Type II diabetes remains a genetic nightmare. The major problem is identifying suitable pedigrees, sib-pairs, and populations for study. Segregation analysis data suggest that type II diabetes is likely to be polygenic, although one or more major genes could also be involved. This and the high prevalence of diabetes affect the strategies for searching for genetic mutations. Linkage analysis in classical type II diabetes pedigrees is unlikely to be successful. In addition, affected sib-pair analysis is limited because both parents are often affected, leading to bilineal inheritance. Sib-pairs with both parents alive are unusual, so identity by descent analysis is rarely feasible. Strategies to reduce bilineal inheritance by identifying sib-pairs with one known nondiabetic parent or with the second sibling having mild subclinical diabetes may be worthwhile. Identification of individuals or pedigrees with an unusual phenotype that suggests a single gene disorder, such as maturity-onset diabetes of the young, will continue to be important, for this allows linkage analysis with markers near candidate genes and exclusion mapping of chromosomal regions using highly polymorphic markers. Population association studies with candidate genes can detect mutations that have a minor role in the majority proportion of diabetic subjects, but large numbers are required and great care must be taken to exclude ethnic group differences between the diabetic and normoglycemic populations. The study of small inbred communities might be helpful because they may have fewer diabetogenic genes than outbred populations, and this would increase the power of sib-pair and population association studies. Direct screening for mutations in candidate genes (with single-strand conformation polymorphism or heteroduplex screening or with direct sequencing) in patients with the appropriate pathophysiological abnormality can be a successful strategy. The identification of well-defined diabetic pedigrees, sib-pairs, and suitable matched diabetic and nondiabetic populations will be key to the discovery of the genes for diabetes.

Type II diabetes is likely to have a major genetic component in view of the different prevalence between ethnic groups, the familial clustering, and the high concordance in monozygotic twins. Molecular biological techniques have identified the genes for several diseases with a clear mode of inheritance, such as cystic fibrosis, muscular dystrophy, and Huntington’s chorea. Five years ago we discussed the difficulties in applying these techniques to type II diabetes, in which the nature of the inheritance remains unknown. The strategy thought most likely to succeed was identifying extended pedigrees of patients with a defined phenotype, such as maturity-onset diabetes of the young (MODY), in which a dominant pattern of inheritance allowed linkage analysis with candidate genes or with random highly polymorphic markers. These approaches have been successful with linkage analysis identifying mutations in the candidate gene glucokinase in French and English pedigrees and with random polymorphic markers finding linkage close to the adenosine deaminase gene (ADA) on chromosome 20 in a U.S. pedigree. Specific phenotypes have also led to the discovery of insulin gene mutations in patients with high proinsulin levels, to insulin receptor mutations in patients with marked insulin resistance, and to mutations in mitochondrial DNA associated with deafness and maternal inheritance. These defects account for only a minor proportion of type II diabetes.

Advances in molecular biological techniques continue to be made, including improved techniques for screening the genome with more closely spaced markers (14) and molecular scanning techniques, such as single-strand conformation polymorphism (SSCP) (15) and heteroduplex analysis (16), and direct sequencing of candidate genes (17). The limiting factor is now predominantly choosing the most appropriate characterized patients and populations for study, rather than molecular biological limitations. A reassessment is needed in view of the growing recognition from clinical genetic studies that type II diabetes is likely to be a polygenic disease (18) and the implications for identifying pedigrees or affected sib-pairs (19) for application of the new techniques. We discuss the nature of the clinical problem and suggest potential strategies. Straightforward study of the easily available type II pedigrees and diabetic sib-pairs is not likely to be successful.
IS TYPE II DIABETES INHERITED?
The strongest data available are from monozygotic twin studies. Barnett et al. (1) showed a near 100% concordance in type II diabetic pairs. Although their twins were obtained by referral to a tertiary center and might have included ascertainment bias, the prospective study of unselected twins by Newman et al. (2) provided supportive data. They found that 58% of monozygotic twins at initial examination were concordant, but when the 15 discordant twin pairs were restudied at the second examination 10 years later, 14 had become concordant. These twin studies strongly suggest the importance of genetic factors in the development of type II diabetes but tell us nothing about the mode of inheritance, e.g., whether monogenic or polygenic inheritance has occurred. The variation in age at onset of diabetes between twins indicates a nongenetic component in the etiology of the disease (2), but since most identical twins have similar environments, lifestyles, and degrees of obesity, it is not possible to discern important environmental contributions. Further studies of the development of diabetes in identical twins reared apart and comparison of monozygotic and dizygotic twins are required (20).

