

---

# Genetics of Obesity and Type 2 Diabetes

## Tracking Pathogenic Traits During the Predisease Period

Pierre Bougnères

The modern generalization of sedentary life and caloric abundance has created new physiological conditions capable of changing the level of expression of a number of genes involved in fuel metabolism and body weight regulation. It is likely that the genetic variants or alleles of these genes have in the past participated in the adaptation of human physiology to its evolutionary constraints. The nature and prevalence of polymorphisms responsible for the quantitative variation of complex metabolic traits may have been different among human populations, depending on their environment and ancestral genetic background. These polymorphisms could likely explain differences in disease susceptibility and prevalence among groups of humans. From complex traits to potentially complex alleles, understanding the molecular genetic basis underlying quantitative variation will continue to be a growing concern among geneticists dealing with obesity and type 2 diabetes, the main fuel disorders of the modern era. Genomics and genetic epidemiology now allow high-level linkage and association studies to be designed. But the pooling of large trans-geographic cohorts may in fact increase the genetic heterogeneity of studied traits and dilute genotype-phenotype associations. In this article, we underscore the importance of selecting the traits to be subjected to quantitative genetic analysis. Although this is not possible for most other multifactorial diseases, obesity and type 2 diabetes can be subjected to a pregenetic dissection of complexity into simpler quantitative traits (QTs). This dissection is based on the pathogenic mechanisms, and the time course of the traits, and the individuals' age, within the predisease period rather than on descriptive parameters after disease diagnosis. We defend that this approach of phenotypes may ease future associations to be established between QTs of intermediate complexity and genetic polymorphisms. *Diabetes* 51 (Suppl. 3): S295-S303, 2002

---

From Service d'Endocrinologie, Unité 561 INSERM, Hôpital Saint Vincent de Paul, Paris, France.

Address correspondence and reprint requests to Pierre Bougnères, Service d'Endocrinologie, Unité 561 INSERM, Hôpital Saint Vincent de Paul, Paris, France. E-mail: bougneres@cochin.inserm.fr.

Received for publication 29 March 2002 and accepted in revised form 7 May 2002.

FFA, free fatty acid; IGT, impaired glucose tolerance; ODDM, obesity-dependent diabetes mellitus; QTL, quantitative trait locus; SNP, single-nucleotide polymorphism.

The symposium and the publication of this article have been made possible by an unrestricted educational grant from Servier, Paris.

**B**ecause they depend on multiple genetic factors as well as environment, obesity and type 2 diabetes are called "multifactorial" or "complex" diseases. In the past two decades, the explosion of knowledge and tools in both molecular and computational technology has enabled the identification of genes for a number of inherited human disorders. These successes, however, have been almost exclusively restricted to simple Mendelian diseases, which are rare and of limited significance in terms of public health. The genetic approach of frequent multifactorial disorders has largely remained unsuccessful. This is particularly true for common obesity and common type 2 diabetes, which are rapidly spreading under the new sedentary conditions and feeding habits of our societies (1,2).

On the one hand, we now have almost the entire human DNA sequence available, at least in silico, and millions of genomic variations between humans are identified and available in databases. On the other hand, definition of the phenotype is a key issue in designing any genetic study for which the goal is to detect genes. To increase our chance of finding relevant susceptibility genes, we may have to look more closely at human phenotypes and start dissecting them according to their pathophysiology. All multifactorial diseases can be considered as "cases," a binary definition versus "controls." Some of these diseases can also be decomposed into descriptive quantitative traits (QTs), often called "partial phenotypes." Only a few diseases, including obesity and type 2 diabetes, can be approached through the quantification of "pathogenic" QTs. This article focuses on juvenile obesity and the subtype of type 2 diabetes that is associated with the long-term evolution of obesity (obesity-dependent diabetes mellitus [ODDM]).

### GENETIC VARIATION IN HUMANS

Human evolution has shaped the genome of modern humans. The genomic differences between individuals are mostly made up of numerous changes in single nucleotides, called mutations or single-nucleotide polymorphisms (SNPs) (3,4). For example, at a given position of the genome, a human has a T allele, whereas others have an A allele. Groups of humans have an A allele if they have the same ancestor; other groups with a different ancestor have a T allele. Proportions of A and T alleles at this position vary from one human group to another, depending on the ancestral genetic history of each group. Mutations are the source of variation, but the process of

mutation per se cannot account for the rapid evolution of the human population. The rate of change in allele frequency resulting from the mutation process alone is very low, because the spontaneous mutation rate is low. The creation of genetic variation by recombination is a much faster process because each meiotic recombination has an opportunity to change the relationships between the neighboring alleles aligned on a particular DNA segment, creating a variety of new haplotypes (5,6). Because of the recombination, linkage disequilibrium between neighboring alleles decays over time. Another important genomic evolutionary force is migration into a population from other populations with different gene frequencies. The survival and reproductive success of humans bearing different mutations have been key to the destiny of these mutations. Random genetic drift and the selection or adaptation to environment have or have not allowed those mutations to be transmitted to the successive generations of hominids. The history of human genes has determined the nature and frequency of genetic variations in a given human population (7).

As a result, not one human is genetically similar to another, although their genome does contain the same set of genes at the same locations. Each human is characterized by his or her own genotype (the unique combination of genetic variations within each gene and genomic region). Within, nearby or far from genes, the human genome contains more than 10 million nucleotide positions that are polymorphic. SNPs occur with a frequency of 1/200 to 1/1,000 bp, with various densities from a genomic region to another. They have been transmitted throughout evolution as separate single entities or as a cluster of alleles ("islands of disequilibrium") (8). There is a whole variety of SNPs (9). Some are located within protein-coding regions and can be synonymous or nonsynonymous, depending on whether they change the encoded amino acid sequence. Out of coding regions, other SNPs do not usually affect gene expression unless they occur in a promoter or regulatory region or they alternatively impair mRNA splicing efficiently.

In numerous genomic regions, a fraction of SNPs shows high frequency of a variant allele (the variant is the allele that has a frequency inferior to 0.50, although it does not mean that it occurred more recently in evolution than the wild-type allele). Allele frequency at these locations varies more or less between human populations. A variant is common in modern humans if it appeared long ago in the human genome, or in prehomonids, and has then been passed by descent to numerous individuals who reproduced at sufficient rates. The vast majority of mutations that occur are either neutral with respect to fitness (defined as the individual's ability to survive and reproduce) or are disadvantageous. If they are disadvantageous, they will tend to be removed from the population because their bearers will be less likely to survive and/or reproduce (negative selection). Occasionally, a new mutation confers a selective advantage and increases the fitness of individuals bearing it, so that it will eventually reach fixation (positive selection). Sadly, even favorable mutations were often lost during evolution (the Darwinian definition of "favorable" is highly dependent on the environment where individuals lived). Selection is thus nothing more than the

differential and nonrandom reproduction of genotypes resulting from the superior or inferior fitness of their associated phenotypes. Gene evolution does not, however, invariably require selection because changes in allele frequency can also occur by chance, owing to random sampling of gametes (genetic drift). Genetic drift can cause rapid changes only in small populations, according to the neutral theory of molecular evolution (10).

