) is generally comparable to the rest of the genome (see association studies below), other focused classes of functional variants are captured poorly by SNP sets aimed at common variation [9]. A large number of genes have been associated with the development of obesity; they have been reviewed recently elsewhere [10, 11].

Genome-Wide Strategies

Alternatively, susceptibility genes can be identified by genome-wide linkage or association scans, which are followed by positional cloning (fig. 2). Positional cloning requires no knowledge and/or judgment of the 'biologically plausible' genetic candidates. Instead, a disease gene is discovered because it resides on a chromosomal region that segregates

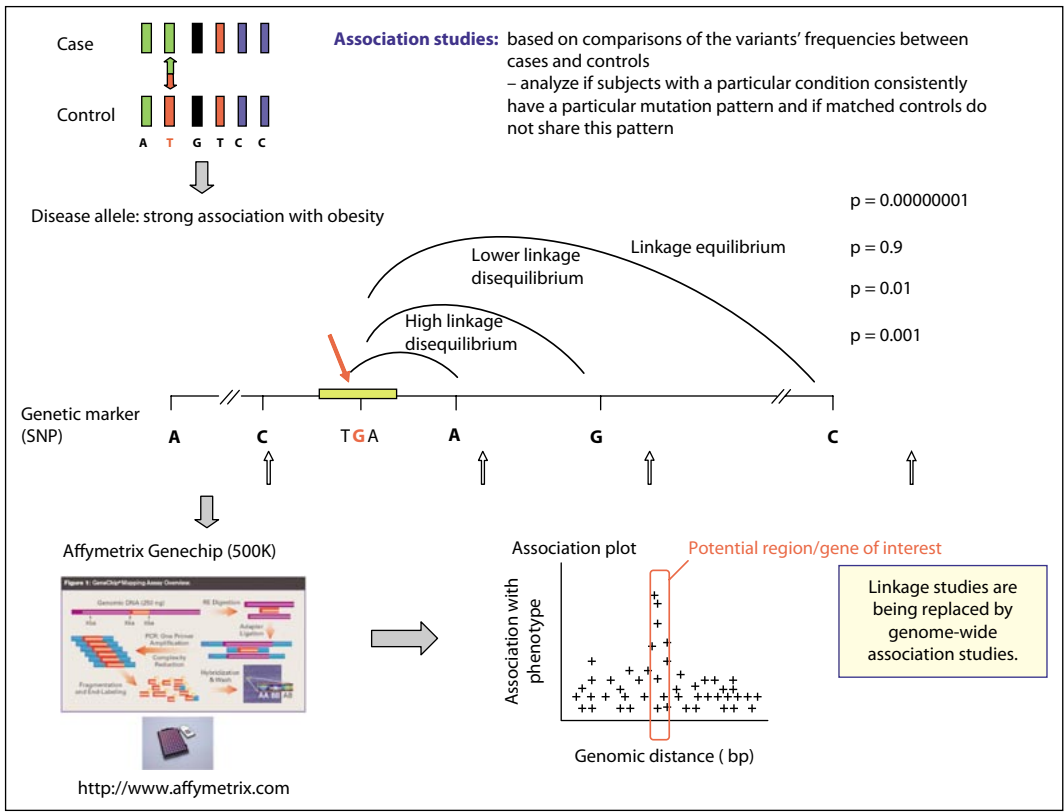


Fig. 2. Genome-wide association studies.

with a phenotype [12]. In the last two decades, genetic studies have focused on the technique of genetic linkage. This study design proved to be efficient for identifying rare, high-risk alleles, i.e. alleles that have a large population attributable fraction (PAF) in rare single-gene mendelian diseases (cystic fibrosis) [13, 14] but do not appear to have large PAF in common diseases. Using linkage studies, researchers attempt to find regions of the genome with a higher than expected number of shared alleles among affected individuals within a family. Since the design is based on closely related individuals within a family, and these individuals share larger regions of the genome, genotyping of a relatively small number of polymorphic markers is sufficient to detect region of linkage. Using positional cloning one may attempt to identify the gene(s), which reside within the linkage region and segregate with the phenotype. Usually, genetic markers (e.g. SNPs) are being selected to provide a high-density map (e.g. every 2 kb) within the region of linkage. The SNPs are then being genotyped in study subjects and analyzed for association with the disease. SNPs with strongest associations may be the causal disease risk variants (direct association) or are in a close proximity and thus in high linkage disequilibrium (non-random correlation between alleles at a pair of SNPs) and possibly indirectly associated with the disease variants [12].