Environmental factors play a major role in the clinical expression of the disease. These include the degree of obesity, the amount of physical activity, and use of drugs such as steroids and antihypertensive agents. In addition, malnutrition in the perinatal period increases the risk of developing impaired glucose tolerance (IGT) (21), although the quantitative contribution may be modest (22).

The likelihood of polygenic inheritance. The nature of inheritance of type II diabetes has been difficult to ascertain because of the late age of onset and increased mortality of the disease (6). By the time patients develop diabetes in their 50s or 60s, one or both parents are often dead. At the same time, the children may still be too young to have developed the disease. The family history is only partially informative because the condition is often subclinical.

The identification of bimodality of glucose tolerance in populations in which diabetes is common (23-25) has been taken to suggest a major single gene influence on the development of diabetes. However, bimodality could also occur if there were enhanced transition from a high normal blood glucose to diabetic levels. A possible mechanism is known, because when hyperglycemia develops, it can secondarily impair both β-cell function and insulin sensitivity (26). This feed-forward effect could induce a discontinuum. Thus, bimodality may be an inevitable feature in the development of diabetes and may not indicate a single major gene. Although a World Health Organization (WHO) Committee in 1985 suggested that a single gene defect was likely to be involved (27), this seems unlikely given the known complexity of the pathophysiology that includes contributions from impaired β-cell function, insulin resistance, and obesity.

Recent studies in Caucasian (18,28) and Asian Indian pedigrees (29) have indicated that a monogenic disorder is unlikely. Cook et al. (18) examined 20 consecutive Caucasian nuclear pedigrees in which both parents were alive and willing to be studied and that were identified with complete ascertainment without bias toward families with a known high prevalence of diabetes. Seven probands were found to have neither parent affected with diabetes or IGT. Only seven probands had a parent with diabetes, and the other six probands had parents with IGT, in spite of being two or more decades older than the proband. The finding that a sizable subgroup of type II diabetic patients have neither parent affected means that the assumption of a single autosomal dominant gene with high penetrance is not supported, although it does not exclude a major gene effect in a minority of pedigrees. Formal segregation analysis of 59 Caucasian pedigrees excluded recessive inheritance and favored a single gene disorder on a polygenic background but could not distinguish statistically between this and a multifactorial model (28). McCarthy et al. (29) studied 64 pedigrees in South India, where there is a high prevalence of diabetes, and found that in 44% both parents were diabetic. Segregation analysis suggested at least a common major gene and modifier gene, and polygenic determinants were also possible. Studies of nuclear families have unavoidable biases, since identifying diabetic patients with both parents alive is likely to lead preferentially to patients who have unaffected parents, because of the increased mortality in parents who have diabetes. Nevertheless, these data suggest that oligogenic or polygenic inheritance is likely.

These data are in accord with earlier observations. Koberling and Tillil (30) studied the family histories of 311 subjects with type II diabetes and applied an age correction to assess the expected prevalence at age 80 by a modified Stromgren method. The ultimate prevalences calculated for siblings and children were 38 and 32%, respectively, and these have often been interpreted as being consistent with dominant inheritance. However, they also found that the prevalence of diabetes in the parents of diabetic subjects was 21%, which is considerably lower than would be expected for a dominantly inherited disease. An increased prevalence of diabetes was observed in the siblings of diabetic patients who were normal weight, rather than those who were overweight. This is also in keeping with polygenic inheritances, since diabetes in obese subjects would be expected to include an environmental obesity-associated insulin resistance as well as genetic determinants, whereas diabetes might only develop in nonobese subjects who have a high genetic load.

O’Rahilly et al. (31) found a high frequency of diabetes or glucose intolerance in the parents of patients who presented with diabetes before age 40 and suggested that an increased gene dose effect may have accounted for the early onset of their disease. This hypothesis was supported by a high proportion (69%) of their siblings being either diabetic or glucose intolerant. The inheritance of determinants from both parents may be a more general phenomenon because studies of type II diabetic pedigrees show a graded relationship between the number of parents affected and the prevalence of affection in siblings (19). It is thus possible that diabetes in patients with neither parent affected may have occurred from inheritance of a combination of determinants from each parent. This also would mean that even when the diabetic patient has only one parent affected, the possibility of inheritance of determinants from the unaffected parent cannot be excluded.

The high incidence of diabetes in the offspring of conjugal diabetic parents is also in accord with polygenic inheritance rather than a major dominant, codominant, or recessive gene effect. Cooke et al. (32) found a prevalence of known diabetes of 4%, but only 12% of the offspring were older than 50 (32). Kahn et al. (33) found a higher prevalence when they performed glucose tolerance tests. Tattersall and Fajans (34)
Autoimmune the time of intercurrent illness but most affected patients probably have a polygenic disease. chance finding, as gestational diabetes or symptomatically at insulin resistance and β-cell function, although possibly to development of diabetes, because it is a major determinant mined that genetic factors contribute to both the development of which a dominantly inherited low-density lipoprotein recep-

Glucokinase mutations. In ~50% of families with MODY, it is due to mutations in glucokinase (42). These induce only slightly raised glucose levels, with a modest increase with age, so that it is often undiagnosed and only presents as a chance finding, as gestational diabetes or symptomatically at the time of intercurrent illness (8,43). It can also present as classical type II (44) (Fig. 1).