#### COMMON DISEASES: MENDELIAN OR QUANTITATIVE GENETICS?

In 1918, Fisher (11) resolved the controversies between the Mendelians and biometricians by pointing out that the variation of continuous traits could be explained by the combined action of a set of individual genes (12).

SNPs are called mutations when they mutilate the protein encoded by a gene enough to significantly impair its function. If this protein is of pathogenic importance, the mutation leads to a monogenic disease, relatively independent from the environment. People bearing these mutations often have health problems, limited survival, or low reproductive ability, explaining the persisting low frequency of such mutations, unless there was some advantage to being a heterozygote.

To be considered common, a disease would involve either a large collection of different monogenic diseases, each of them characterized by a unique mutation (one mutated gene, one altered protein, one disease). A common disease can more likely be a combination of common SNP variants (3,13,14) responsible for quantitative variations of a common phenotypic trait. In natural populations, variations in most characters takes the form of a continuous phenotypic range rather than discrete phenotypic categories. Variations are quantitative, not qualitative. Most phenotypic characteristics, such as weight, height, shape, metabolic activity, reproductive rate, and color, vary continuously. This does not mean, however, that their variation is the result of some genetic mechanisms different from the Mendelian genes with which we are familiar. The continuity of phenotypes is the result of two phenomena. First, each genotype does not have a single phenotypic expression, but a "norm of reaction" (15) that covers a wide genotypic range. As a result, the phenotypic differences between genotypic classes become blurred, so that one cannot assign a particular phenotype unambiguously to a particular genotype. Second, many segregating loci may have alleles that make a difference in the phenotype under observation. Thus, many different genotypes may have the same average phenotype. At the same time, because of environmental conditions, two humans of the same genotype may not have the same phenotype. This lack of one-on-one correspondence between genotype and phenotype obscures the underlying genetic mechanism.

The study of the genetics of continuously various characters in a population is called "quantitative genetics." In experimental organisms (plants, flies, cattle), it is relatively simple to determine whether there is genetic influence on a continuous trait, but extremely laborious approaches are required to localize the genes (16). In humans, it is already difficult to be sure of a genetic influence on continuous traits because it is almost impos-

sible to separate environmental (mostly within a single family or sib-ship) from genetic effects (16). As a consequence, although the genetics of bristle number in *Drosophila* were already well advanced 40 years ago (17), and despite significant progress (18), the quantitative genetics of complex human traits have remained largely obscure.

#### GENES THAT INFLUENCE QUANTITATIVE CHARACTERS

The critical difference between Mendelian and quantitative traits is the size of phenotypic differences between genotypes compared with the variation between individuals within genotypic classes. A quantitative character is one for which the average phenotypic differences between genotypes are small compared with the variation between individuals within genotypes.

The difference between Mendelian and quantitative traits is not the number of responsible loci. Continuous variation in a character can be caused by a large number of segregating genes. It may as well be caused by one gene with two alleles if the difference between genotypic means is small compared with the environmental variance.

If many segregating loci influence a character, this character is expected to show continuous variation because each allelic difference must account for only a small difference in the trait. This “multiple-factor hypothesis” (a large number of genes, each with a small effect, produce quantitative variation) is a classic model in quantitative genetics. Any variable character that depends on the additive action of a large number of individually small independent causes will be distributed in a Gaussian or near-Gaussian distribution in the population. Such “polygenes,” or “oligogenes,” are in apparent contrast with the genes of simple Mendelian analysis because of their small-but-equal effect on phenotypes. It takes only a few genetically various loci to produce variation that is indistinguishable from the effect of many loci of small effect. There is no real dividing line between polygenic traits and other traits. No phenotypic trait above the level of the amino acid sequence in a protein is influenced by only one gene allele. Traits influenced by several genes are not equally influenced by all of them. Some gene alleles have major effects on a trait; others have minor effects. Some medical investigators and investors in biotech companies have popularized the belief that common diseases are caused by “major” genes. No one gene has yet met their expectations. It is much more likely that common diseases are caused by common SNPs (19,20). Such SNPs can be both common and “minor.” Despite the landmark article by Fisher (11) and the considerable work since, including the availability of fancy genetic epidemiology methods (18,21–23), there remains a large gap between the theory of QTs (24) and the identification of classic Mendelian genes. Table 1 shows an example of the formidable task of demonstrating gene effects in oligogenic conditions.

#### THE SEARCH FOR OBESITY AND TYPE 2 DIABETES GENES

The molecular basis of a few single-gene human disorders resulting from loss of function has been clarified for rare causes of obesity (leptin, leptin receptor, melanocortin-3 receptor, proopiomelanocortin) (25) or hyperglycemia (maturity-onset diabetes of the young [MODY]) (26). But

TABLE 1  
Simulation results for oligogenic models

	10% prevalence		1% prevalence	
	25	50	25	50
Heritability (%)	25	50	25	50
Number of loci				
4	4,112	1,815	870	466
6	9,048	6,264	2,072	834
10	63,078	20,075	7,118	2,624

From Rice et al. (24). The table shows number of affected sib-pairs required to detect linkage at a significance level of  $\alpha = 0.0001$  and 80% power for a phenotype with heritability of 25 or 50% and prevalence of 1 or 10% of the general population. These values illustrate values usually contemplated for obesity and type 2 diabetes.

the likelihood that common obesity or type 2 ODDM is due to a disparate collection of monogenic diseases appears very low with respect to the rapid spreading of these two conditions in modern human societies.

In mice models, gene knockout experiments have targeted a large amount of molecules involved in the physiology of weight regulation or glucose homeostasis (27). Modified mice allowed investigators to observe a whole set of artificial obesity or diabetes phenotypes based on single-gene defects. Gene invalidation either replicates more or less the human alteration of the phenotype, or fails to do so, for many potential reasons that go from species-specific physiology or developmental patterns to the genetic background of species as distant as humans and mice. Sometimes the different genetic backgrounds of different mice lines are sufficient to allow or prevent the phenotypic expression of the invalidated gene. Conditional knockout experiments are developing, with the creation of “hypomorph” animals characterized by the partial expression, at a given time point, of a given gene. Often, at the end of articles reporting experiments in genetically manipulated mice, the authors suggest that their observations could provide a useful basis to understanding human pathology. Common human diseases, however, still escape this optimistic view. For obesity and type 2 ODDM, the genetic manipulation of gene expression in murine models is still far from recreating the complexity of genetic factors resulting from human evolution (28).