In contrast to linkage approaches, association analyses are expected to be more powerful in identifying alleles that confer modest risk for developing a complex disease. Common modest risk variants account for a larger PAF than do rare high-risk alleles and this is often referred to as disease-common variant hypothesis [15]. The advantage of association analyses is based on the fact that for modest-risk alleles the patterns of sharing between related affected individuals are less striking than patterns of sharing between unrelated affected individuals [16]. This may partially explain the limited success of linkage studies in identifying genes for common diseases such as obesity. Another advantage of association analyses is that it is much easier to recruit large cohorts of unrelated individuals than collecting large pedigrees. However, since the shared region among unrelated individuals is much smaller than among the family members, association analyses require higher marker densities than linkage analyses [16]. This seems to be given by recent advances in the field of high throughput genotyping of SNPs, enabling a high-density coverage of the genome. Besides the marker coverage, another point that requires consideration in association analyses based on comparisons of the variants' frequencies between cases and controls is the power. Most studies are underpowered, especially when considering that in order to detect associations with an odds ratio of 1.2 one would need 1,000 cases and 1,000 controls [17]. Therefore, large collaborating networks enabling replications of initial findings are crucial for successful association analyses. This was impressively represented by a very recent genome-wide analysis involving research centers from Europe and including more than 35,000 study subjects, which led to discovering *FTO* as the strongest predictor of human polygenic obesity seen so far [18, 19].

Isolated Populations

Along with the polygenic nature and pathophysiologic complexity of diseases such as obesity or diabetes, another major problem is the genetic heterogeneity of modern populations. This means that healthy and afflicted subjects are likely to have very different sequences throughout the genome, not only in the area with the disease gene(s). To reduce genetic heterogeneity, one can use crossing studies with experimental models, which are inbred and so genetically uniform. Furthermore, their environment can be standardized to overcome gene-environment interaction. Linkage or quantitative trait analyses using experimental crosses may lead to chromosomal regions associated with a phenotype of interest, which may then point to syntenic regions in humans and so suggest novel human candidate genes. However, very often it proves to be extremely difficult to find an experimental model completely resembling pathophysiological aspects/patterns of human diseases. Therefore, one may attempt to reduce genetic heterogeneity by studying populations with limited genetic variability [20]. Isolated populations have already contributed to identification of mendelian variants of complex diseases, such as Hirschsprung disease in

the Amish or nonsyndromic hearing loss in Beduins [21–23]. There are several reasons to believe that studying genetics in isolated populations will result in genes involved in susceptibility for complex diseases: (a) genetic homogeneity makes genetic differences between healthy and afflicted individuals more pronounced; (b) the number of disease-causing mutations is much smaller because it can be traced back to very few ancestral carriers ('founders'); (c) more uniform environment (reducing the effects of gene-environment interactions); (d) higher prevalence for some diseases; (e) good genealogical records.

Indeed, current genetic projects on isolated populations are providing a good reason to share this optimism. The project receiving the most attention is the deCODE project, founded in 1996, which investigates the genetically isolated population of Iceland. The country of ca. 275,000 Icelanders has an extensive Icelandic genealogical database that can be traced back over 1,000 years. This unique resource together with the extensive high throughput genotyping led to the discovery of several genes controlling complex diseases such as prostate cancer [24] or stroke [25]. The Icelandic cohort was also one of the first populations in which type 2 diabetes (T2D) susceptibility genes have been reported from genome-wide association studies [26]. Another promising population in the field of metabolic disorders seems to be the population of the Island of Kosrae in Micronesia, in which a comprehensive epidemiological and genetic study has been undertaken [27, 28].