Autoimmune type I diabetes. Although autoimmune type I diabetes usually presents in childhood or early adult life, patients exhibiting the disease in late or old age are well known. Type II diabetic patients with positive islet antibo-


diabetes that was probably only due, in part, to the glucokinase mutation. Two of her diabetic daughters did not have the glucokinase mutation and had classical type II diabetes that had become insulin-requiring; they may have inherited a classical diabetes gene from their mother. The third daughter with the glucokinase mutation had a less severe phenotype and was treated with diet. This affected sib-trio thus included heterogeneity. Without previous identification of the glucokinase mutation, linkage in the whole pedigree with a glucokinase polymorphism gave a negative LOD score.

suggesting that the cumulative risk of diabetes or IGT was 60% by age 60. Ganda and Soeldner (35) have suggested that by age 85, 33% of offspring would have diabetes and 50% would have an abnormality of glucose tolerance.

Insulin resistance, impaired β-cell function, and obesity. Type II diabetes is recognized to arise from a combination of insulin resistance and impaired β-cell function, as discussed in a recent perspective (36). In Pima Indians (37) and in Mexican-Americans (38), insulin resistance is the first identifiable feature, whereas in Caucasian populations β-cell deficiency appears to be more marked at an early stage in the development of diabetes (39,40). In all populations it is likely that genetic factors contribute to both the development of insulin resistance and β-cell function, although possibly to differing extents. In addition, obesity is a major factor in the development of diabetes, because it is a major determinant of insulin resistance. Because obesity is genetically determined (41), separate genetic determinants of insulin sensitivity, β-cell function, and appetite control are likely to contribute to type II diabetes. Each involves complex regulatory pathways, giving rise to the likelihood of many contributing genetic factors, quite apart from environmental influences.

Genetic heterogeneity. Already several different genetic determinants have been described. They provide examples of heterogeneity analogous to hypercholesterolemia, in which a dominantly inherited low-density lipoprotein receptor abnormality accounts for a small proportion of patients, but most affected patients probably have a polygenic disease. Glucokinase mutations. In ~50% of families with MODY, it is due to mutations in glucokinase (42). These induce only slightly raised glucose levels, with a modest increase with age, so that it is often undiagnosed and only presents as a chance finding, as gestational diabetes or symptomatically at the time of intercurrent illness (8,43). It can also present as classical type II (44) (Fig. 1).

Autoimmune type I diabetes. Although autoimmune type I diabetes usually presents in childhood or early adult life, patients exhibiting the disease in late or old age are well known. Type II diabetic patients with positive islet antibo-

ies and/or human leukocyte antigen (HLA) DR3/DR4 alleles can have a more severe insulin-requiring phenotype (45), although many have a normal type II diabetic phenotype (46). The development of assays for anti-glutamic acid dehy-
drogenase suggests that 5–10% of type II diabetic patients may have autoimmune disease (47).

Mitochondrial mutations. The association of type II diabetes with deafness in families with inheritance through the mother led to the discovery of a point mutation in position 3243 of the mitochondrial gene encoding tRNA Leu(UUR) (13), but preliminary studies suggest a diabetes prevalence of <1% in Caucasian subjects.

Insulin receptor mutations. The majority of insulin receptor mutations have been identified in patients with severe insulin resistance syndrome (12) with acanthosis nigricans and, in women, by androgenization and polycystic ovary syndrome. Most do not have diabetes, but mutations that effect the function of insulin receptors to a less marked extent could make a contribution to diabetes in some patients (48).

Insulin receptor substrate 1. Polymorphisms have been identified in the insulin receptor substrate 1 (IRS-1) gene (49), but a similar frequency in the normal population suggests that they may not be pathogenic mutations (50).