In recent years, the genetic approach of complex traits has been more frustrating than previously thought by medical investigators. Possibly, quantitative geneticists working on plants or flies would have predicted these difficulties. The schizophrenia debacle has led to some reevaluation of the way linkage analysis should be used in complex diseases (29). General causes for making the genetic approach of these traits so difficult have been the subject of many reviews (29–32). Remedies have been searched in several directions: refinement of phenotype definition (33), optimization of study designs (34), increased number of subjects and pooling of cohorts (35), development of sophisticated genetic epidemiology statistics (36), high throughput genotyping facilities, up-to-date linkage, and association approaches with random genomic markers or polymorphic candidate gene regions. For example, the first empirical applications of variance component analysis to QTs in human families appeared in 1996. Since then, a large number of applied quantitative trait

locus (QTL) mapping studies using variance component analysis has appeared (37).

### THE CASE-CONTROL APPROACH OF OBESITY AND TYPE 2 DIABETES

Overweight, obesity, and impaired glucose tolerance (IGT) or diabetes are defined by thresholds on the distribution of continuous traits in well-defined human populations. This definitions are usually given by a consensus conference held under the auspices of a professional group of medical and public health experts.

In young subjects, overweight is defined by a BMI (ratio of body weight to the square of body height) exceeding the 85th percentile for age and sex. Obesity is defined by a BMI exceeding the 90th percentile for age and sex (38). Using this definition, >20% of the Western adult population is overweight, and 10% is obese. Excess fat has become a population phenomenon.

IGT is defined according to the American Diabetes Association guidelines as a fasting plasma glucose level of <126 mg/dl and a 2-h plasma glucose level of 140–200 mg/dl. Type 2 diabetes is defined as a fasting glucose level of  $\geq 126$  mg/dl or a 2-h plasma glucose level of >200 mg/dl (39). It now affects >5% of the U.S. population, and type 2 ODDM represents a large fraction of it (40).

These definitions, familiar to all investigators in the obesity-diabetes field, define individuals as “cases” or “controls” according to a binary classification (those who are below or above the selected threshold). This definition of “disease” is entirely based on public health data. The threshold defining the dividing line between cases and controls reflects the value of the continuous trait (BMI, plasma glucose) above which people are expected to develop increased morbidity. This view of diseases, whereas useful to the medical approach of risk factors, has no direct link with pathogenesis. In this respect, why not define, for example, the diabetic condition using free fatty acid (FFA) levels in plasma instead of glucose (41)? And if so, what would then be the relevant threshold?

Another problem when performing genetic studies with a case-control approach is the stratification of human populations. There may be significant differences in genetic backgrounds between the case group and the control groups that have nothing to do with the considered phenotype (42). Despite the humorous caveat of Lander and Schork (29), many disease marker association studies in the diabetes-obesity field are not yet guarded against the risk of false positivity due to stratification (42).

### DISSECTION OF OBESITY AND TYPE 2 DIABETES INTO DESCRIPTIVE TRAITS

It is not easy to dissect asthma, Crohn's disease, or type 1 autoimmune diabetes into QTs. It is simpler to dissect obesity or type 2 diabetes into several partial phenotypes, which we call “descriptive traits,” that could then be used as QTs for genetic analysis. This approach has been widely used in genome scan research for new loci. Traits can be anthropometric or body composition parameters. For example, instead of being defined as obese, a young individual can be characterized as having a BMI of 39 kg/m<sup>2</sup>, a fat mass of 49 kg, a subcutaneous depot of 41 kg, a waist-to-hip ratio of 1.21, etc. This approach refines the description

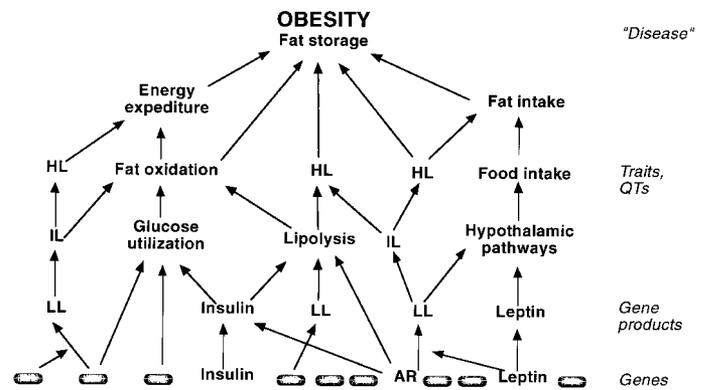


FIG. 1. Schematic diagram of the relationships of genes and their products to intermediate phenotypes concurring with weight regulation and body adiposity. Specific environmental factors are not figured because of lack of knowledge and complexity. The strength of the contribution of a lower-level factor to a higher-level factor, although depicted by arrows of similar thickness, is likely to be variable. AR, adrenergic receptor; HL, high-level phenotypes; IL, intermediate phenotypes; LL, low-level phenotypes (see text).

of the obesity status. The status of obesity can also be characterized biologically in such patients with fasting serum levels of leptin at 45 pmol/l, insulin at 33  $\mu$ U/ml, FFAs at 0.73 mmol/l, etc., with the hope that these QTs could partially reflect disease pathogenesis. The difficulty is that high leptin, insulin, or FFA levels can be implicated in both the causal mechanisms and the consequences of the obese status.

### STUDYING LOW-LEVEL PHYSIOLOGICAL TRAITS

To overcome problems plaguing genome-wide searches or associations for complex diseases, it is necessary to reduce the effects of other factors surrounding the effect of individual genes. Most physiological systems have a hierarchical component to them, leading from the gene to its product, to intermediate phenotypes of greater complexity, to the ultimate phenotypes used to diagnose disease (Fig. 1). Genetic analysis of a low-level phenotype is eased by the greater proximity to the genetic factor. It is important that the chosen intermediate phenotype relate to the ultimate phenotype of interest—in our case, excess body fat and increased blood glucose. Candidate genes can more accurately be tested against physiological traits that are close to them.

This approach also reduces the confusing effects of genetic heterogeneity at the level of disease (43–45). It is likely that obesity or diabetes is not associated with the same genetic factors among individuals or populations. The study of low-level or intermediate pathogenic traits remains meaningful in these conditions, even if these traits do not concur at a similar degree to the ultimate level of complex diseases.