Studies in Children

Children represent an interesting population for identifying such primary genetic determinants involved in the susceptibility to complex polygenic diseases, since unlike in adults, phenotypes are less influenced by co-morbidities, their treatment, and environmental factors. In addition, the detailed evaluation of parameters of glucose and insulin metabolism at early stages of metabolic impairment may help to understand the sequence of events leading to overt pathology and diabetes.

Fatty Acid Synthase and Pima Indians

To identify genetic determinants of human polygenic obesity, researchers at the National Institutes of Health in Phoenix have focused on the relatively genetically and environmentally homogeneous Pima Indian population of Southern Arizona. The Pima Indians of Arizona are one of the most obese populations in the world and also have the highest reported prevalence of T2D [29]. Their diabetes is characterized by obesity, insulin resistance, insulin secretory dysfunction and increased rates of endogenous glucose production [30, 31]. To search for obesity susceptibility genes, a genome-wide linkage scan in Pima Indians was previously completed [32, 33]. The

strongest evidence for linkage with body mass index (BMI) was on chromosome 11q23–24 (LOD = 3.6) [32], while the strongest evidence for linkage with percentage of body fat was on chromosome 17q25 (LOD = 1.9) in a multipoint sibling-based variance component analysis [33]. The region on chromosome 17q35 seemed to be particularly interesting since the human fatty acid synthase (*FAS*) gene was positioned at 135 cM, within 7 cM of the peak of linkage to percentage of body fat [33]. The *FAS* enzyme is necessary for the de novo synthesis of long-chain fatty acids from acetyl-CoA, malonyl-CoA and NADPH [34]. Recent physiologic studies have shown that inhibition of the *FAS* gene induces a rapid decline in fat stores in mice, suggesting a role for *FAS* in energy homeostasis [35, 36]. Based on the chromosomal location and known physiology of *FAS*, the *FAS* was investigated as a candidate gene for determining body weight and percentage of body fat in Pima Indians. In these studies, a novel Val1483Ile polymorphism was identified, which was associated with percentage of body fat and 24-hour substrate oxidation rates measured in a respiratory chamber [37]. These findings indicate that the Val1483Ile substitution in *FAS* is protective against obesity in Pima Indians, an effect possibly explained by the role of this gene in the regulation of substrate oxidation. The effects of Val1483 on obesity have been recently also investigated in Caucasian children and adolescents. In a cohort of 738 Caucasian children and adolescents and 205 obese children from Leipzig, Germany, a significant interaction effect between gender and genotype was observed [38]. The findings in Caucasian children suggest a gender specific protective effect of the Val1483Ile polymorphism in *FAS* for obesity and lipid phenotypes in Caucasian boys. The story of *FAS* is an example of how combining different approaches may ultimately lead to novel genetic targets possibly involved in the pathophysiology of human obesity. *FAS* is not only an excellent candidate gene based on its known biological function, but also a positional candidate mapped within a region of linkage to percentage of body fat. In addition, studies in Pima Indians prove that genetically isolated populations may definitely help in the search for new obesity risk candidate genes.

Candidate Genes

Peroxisome Proliferator-Activated Receptor Gene

The peroxisome proliferator-activated receptor (PPAR)- γ is a transcription factor with a key role in adipocyte differentiation, susceptibility to obesity and insulin sensitivity. The common Pro12Ala polymorphism in PPAR γ is caused by a missense mutation in exon B of the adipocyte-specific $\gamma 2$ isoform. It was identified in 1997 and is thought to confer reduced transcriptional activity. The Ala allele is associated with a reduced risk for T2D. The prevalence of the Ala allele varies from about 4% in Asian populations [39] to about 28% in Caucasians [40]. Several more genetic variants in PPAR γ are known but are much less frequent. For example a very rare gain of function mutation (Pro115Gln) associated with obesity but not insulin resistance and a