DEFINING THE PHENOTYPE FOR GENETIC AND MOLECULAR BIOLOGICAL ANALYSES

What Criteria Should Be Used for Identifying Diabetes? Type II diabetes is primarily a subclinical disease and usually only presents clinically with symptoms of hyperglycemia when the fasting plasma glucose (FPG) level is ~12 mmol/l, compared with a normal 95th percentile range of ~3.5–5.7 mmol/l. Affected subjects with FPG levels in the range of 6–12 mmol/l are often asymptomatic, and the disease can only be detected by screening tests. The definition of the phenotype is important but difficult in view of the continuum from normality through IGT to diabetes and the need to produce arbitrary categorical criteria for most analyses. In view of the late onset of the disease, phenotypic distinction between affected and nonaffected patients is particularly difficult in early middle age. In the general population, glucose tolerance declines with age in unaffected as well as affected subjects, due, in part, to physical inactivity and obesity. It is inappropriate to define abnormality on the basis of an arbitrary fixed percentage in the upper tail of population distribution, since the clinically diabetic population ranges from as little as a few tenths of a percentage in young adult life to 10–15% or more in old age (51). In practice, at present, a single non–age-related threshold is most appropri-

ate. A low threshold will be sensitive but not specific, whereas a high threshold will be insensitive but specific. For population association studies comparing allele frequency of unaffected and affected groups, specificity is of greater importance than sensitivity, and a relatively high threshold for classification is appropriate. Specificity is increased by a repeat test. Two FPG values >6 mmol/l, measured with precise quality-controlled laboratory assays, were the entry criterion for the U.K. Prospective Diabetes Study (52) and may be suitable for many purposes. This level may be sufficiently high to exclude an increase due to either obesity per se or IGT induced by inadequate perinatal nutrition. It is imprudent to define patients as affected/nonaffected on the basis of a single test unless the result is grossly abnormal and
there is other corroborative confirmatory data, such as diabetic symptoms or complications.

In linkage studies within pedigrees, a lower threshold may be appropriate, because using Bayesian logic, the predictive power of a glucose concentration at the upper end of the normal distribution is high since the likelihood of having an affected relative is high. It may then be appropriate to require two values >5.7 mmol/L. The development of methods of linkage analysis that evaluate quantitative traits will be more appropriate than use of a categorical affected/nonaffected dichotomy.

**Which is the most suitable test?**

**Oral glucose tolerance test.** The oral glucose tolerance test is the standard test for diabetes, with a WHO criterion of a plasma glucose concentration >11.1 mmol/L at 2 h after a 75-g oral glucose load. The WHO criteria for diabetes relate to the level of glycemia that gives a particularly high risk of developing microvascular complications, such as retinopathy and albuminuria (53,54), and the IGT criteria >7.8 mmol/L at 2 h may be more appropriate for identifying the phenotype in pedigrees. However, poor reducibility is a major problem, with a coefficient of variation on the order of 15-20% (55). The variation is due, in part, to day-to-day differences in the gastric emptying rate, leading to variable absorption of glucose and a variable stimulus. The imprecision means that two abnormal tests are particularly required.

**FPG.** FPG has a day-to-day coefficient of variation of 6% and, if care is taken with quality control of assays, it provides a suitable, precise measure of the severity of diabetes (56). The WHO fasting criterion, an FPG concentration of ≥7.8 mmol/L, is too high for defining the phenotype of patients with the glucokinase syndrome, who often have FPG levels of 6.0-7.8 mmol/L (8,42). Patients with glucose intolerance in this range have to be described as "affected" in genetic analyses. The criterion of FPG concentration >6 mmol/L has been used in defining MODY (8,42). The advantage of the FPG test is that it is cheaper and easier to use than stimulation tests and provides a simple-to-understand precise estimate with similar quantitative information to oral glucose tolerance testing. It is the most appropriate test for general purpose use, as long as a reliable, quality-controlled laboratory assay is used, preferably with measurement on two separate days to help decrease the likelihood of an anomalous result.

**Should the Pathophysiology Rather Than Glucose Tolerance Be Studied?** Because impaired insulin sensitivity and impaired β-cell function are likely to have separate genetic determinants, in theory, the search for the genes for these two major pathophysiological conditions should be done separately, by identifying patients, pedigrees, or populations who have predominantly insulin resistance or β-cell dysfunction.

Identification of particularly insulin-resistant type II diabetic patients has been used in the search for mutations of GLUT4 (57) and the insulin receptor (58). In Pima Indians, analysis of sib-pairs who have increased insulin resistance, rather than diabetes, found an association with markers on chromosome 4q (59). The current tests are not as precise or easy to interpret as one would like, and the distinction of abnormality from normalcy is less well defined than for glucose tolerance tests.

**Which is the most suitable test?**

**Euglycemic and hyperglycemic clamps.** These are the current gold standard tests, but because they are labor-intensive, they are more suited for clinical research in small cohorts than for large-scale family or epidemiological studies. Few studies of their day-to-day variability have been done.