Studying low-level phenotypes can also bridge the gap between the two strategies of contemporary complex disease research: the “top-down” approach, linking complex phenotypes with genotypes, and the “bottom-up” approach, which takes genes as a starting point and then works up to the phenotype (46). Genome screens are basically a top-down approach, whereas most candidate gene studies are characteristic of the bottom-up approach. Trying to relate candidate genes bearing common variants to proximal phenotypes is an attempt to make sense of

what is going on between the top and the bottom. We used this approach for association studies between insulin levels and insulin gene polymorphisms (47), leptin levels and leptin gene (48), insulin sensitivity, and *IRS1* and *IRS2* polymorphisms (49). The research work then proceeds by involving genes in low-level phenotypes (50), then attempts to relate the phenotypes studied with higher-level phenotypes. To match the term “genomics,” associated with purely genetic research, Schork calls “phenomics” the delineation of connections among genes, gene products, intermediate traits, physiological networks, and ultimately diseases (51). This concept supports the recent advances obtained in the establishment of genomic system biology maps (52), in which 239 phenotypic traits can be analyzed in each animal and compared between different genotypic groups of rats.

The study of low-level physiological traits, proximal to the function of candidate gene products, may also help to cope with both epistatic interaction between genes (53) and pleiotropy (one locus is associated with several physiological functions) (42).

#### MEASURING “THRIFTY” PATHOGENIC TRAITS DURING YOUTH

The study of young subjects is generally recommended for the genetic study of complex diseases (29). There are several reasons to consider that this could be particularly important in the obesity-diabetes field.

The first reason is that evolutionary forces may have shaped the human genome according to mechanisms (fat storage and mobilization, insulin secretion and sensitivity, leptin signaling, weight and body composition regulation, availability of glucose to the brain, etc.) that are now directly involved in the pathophysiology of juvenile obesity and associated changes in insulin–fuel homeostasis. These physiological functions and traits were of major importance during the infancy, childhood, and puberty of ancestors for metabolism, development, and growth. It is likely that prehistoric metabolic genes welcomed new mutations, provided that they favored the storage of calories. The notion of the thrifty genotype (54–56) covers all kinds of genes that could help early humans adapt to their hostile environment, when food was scarce and rather unpredictable, but nevertheless crucial for fitness and reproduction. It is likely that gene alleles favoring fat accumulation have been selected by humans and are now turning their bad effects to modern subjects because of an unexpected caloric richness and sedentary environment. Similarly, it is possible that insulin sensitivity underwent evolutionary changes toward increased channeling of glucose to the large human brain rather than to the insulin-sensitive muscle mass. Measuring these phenomena early in life rather than in adulthood may more closely reflect their evolutionary tendencies. In addition, the life span of early humans was limited, and evolution has therefore mostly worked on the physiology of young people.

The study of young individuals meets the goals of predictive genetic epidemiology because it allows the follow-up of genotyped patients through later phenotype evolution as well as clinical trials. We have also observed that the motivation and sampling of siblings and parents for genetic analysis are facilitated when medical traits,

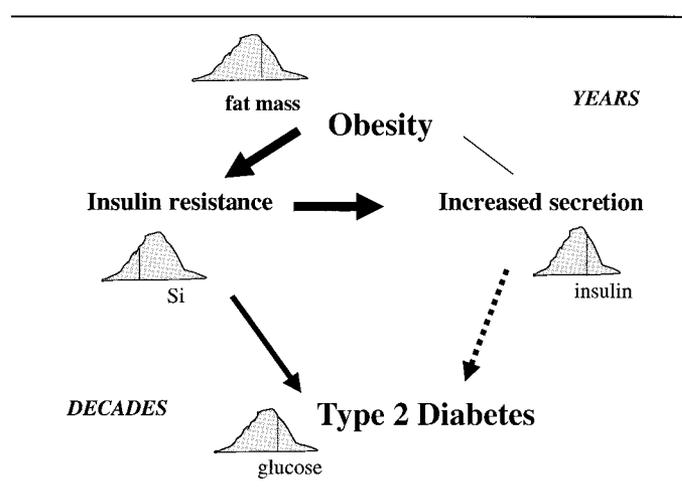


FIG. 2. Simplified scheme of the evolution of juvenile obesity toward type 2 diabetes. Dotted lines indicate the decrease of insulin secretion with evolution and aging.

such as obesity, are detected in children and adolescents. The study of nuclear families with young sib-ships favors the analysis of sib-pairs in a comparable environment, as well as transmission disequilibrium tests (57).

#### MEASURING PATHOGENIC TRAITS BEFORE DISEASE STARTS

When patients have been exposed to a disease status for a prolonged period, their physiology follows new rules, aimed at the homeostatic compensations of various dysfunctions. At the frontline of type 2 diabetes and obesity, insulin secretion and insulin sensitivity are two pillars of pathophysiology (Fig. 2). Insulin resistance and failure of insulin secretion vary largely from one patient to another, in lean as well as in obese people (58). In a large fraction of patients, they combine their effects to result in IGT, type 2 diabetes, insulin resistance syndrome, etc.

Figure 3 depicts an artificial example of the simultaneous evolution of insulin sensitivity and insulin secretion in an obese woman, who became obese at adolescence and remained so during later life. Time- and age-dependent variations of insulin sensitivity are observed depending on many factors: physical activity at different periods, reproduction cycle (from childhood to puberty, pregnancy, menopause), aging, weight changes, adjustment to decades of hypercortisolemia, hyperinsulinemia, hyperlipidemia, etc. In the meantime, insulin secretion also follows various changes. Depending on the time of observation, the investigator will get a very different view of the traits. After diabetes is diagnosed, both traits follow a course that is influenced by both disease complications and therapeutics.

An investigator quantifying insulin sensitivity or action in a cohort of women with obesity and diabetes could use statistical tools to normalize the studied trait for age, sexual parameters, physical activity, accumulated fat at the time, insulin or fatty acid, etc. However, he or she will not be able to catch *bona fide* parameters reflecting the real natural history of the trait.

The apparent phenotypic concordance between individuals, to be used in a linkage or an association approach, can also be misleading if defined in terms of disease. The

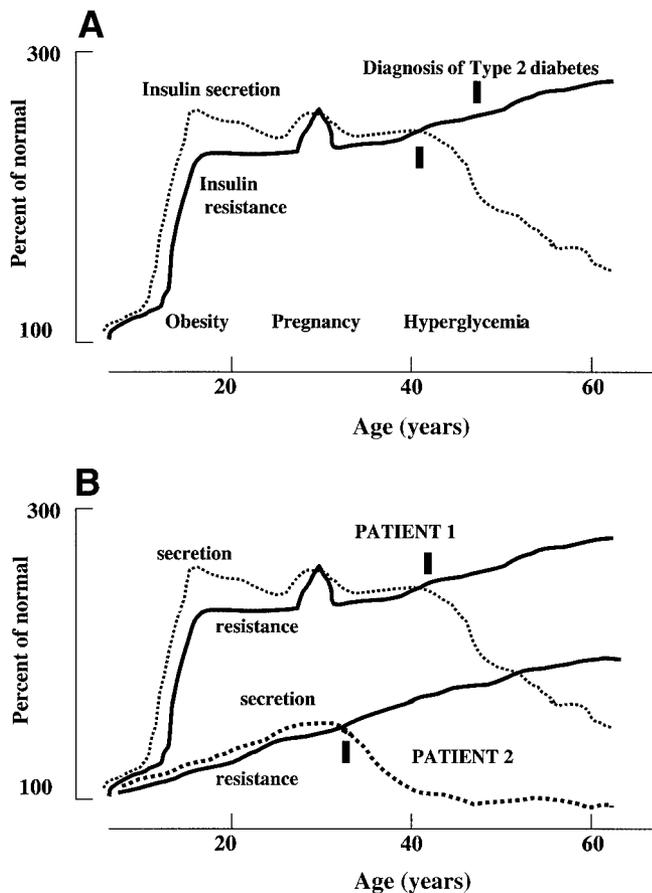


FIG. 3. Evolution of insulin resistance and secretion with age in a putative case (Patient 1) of juvenile obesity leading to type 2 ODDM in mid-adulthood (A) and in a sibling who also developed type 2 diabetes (Patient 2) (B).

two individuals in Fig. 3, although both were diagnosed during adulthood, are in fact highly discordant for both insulin secretion and sensitivity, the studied QTs.

An appealing approach is to measure the traits before they become seriously altered by pathological processes. Obviously, this requires the study of subjects while they are still in a predisease state. In a Westernized population where juvenile obesity is known to be frequent, the early accumulation of weight in the first years of life (as kilogram fat mass per year) may be a suitable QT to approach the genetic predisposition to obesity on a large population scale.

Another example is the study of traits in a population at high risk for future type 2 ODDM. Adolescents with recent and severe obesity suit this purpose. Their metabolic features have been characterized (48,59,60). According to epidemiology, their risk of future IGT and type 2 ODDM is important (61). Indexes of insulin secretion and insulin resistance can be measured at a population level in these obese adolescents with simple oral glucose tests (62), and an association can be searched for between these QTs and genetic factors.

#### QUANTITATIVE TRAITS AS DYNAMIC DEVELOPMENTAL TRAITS

Many traits, such as weight, exhibit marked developmental or time-dependent trends. The time course and develop-

mental pattern exhibited by these traits is typically influenced by a number of genetic and environmental factors. Identifying the effects of these factors on the development of such traits will be difficult unless relevant time course or longitudinal data have been collected. Even if such data have been collected at multiple time points, however, one must appropriately model or accommodate the developmental patterns exhibited by the trait in an analysis to draw compelling inferences (63). Time is important to the evolution of most physiological traits and possibly modifies their association with genetic factors (64,65).

In an age-variable study, statistics can control for the effect of covariates, such as age or puberty, thereby improving the power to detect genotype-phenotype associations (66). However, the statistical normalization to age over a wide range of ages does not allow us to reconstitute the real age-dependent values of a developmental trait, unless this trait has followed uniform evolution in various individuals. For example, one cannot know what would have been the fat mass or insulin sensitivity at age 12 years in two subjects who would undergo measurements of this parameter at 15 and 21 years of age, respectively.

There are solutions opened to clinical research that can reduce age-dependent variability. One solution is simply to measure the QT of interest at a given age, in all subjects, using a cross-sectional design. This could be of major interest in traits that are developmentally regulated, such as those pertaining to growth and physiological maturation. For the phenotypic analysis of sib-pairs, the quantification of differences between sibs' traits could allow a more precise evaluation than the definition of concordance (67) or discordance (68) used in many genome scan attempts or in association studies. Another solution is to measure the trait throughout life at multiple time points. It is only possible currently for such traits as weight and height. For example, one can quantify the early increase of weight versus time (as kilogram per year) in countries where young children are frequently weighed by pediatricians for general public health purposes.

Another remark is that obesity and type 2 diabetes are considered polygenic clinical conditions that result from the cumulative effect of a large number of genes, each with individual small effects. Because they affect phenotypes progressively over decades, the effects of these genes may prove to be quite small and therefore difficult to demonstrate. What is small, in fact, after integration during 10–40 years of trait evolution?

#### JUVENILE OBESITY ALLOWS TO AMPLIFY METABOLIC PHENOTYPES

A QTL is a gene for which variant alleles contribute to quantitative variation for some character. A QT or character is one for which the average phenotypic differences between genotypes are small compared with the variation between individuals within genotypes. Therefore, the general population may not offer a sufficiently wide range of values to allow the characterization of genotypic effects on quantitative traits, unless huge numbers of participants are contemplated. For example, in normal 12-year-old male adolescents, BMI ranges from 14.5 to 20 kg/m<sup>2</sup> (mean  $\pm$  2 SD), fat mass from 2 to 10 kg, and plasma insulin from 4 to 14  $\mu$ U/ml (P.B., personal observation). This result leaves

TABLE 2.  
Main clinical and body composition characteristics of the current Saint Vincent cohort

<i>n</i>	1,006
M/F	407/599
Age (years)	11.8 ± 0.1
BMI (kg/m <sup>2</sup> )	29.8 ± 0.2
BMI (% <i>n</i> )	167 ± 1
Fat mass (kg)	31 ± 0.5
Number of sib-pairs available*	423
Prevalence of obesity in sib-ship†	0.29

Data are means ± SE unless otherwise stated. \*Concordant or not for adiposity; †λs of 9.7, which is higher than that in other obese cohorts (71).

little space for detecting genotypic effects on these traits, unless thousands of children are enrolled in each genotypic group. Two additional problems would arise for insulin measurements in the general population. First, these measurements would be of little value in the uncontrolled dietary and physiological environment of real life where many genetic studies of metabolic traits take place. Second, the expectable genotypic variation of insulin levels would possibly fall within the limits of precision of the radioimmunoassay measurement.

If one postulates that the studied trait depends on the same genotypic effects in lean and obese children, which may not always be the case, then obesity would offer a range of phenotypic variation favorable to the detection of these effects. One could also say that obesity offers an opportunity to study the unrestrained phenotypic expression of fat storage alleles, of which the modern environment favors the penetrance. In a recently studied cohort of obese children, BMI ranged from 21 to 58 kg/m<sup>2</sup>, fat mass from 14 to 55 kg, and fasting insulin concentration from 4 to 82 μU/ml (47). Such variations allow study of each individual QT over a wide distribution of values, as well as a more precise definition of regression equations between various QTs (47), and pave the way for a more sophisticated phenotypic analysis such as phenotypic correlation matrixes (42,52). In addition, the control of experimental condition is often easier in a patient population that can be studied in clinical research centers.