**Minimal model: frequently sampled intravenous glucose tolerance test.** This test provides assessment of both β-cell function and insulin sensitivity. The test requires a bolus of intravenous glucose followed 20 min later by a bolus of insulin and is labor-intensive with 31 blood samples over a 3-h period (60). The pathophysiological assessment requires detailed computation to provide indexes of insulin secretion and insulin sensitivity. It has been used quite extensively and seems to give clinically meaningful answers in subjects with normal glucose tolerance or IGT, but there is little experience in type II diabetes. It is uncertain whether the reduction in plasma glucose in the 20 min after the glucose bolus and before the insulin bolus provides valid information on glucose tolerance.

**Homeostasis model assessment (HOMA).** An analysis of the interaction between FPG and insulin or C-peptide measurement with the aid of a structural mathematical model allows an assessment of the degree of impaired β-cell function and insulin resistance in a person (61). The mean of values from three samples taken at 5-min intervals reduces noise from the 13-min plasma insulin oscillations that occur in normal humans. In comparison with clamps, HOMA has been shown to be accurate, but moderately imprecise, with a coefficient of variation ranging from 20 to 25%. It is thus more suitable for population studies than for assessment of individuals.

**Continuous infusion of glucose.** This test uses the same model as HOMA, but the assessment is made after a low-dose continuous infusion of glucose (0.5 g·kg ideal body wt \(^{-1}\)·min \(^{-1}\) for 1 h) that increases the glucose, insulin, and C-peptide levels, thus providing more precise assay measurement (62,63). β-cell function and insulin sensitivity are more precise, on the order of 15-20%, than can be obtained by HOMA. The test also assesses glucose tolerance and provides a relatively simple test that can be used in large numbers.

**Oral glucose tolerance test.** The increase in insulin from baseline to 30 or 60 min, when expressed in terms of the glucose increment, provides a measure of β-cell function (64). A validated measure of insulin sensitivity has not been defined.

**Insulin tolerance test.** The glucose response over 20 min following an intravenous insulin bolus provides a simple test of insulin sensitivity (65). An infusion of insulin, glucose, and somatostatin for 2 h is an alternative (66).

**Interpretation of results.** In practice, measurements of β-cell function and insulin sensitivity are not easy to determine because they are influenced by interacting effects of obesity, the amount of exercise and food intake in the previous few days, and the degree of hyperglycemia. β-cell function has to be assessed in relation to the prevailing obesity or measured insulin resistance (63,67). In addition, classical type II diabetes the degree of β-cell function deteriorates with age (68), but adjustment is problematic, as the mathematical relationships have not been evaluated. The simplest model may be to assume a linear decline from age 20 (69), but other functions might be more appropriate. Adjustment for measured insulin resistance needs to take into account the prevailing obesity, and this can also be done.
by regression models (59,63). The optimal tests and methods of interpretation are still uncertain.

Because most type II diabetic subjects have both insulin resistance and impaired β-cell function and their measurement is still controversial, in practice, at present, it is still reasonable to use indexes of glucose tolerance, such as FPG, as a surrogate index for impaired β-cell function. On the other hand, the fasting insulin is a more appropriate simple index of insulin resistance.

**POTENTIAL PROBLEMS WITH THE STANDARD APPROACHES FOR IDENTIFYING GENE LOCI**

**Linkage Studies in Large Pedigrees.** Linkage studies to discern co-segregation of genetic markers with a specific phenotype in pedigrees are the major method of identifying the locus of genetic mutations that have high penetrance. Both polymorphic markers near candidate genes and spaced dinucleotide repeats for exclusion mapping are used. One extended family with 18 informative meioses would, on its own, have an 80% chance of detecting linkage with 10% recombination between the markers and a gene causing diabetes (70). Glucokinase mutations and an ADA-linked abnormality explain only a proportion of MODY pedigrees, and linkage analysis in other MODY pedigrees is likely to be successful.

**Disadvantages.** Large informative pedigrees for linkage analysis of classical type II diabetes are virtually nonexistent. By the time diabetes is diagnosed in middle age, the parents have usually died (often from diabetes-related atheroma), so nuclear families with both parents alive are unusual. Families with diabetes in three generations who are alive do not exist except for those with MODY. This is confounded by further specific problems.

1. Two affected siblings have a high prevalence of two affected parents. If one searches for genetically interesting pedigrees with two affected sibs, often both parents are affected even when the second affected sib has IGT rather than diabetes (19). If one chooses sib-pairs in whom both had overt diabetes, it is likely that both parents are affected in the majority of pedigrees. Linkage studies give little useful information when both parents are affected because of the likelihood of bilineal inheritance.

   Linkage studies require a diabetic and a normal parent and two or more affected offspring to become informative. This constellation is unusual because by the time the second sib has developed clinically apparent diabetes, the whole family will be older, and at least one parent is likely to have died from diabetes-related premature mortality.