#### THE SAINT VINCENT COHORT OF YOUNG OBESE PATIENTS AND THE VNTR CASE

Our studies took place in a cohort of hundreds of young obese patients, whose recruitment has been continuous since 1999 (Table 2). This cohort fulfills many of the recommendations for the genetic approach of complex diseases (44), i.e., obesity. Proband's are young, phenotypic traits are marked, prevalence of obesity in siblings is high, and geographical origin and ethnicity is known for the grandparents. In addition, epidemiological prediction allows us to consider the obese adolescents as being in a pre-type 2 situation, with nearly 20–30% expected to develop type 2 ODDM in adulthood (61,69). This cohort thus offers an opportunity to study some traits that will potentially lead to impairments of glucose homeostasis and insulin resistance (Fig. 4) long before these traits themselves become altered by the diabetic status. Only obese children whose body weight curve had remained monotonic since birth to time of study, i.e., who had never been subjected to

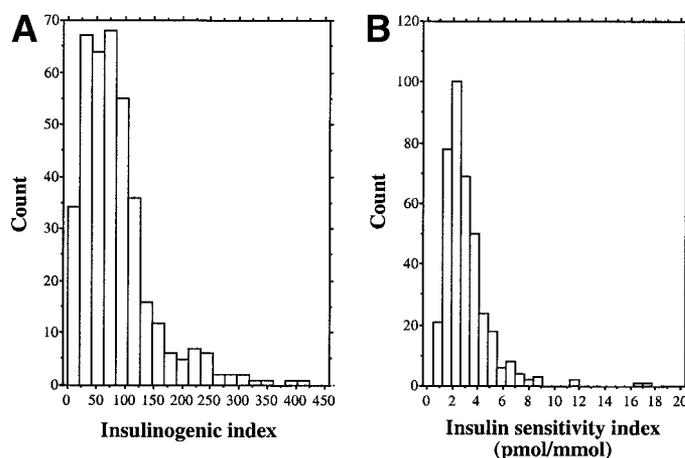


FIG. 4. Two quantitative traits reflecting insulin secretion and sensitivity in the Saint Vincent cohort of obese children. Distribution of values for an index of insulin secretion (A) (insulinogenic index; see Wareham et al. [72] for calculation and validation) and an index of insulin sensitivity (B) (composite insulin sensitivity index; see Matsuda and DeFronzo [73] for calculation and validation).

significant weight loss, are included. This period reflects as closely as possible the dynamics of early fat accumulation and of the natural, partially genetic, metabolic and hormonal adaptations to increasing adiposity.

Our focus was initially to understand how genetic factors could allow insulin secretion to cope with fat accumulation and mounting insulin resistance. Findings from early investigators supported that insulin secretion is an heritable trait (70). As candidate genetic factors, we selected polymorphisms of the insulin gene region called the insulin VNTR (variable number of tandem repeats). These SNPs were genotyped, and the VNTR alleles were studied as the main genomic markers reflecting the polymorphism of the promoter region of the insulin gene. A strong linkage disequilibrium exists in Caucasians between VNTR alleles and neighboring  $-23$  *HphI* and  $+123$  *MspI* SNPs, as well as with other nearby SNPs. Insulin VNTR alleles were classified as short (class I) or long (class III) depending on their number of tandem repeats. Through the binding of specific transcription factors, the G-rich VNTR alleles are known to modulate the transcription of the insulin gene according to studies in human pancreata as well as in  $\beta$ -cell lines.

We measured insulin in carefully codified nutritional and experimental conditions to make sure it truly reflects each individual's characteristics. We found that the polymorphisms in the insulin promoter region were strongly associated with insulin levels (47). Only obese patients with short VNTR alleles proved capable of a strong insulin compensatory secretion to obesity and insulin resistance (47).

We also observed that class I homozygous genotypes aggravate fat gain during adolescence (Fig. 5), which we interpreted as a consequence of more insulin being synthesized and secreted.

We also found that paternal VNTR class I alleles (those that are expressed in human fetuses due to parental imprinting) were associated with a predisposition to early fat accumulation in humans (57). These alleles gave a relative risk of 1.7 for childhood obesity. Because nearly 75% of Caucasian neonates have a class I VNTR paternal allele, this polymorphism can be considered a common

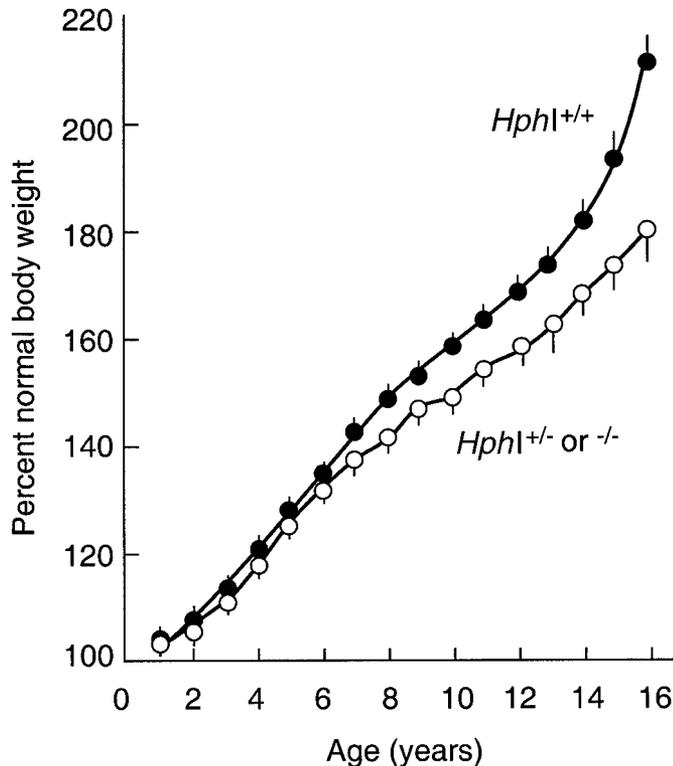


FIG. 5. Evolution of adiposity at adolescence in two genotypic groups of obese children and adolescents in the Saint Vincent cohort (47). The *HphI* T/T VNTR I/I obese patients showed a more rapid weight gain in late childhood and adolescence. Averaged longitudinal weight curves (normalized to normal weight value for age and height) are plotted versus chronological age in the two genotypic groups of obese children. Curves were constructed from the study of 332 patients whose yearly individual data could be collected from personal health records, from birth until time of study. Statistical analysis using repeated measures ANOVA as well as randomization-permutation tests indicated a highly significant difference between genotypic groups ( $P < 0.001$ ). Reprinted with permission from Le Stunff et al. [47]).

QTL for early fat accumulation. It is an example of an "oligogene" predisposing a large fraction of the Caucasian population to a relatively small phenotypic effect.

One should not be surprised by the pleiotropic effects arising at the insulin locus. First, insulin itself is a hormone with numerous pleiotropic effects on metabolism, anabolism, growth, glucose homeostasis, etc. Second, the insulin VNTR polymorphism may be in linkage disequilibrium with other functional genomic variants in the *IGF2* or tyrosine hydroxylase genes, as well as other loci in the 11p15 region involved in early growth and metabolism and subjected to various parental imprinting patterns.