2. Early age-at-onset diabetic patients have a high prevalence of two affected parents. If one searches for diabetic probands whose disease became apparent in early middle age, so that both parents might be alive and available for study, they are often both affected (31).

3. High gene frequencies are problematic for extended families and nuclear families. The high prevalence of the disease means that there is a possibility that several different genes from different forebears are being expressed. For example, cousins share one-eighth of their genes. If there were a single major dominant diabetogenic gene with a gene frequency of 15%, cousins with diabetes would be as likely by chance to express different affected alleles inherited from different grandparents as a single mutation inherited from their common grandparent.

   In addition, when the disease is common and a high gene frequency is expected, one affected parent may have more than one diabetic determinant, including a homozygous dominant gene. Then, a family may appear to have dominant inheritance, with one diabetic and one nondiabetic parent having two diabetic offspring, but it would not be helpful for linkage analysis if the two offspring had inherited different affected alleles from the diabetic parent.

4. Heterogeneity within a pedigree and between pedigrees. If diabetes is a polygenic disorder, heterogeneity is likely, even within pedigrees (Fig. 1). Complex genetic interactions could occur, as in type I diabetes (71). Because each diabetic subject may have several interacting gene mutations operating, there is no guarantee that two diabetic subjects within a pedigree share the same genotype. When there is heterogeneity between families, combining results of logarithm of odds (LOD) scores between families will be misleading because different families are likely to carry different diabetogenic genes.

In summary, the apparently direct approach of identifying multiplex type II diabetic pedigrees is not applicable. It is unlikely that linkage analysis will be informative unless a specific phenotypic subtype can be identified in a family, e.g., MODY.

**Linkage Studies in a Large Number of Small Nuclear Families.** By conducting large-scale screening of diabetic families, it is possible to identify pedigrees in which the likelihood of bilineal polygenic inheritance is minimized, i.e., by detecting diabetic probands with two living parents, one of whom has type II diabetes and the other being normoglycemic. Screening the proband's siblings for undetected type II diabetes can provide pedigrees in which linkage analysis can be done to determine whether a candidate gene could have a mutant behaving as a dominant with high penetrance. If nonlinkage can be demonstrated in many small pedigrees, the candidate gene is unlikely to be a prevalent dominant gene accounting for a major proportion of type II diabetes (72).

**Disadvantages.** Although exclusion of linkage to a major gene is often feasible and is useful initial information, identifying co-segregation is less easy. In small families, even if the candidate gene does cause the disease, the LOD score would only be weakly positive and could not be distinguished from chance alone without additional mutation screening methods (see below).

**Affected Sib-Pair Analysis.** Sib-pair analysis relies on the greater sharing than would be expected by chance of parental genetic markers, thus implicating a potential genetic determinant for the disease (73). The analysis does not require assumptions about the mode of inheritance of diabetes. Because only diabetic patients are analyzed, the question
of defining the phenotype in apparently normal subjects does not occur.

1. **Identity by descent.** Establishing sharing of specific alleles by identity by descent is a powerful analysis that requires availability of both parents for typing. 1) to define phase, i.e., determining whether the affected sib-pair share a specific maternal or paternal allele; 2) to exclude families in which both parents are affected, because the sibs may each have inherited different determinants and be heterogeneous; and 3) to exclude nonpaternity.

**Disadvantages.** The problems affecting the availability of families of linkage studies (see above) also apply to sib-pair analysis.

1. In practice, it is extremely difficult to find a large number of sib-pairs who have both parents alive (19). When identifying younger patients with both parents alive, in a high proportion both parents are affected (31) (with the possibility of bilineal inheritance of different determinants to provide a within-sib-pair heterogeneity).

2. Sib-pair analysis has the most power for detecting a recessive gene, but segregation analysis indicates that this is unlikely (28,29). Large numbers of sib-pairs are required to achieve reasonable statistical power for detecting dominant effects, particularly when there is heterogeneity. Figure 1 shows that heterogeneity can occur between a sib-pair, as well as between sib-pairs. If a specific genetic determinant contributed to diabetes in the majority of patients, it could be detected by studying sufficient numbers of sib-pairs. If different sib-pairs have completely different genetic determinants, the number of sib-pairs required to detect determinants becomes very large.

2. **Identity by state.** When parents are not available, sib-pair analysis assesses the sharing of alleles in relation to the known frequency of the alleles in the general population (rather than in relation to the sharing of specific parental alleles in identity by descent). While this method is less powerful than identity by descent (74), it has the apparent advantages that affected sib-pairs can be obtained in large numbers if parental availability is not a requirement.

**Disadvantages.** Many of the caveats for linkage analysis and identity by descent analysis apply to identity by state analysis. Thus, when a sib-pair with clinically apparent diabetes is identified, a high prevalence of bilineal inheritance is likely. Many sib-pairs will have diabetes that has arisen from nonshared as well as shared genetic determinants. In addition to heterogeneity within sib-pairs, there is likely to be heterogeneity between sib-pairs. To overcome the effects of this noise, it is likely that several hundred or even a thousand sib-pairs will be required. If diabetes is so polygenic that each causative mutation occurs in only, say, 25% of diabetic patients, even then it may not necessarily be shared between sib-pairs.

As for population association studies (see below), any differences in genetic admixture between the diabetic sib-pairs and the reference normal population can give false-positive results. Identifying true-positive from low-order significance false-positive results, which are obtained when studying large numbers, is bound to become a problem.

**Population Association Studies.** Population association studies help to determine the likely role of a candidate gene in a polygenic disease, and like sib-pair analysis, no assumption about the mode of inheritance is required (75). They have, for instance, been helpful in determining a role for the insulin gene in type I diabetes (76).

The method could detect a single common mutation contributing to a majority of the diabetic patients being studied. Given the likely heterogeneity of the disease and the fact that only a small proportion of type II diabetic subjects may share the same gene mutation, the sensitivity of population association studies is likely to be low. Even larger numbers than for sib-pairs are required.

**Disadvantages.** In type II diabetes, the published literature has been predominantly unhelpful. This is because of presumably false-positive reports that have not been repeated by other workers, probably reflecting

1. the difficulty of excluding a genetic admixture between diabetic and normal populations. In populations that are known to have mixed heritage, e.g., American blacks, there is an increased risk of modest differences in ethnic background to produce a false-positive result. Thus, in Nauruans, the apparent relationship between HLA and diabetes was found to be due to the degree of admixture of Caucasian and Nauruan ancestry (77).

2. a bias to overreporting of positive results, some of which may be type I errors.

**STRATEGIES IDENTIFYING GENETIC DETERMINANTS OF TYPE II DIABETES**

Since no approach is clearly superior for classical type II diabetes, more than one type of study is needed for both candidate genes and exclusion mapping. In view of the major problems of identifying suitable type II diabetic pedigrees, linkage analysis in pedigrees obtained in the general population is not recommended because of the likelihood of heterogeneity arising from bilineal or multineurine inheritance with a family. False-negative results can occur in large pedigrees, even when there is a major gene contributing to a high proportion of the diabetic members, as in Fig. 1.

**Study of candidate genes.** The candidate gene approach has the advantage that the availability of a polymorphism close to the gene makes it possible to report clear positive or negative results because recombinants between the marker and the gene are unlikely. However, only those genes that have been characterized can be studied, and most of the methods have limited power in assessing minor rather than major gene effects. Six approaches are feasible:

1. **Linkage analysis in MODY pedigrees, or in other pedigrees with a defined phenotype.** Analysis of MODY pedigrees continues to be the major method for studying candidate genes thought to impair β-cell function. The pedigrees should probably have at least three subjects (sibs, parents, or cousins) with an early age at onset to define MODY. Identification of other specific pedigrees in which a defined phenotype appears to be inherited in a dominant fashion will be very helpful.

2. **Population association studies between diabetic and nondiabetic populations.** In the general population, population association studies should use large numbers (at least 200/group), and the type II diabetic and normal control
Subjects should have similar age distributions obtained from the same geographical population to try to exclude ethnic admixture. Asking each diabetic patient to nominate a spouse or friend provides a simple initial means of obtaining a control group that has many of the required characteristics. Large numbers, e.g., 500 vs. 500 or even 1,000 vs. 1,000, would probably be required to give reasonable power of detecting a minor gene in a substantial proportion of a polygenic diabetic disease. Little reliability should be given to the first report of a positive association study, because until it is confirmed in separate populations, it is regrettably likely to be a false-positive.

Populations with low prevalences of diabetes, e.g., Caucasian, are likely to be more informative than populations with a high diabetes prevalence, e.g., Mexican-Americans, in which it might be expected that the general population, as well as the diabetic subjects, will have a high background prevalence of at-risk genes. In addition, populations with little intermarriage between ethnic groups will be advantageous for ethnic matching of diabetic and nondiabetic subjects. Thus, southern European, Mexican-American, and black American populations may require more efforts to ensure suitable control populations than certain north European or Asian communities.

An advance would be to examine subgroups of type II diabetes that have the specific pathophysiology relevant to the candidate gene being studied. For instance, only diabetic subjects with marked β-cell impairment or reduced insulin sensitivity could be identified for comparison with the general population (or even a specific subgroup of the general population without that phenotype, e.g., patients with insulin-resistant diabetes versus the insulin-sensitive general population).