In conclusion, this article suggests a systematic approach to common juvenile obesity or type 2 ODDM genetics based on measurement of related traits in a pathogenic rather than descriptive perspective. To do so, subjects should be studied long before disease starts, in a situation when traits can be measured "intact." This pre-genetic dissection of diseases could facilitate the genetic elucidation, step by step, of the mechanisms of these complex diseases. Another simple principle is to favor, whenever possible, the study of phenotypic traits proximal to candidate gene products. These choices may help to cope with the suspected genetic heterogeneity of complex diseases in different individuals and different populations. Our approach contrasts with case-control designs as well

as with statistical efforts to link or associate complex traits directly with candidate genes or loci. We are almost certain that we would not have seen any relevant result had we searched for a direct association of VNTR polymorphism with common obesity or type 2 diabetes in a large heterogeneous cohort of adult patients.

## REFERENCES

- Hill JO, Peters JC: Environmental contributions to the obesity epidemic. *Science* 280:1371-1377, 1998
- Wickelgren I: Obesity: how big a problem? *Science* 280:1364-1367, 1998
- Collins FS, Guyer MS, Chakravarti A: Variations on a theme: cataloging human DNA sequence variation. *Science* 278:1581-1581, 1997
- Lander ES: The new genomics: global views of biology. *Science* 274:536-539, 1996
- Weiss MK, Clark AG: Linkage disequilibrium and the mapping of complex human traits. *Trends Genet* 18:18-24, 2002
- Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, Messer CJ, Chew A, Han JH, Duan J, Carr JL, Lee MS, Koshy B, Kumar AM, Zhang G, Newell WR, Windemuth A, Xu C, Kalbfleisch TS, Shaner SL, Arnold K, Schulz V, Drysdale CM, Nandabalan K, Judson RS, Ruano G, Vovis GF: Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 293:489-493, 2001
- Cavalli-Sforza LL, Menozzi P, Piazza A: *The History and Geography of Human Genes*. Princeton, NJ, Princeton University Press, 1994
- Goldstein DB: Islands of linkage disequilibrium. *Nat Genet* 29:109-111, 2001
- Cooper DN: Genes, mutations and diseases. In *Human Gene Evolution*. Oxford, UK, BIOS Scientific Publishers, 1999, p. 34-37
- Kimura M: Recent developments of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. *Proc Natl Acad Sci U S A* 88:5969-5973, 1991
- Fisher RA: The correlation between relatives on the supposition of Mendelian inheritance. *Trans R Soc (Edinburgh)* 52:399-433, 1918
- Quantitative genetics and the paradigms of Mendel and Galton. In *Human Genetics, Problems and Approaches*. Vogel F, Motulsky AG, Eds. Berlin, Springer Verlag, 1982, p. 154-156
- Risch NJ: Searching for genetic determinants in the new millennium. *Nature* 405:847-856, 2000
- Gura T: Can SNPs deliver on susceptibility genes? *Science* 293:593-595, 2001
- Lewontin RC: Quantitative genetics. In *An Introduction to Genetic Analysis*. 7th ed. Griffiths AJF, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM, Eds. New York, W.H. Freeman, 2000, p. 756-757
- Barton NH, Keightley PD: Understanding quantitative genetic variation. *Nat Rev Genet* 3:11-21, 2002
- Thoday JM: Location of polygenes. *Nature* 191:368-370, 1961
- Haseman JK, Elston RC: The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 2:3-19, 1972
- Weiss KM: *Genetic Variation and Human Disease: Principles and Evolutionary Approaches*. Cambridge, UK, Cambridge University Press, 1993
- Chakravarti A: Population genetics: making sense out of sequence. *Nat Genet* 21:56-60, 1999
- Elston R: Linkage and association. *Genet Epidemiol* 15:565-576, 1998
- Rao DC: *Genetic Dissection of Complex Traits*. San Diego, CA, Academic Press, 2001
- Mackay TFC: The genetic architecture of quantitative traits. *Annu Rev Genet* 35:303-339, 2001
- Rice JP, Saccone NL, Rasmussen E: Definition of the phenotype. In *Genetic Dissection of Complex Traits*. Rao DC, Province MA, Eds. San Diego, CA, Academic Press, 2001, p. 69-76
- Barsh GS, Farooqi IS, O'Rahilly S: Genetics of body weight regulation. *Nature* 404:644-651, 2000
- Fajans SS, Bell GI, Polonsky KS: Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *N Engl J Med* 345:971-980, 2001
- Robinson SW, Dinulescu DM, Cone RD: Genetic models of obesity and energy balance in the mouse. *Annu Rev Genet* 34:687-745, 2000
- Moore KH, Nagle DL: Complex trait analysis in the mouse: the strengths, the limitations and the promise yet to come. *Annu Rev Genet* 34:652-686, 2000
- Lander ES, Schork NJ: Genetic dissection of complex traits. *Science* 265:2037-2048, 1994

30. Rao DC, Province MA (Eds.): *Genetic Dissection of Complex Traits*. San Diego, CA, Academic Press, 2001
31. Lalouel JM: From genetics to mechanism of disease liability. In *Genetic Dissection of Complex Traits*. Rao DC, Province MA, Eds. San Diego, CA, Academic Press, 2001, p. 517–534
32. Morton NE: Complex inheritance: the 21st century. In *Genetic Dissection of Complex Traits*. Rao DC, Province MA, Eds. San Diego, CA, Academic Press, 2001, p. 535–544
33. Ghosh S, Majumder P: Deciphering the genetic architecture of a multivariate phenotype. In *Genetic Dissection of Complex Traits*. Rao DC, Province MA, Eds. San Diego, CA, Academic Press, 2001, p. 323–350
34. Gu C, Rao DC: Optimum study designs. In *Genetic Dissection of Complex Traits*. Rao DC, Province MA, Eds. San Diego, CA, Academic Press, 2001, p. 439–458
35. Dahlman I, Eaves IA, Kosoy R, Morrison VA, Heward J, Gough SC, Allahabadia A, Franklyn JA, Tuomilehto J, Tuomilehto-Wolf E, Cucca F, Guja C, Ionescu-Tirgoviste C, Stevens H, Carr P, Nutland S, McKinney P, Shield JP, Wang W, Cordell HJ, Walker N, Todd JA, Concannon P: Parameters for reliable results in genetic association studies in common disease. *Nat Genet* 30:149–150, 2002
36. Weeks DE, Lathrop GM: Polygenic disease: methods for mapping complex disease traits. *Trends Genet* 11:513–518, 1995
37. Blanger J, Williams JT, Almasy L: Variance component methods for detecting complex trait loci. In *Genetic Dissection of Complex Traits*. Rao DC, Province MA, Eds. San Diego, CA, Academic Press, 2001, p. 151–182
38. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH: Establishing a standard definition for child overweight and obesity worldwide: international survey. *Br Med J* 320:1–6, 2000
39. Expert Committee on the Diagnosis and Classification of Diabetes Mellitus: Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 22 (Suppl. 1):S5–S9, 1999
40. Mokdad AH, Ford ES, Bowman BA, Nelson DE, Engelgau MM, Vinicor F, Marks JS: Diabetes trends in the U.S.: 1990–1998. *Diabetes Care* 23:1278–1283, 2000
41. McGarry JD: What if Minkowski had been ageusic? An alternative angle on diabetes. *Science* 258:766–770, 1992
42. Schork N, Fallin D, Thiel B, Xu X, Broeckel U, Jacob HJ, Cohen D: The future of genetic case-control studies. In *Genetic Dissection of Complex Traits*. Rao DC, Province MA, Eds. San Diego, CA, Academic Press, 2001, p. 191–211
43. Province MA, Shannon WD, Rao DC: Classification methods for confronting heterogeneity. In *Genetic Dissection of Complex Traits*. Rao DC, Province MA, Eds. San Diego, CA, Academic Press, 2001, p. 273–286
44. Lander ES, Schork NJ: Genetic dissection of complex traits. *Science* 265:2037–2048, 1994
45. Schork NJ: Detection of genetic heterogeneity for complex quantitative phenotypes. *Genet Epidemiol* 9:207–223, 1992
46. Sing CF, Haviland MB, Reilly SL: Genetic architecture of common multifactorial diseases. In *Variation in the Human Genome*. Sing CF, Ed. New York, Ciba Foundation, p. 211–232
47. Le Stunff C, Fallin D, Schork N, Bougnères P: The insulin gene VNTR, fasting insulin levels and development of common juvenile obesity. *Nat Genet* 26:444–446, 2000
48. Le Stunff C, Le Bihan C, Schork N, Bougnères P: A common promoter variant of the leptin gene is associated with changes in the relationship between serum leptin and fat mass in obese girls. *Diabetes* 49:2196–200, 2000
49. Le Fur S, Le Stunff C, Bougnères P: Increased insulin resistance in obese children having both 972 *IRS-1* and 1057 *IRS-2* polymorphisms. *Diabetes* 51 (Suppl. 3):S304–S307, 2002
50. Damerval C, Maurice A, Josse JM, de Vienne D: Quantitative trait loci underlying gene product variation: a novel perspective for analyzing regulation of genome expression. *Genetics* 137:289–301, 1994
51. Schork NJ: Genetics of complex disease: approaches, problems and solutions. *Am J Respir Crit Care Med* 156:S103–S109, 1997
52. Stoll M, Cowley AW, Tonellato PJ, Greene AS, Kaldunski ML, Roman RJ, Dumas P, Schork NJ, Wang Z, Jacob HJ: A genomic-systems biology map for cardiovascular function. *Science* 294:1723–1726, 2001
53. Frankel WN, Schork NJ: Who's afraid of epistasis. *Nat Genet* 14:371–373, 1996
54. Neel JV: Diabetes mellitus: a 'thrifty' genotype rendered detrimental by 'progress'? *Am J Hum Genet* 14:353–362, 1962
55. Boyd-Eaton S, Konner M, Shostak M: Stone-agers in the fast lane: chronic degenerative diseases in evolutionary perspective. *Am J Med* 84:739–749, 1988
56. Ritenbaugh C, Goodby CS: Beyond the thrifty gene: metabolic implications of prehistoric migration into the New World. *Med Anthropol* 11:227–223, 1989
57. Le Stunff C, Fallin D, Bougnères P: Paternal transmission of class I INS VNTR alleles predisposes to childhood obesity. *Nat Genet* 29:96–99, 2001
58. Ferrannini E, Natali A, Bell P, Cavallo-Perrin P, Lalic N, Mingrove G, on behalf of the European Group for the Study of Insulin Resistance (EGIR): Insulin resistance and hypersecretion in obesity. *J Clin Invest* 100:1166–1173, 1997
59. Le Stunff C, Bougnères P: Early changes in postprandial insulin secretion, not in insulin sensitivity, characterize juvenile obesity. *Diabetes* 43:696–702, 1994
60. Le Stunff C, Bougnères P: Time course of increased lipid and decreased glucose oxidation during early phase of childhood obesity. *Diabetes* 42:1010–1016, 1993
61. Must A, Jacques PF, Dallal GE, Bajema CJ, Dietz WH: Long-term morbidity and mortality of overweight adolescents: a follow-up of the Harvard Growth Study of 1922 to 1935. *N Engl J Med* 327:1350–1355, 1992
62. Hanson RL, Pratley RE, Bogardus C, Narayan KM, Roumain JM, Imperatore G, Fagot-Campagna A, Pettitt DJ, Bennett PH, Knowler WC: Evaluation of simple indices of insulin sensitivity and insulin secretion for use in epidemiologic studies. *Am J Epidemiol* 151:190–198, 2000
63. Zeger SL, Liang KY, Albert PS: Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 44:1049–1060, 1988
64. Weder AB, Schork NJ: Adaptation, allometry and hypertension. *Hypertension* 24:145–156, 1994
65. Kovacs P, vandenBrandt J, Kloting I: Effects of quantitative trait loci for lipid phenotypes in the rat are influenced by age. *Clin Exp Pharmacol Physiol* 25:1004–1007, 1998
66. Boerwinkle E, Sing CF: The use of measured genotype information in the analysis of quantitative phenotypes in man. *Ann Hum Genet* 51:211–226, 1987
67. Risch N, Merikangas K: The future of genetic studies of complex human diseases. *Science* 273:1516–1517, 1996
68. Risch N, Zhang H: Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science* 268:1584–1589, 1995
69. Brancati FL, Wang NY, Mead LA, Liang KY, Klag MJ: Body weight patterns from 20 to 49 years of age and subsequent risk for diabetes mellitus. *Arch Intern Med* 159:957–963, 1999
70. Iselius L, Lindsten J, Morton NE, Efendic S, Cerasi E, Haegermark A, Luft R: Genetic regulation of the kinetics of glucose-induced insulin release in man: studies in families with diabetic and non-diabetic probands. *Clin Genet* 28:8–15, 1985
71. Allison DB, Faith MS, Nathan JS: Risch's lambda values for human obesity. *Int J Obes* 20:990–999, 1996
72. Wareham N, Phillips D, Byne C, Hales C: The 30 minute insulin incremental response in an oral glucose tolerance test as a measure of insulin secretion. *Diabet Med* 12:931–937, 1995
73. Matsuda M, DeFronzo RA: Insulin sensitivity indices obtained from oral glucose tolerance testing. *Diabetes Care* 22:1462–1470, 1999