In isolated inbred communities, the likelihood of less heterogeneity than in an outbred community is likely to improve the sensitivity of the population association approach. On the other hand, as shown with Nauruans, even minor out-breeding can give false-positive results.

3. Sib-pair analysis in type II diabetes. As for population association studies, identity by state sib-pair analysis is only recommended in the general population if large numbers are used, probably on the order of 400 or more sib-pairs, to overcome the heterogeneity within and between sib-pairs. In addition, careful ethnic group, geographical, and age matching of the diabetic and nondiabetic population is needed, and populations with a minimal genetic admixture and low disease prevalence may be advantageous.

In the general population, although identification of known affected sib-pairs is the easiest option, it would be advisable to be more selective to try to reduce the likelihood of heterogeneity with a sib-pair from bilineal inheritance.

In theory, the best option might be to search for sib-pairs with a living parent who can be tested and found to be normoglycemic. The assumption is made that the majority of the genetic determinants come from the other parent, who will often be dead. Searching for sib-pairs with both parents alive is excluded by the infrequency with which this occurs.

Since bilineal inheritance in known sib-pairs is common, it is advisable to study sib-pairs who are obtained by screening the sibs of a known diabetic proband for occult, previously undiagnosed subclinical disease, so that the sib-pair consists of the diabetic proband and a sibling with IGT or mild diabetes. At the same time, the identification of further nonaffected siblings can improve the power of the analysis, both by providing control nondiabetic siblings and by providing more information on the likely parental genotypes. As for population association studies, parallel recruitment of the nondiabetic spouse or friends of subjects helps to identify a suitable geographical-, age-, and ethnic group-matched reference population.

In isolated inbred communities, as for population association studies, these communities may have the advantage that there are fewer diabetogenic genes involved than in the general population. Thus, in Pima Indians, linkage of markers on chromosome 4q with insulin resistance were identified by a sib-pair analysis.

4. Linkage analysis in many small nuclear families. A study of many nuclear families in which only one parent is affected can help exclude a major gene effect for a candidate gene. Individual pedigrees in which positive LOD scores are found, even though not statistically significant, could be used for direct screening for mutations (see below).

5. Direct detection of mutations in diabetic subjects. When the sequence of a candidate gene is known, a direct search for mutations in the coding and promoter regions is a powerful test. Mutations giving stop codons in an intron or mutations in the regulatory regions up to 1,000 base pairs away from the gene would not be detected.

Direct sequencing. For small genes, direct sequencing of the whole gene is the most efficient approach, as done by Nishi et al. (17) for the IAPP gene. For larger genes, a prior indication of a possible mutation is probably required, such as a positive LOD score in a nuclear pedigree or the presence of a specific appropriate phenotype.

Molecular scanning for mutations. For large genes, direct sequencing may be impractical, in which case a search for mutations by a scanning technique, such as SSCP (18) or heteroduplex scanning (16), can be used. SSCP allows 300–400 base pairs to be screened directly, but cannot completely exclude a mutation since the sensitivity is probably ~70–80% and some mutations will remain undetected.

SSCP and direct sequencing detect silent polymorphisms (base-pair changes that induce no change in the expressed amino acid) and conservative mutations (that induce an amino acid change but do not affect function). Further study is required (see below) before it can be assumed a causative mutation has been found.

Exclusion mapping with many polymorphic markers. Exclusion mapping has had notable successes in many conditions, such as polycystic kidney disease and cystic fibrosis. It allows the exclusion of large regions of the genome. Three approaches are feasible.

1. Linkage analysis in MODY pedigrees. The availability of large-scale screening with markers covering most of the whole genome allows linkage analysis with as yet undescribed genes. This approach identified the chromosome 20 localization of the gene for the RW MODY pedigree (9,10). It has so far not identified other genes for MODY. The method should only be applied in large pedigrees with a defined phenotype, such as MODY.

2. Sib-pair analysis. As for candidate genes, the exclusion mapping approach can be applied to the most optimal collection of affected sib-pairs available, as shown in type I diabetes (78). As for any large-scale screening technique, false-positive low-order statistically significant results will be obtained, and considerable investigation with adjacent mark-
ers will be needed to determine which positive associations are real and which are by chance. Finding similar results in subsequent collections of sib-pairs provides important verification.

3. Population association study in isolated inbred communities. Since an isolated inbred community has a greater chance than outbred communities of having a high proportion of diabetic subjects being derived from a single common mutation, it may be feasible to pursue the exclusion mapping approach with many markers to detect linkage disequilibrium. Whereas in the general population, a false-positive linkage disequilibrium is more likely than detection of a causative gene, in inbred communities the balance might be reversed, since a single gene mutation might account for a greater proportion of the diabetes.

ASSESSMENT OF CLINICAL SIGNIFICANCE OF GENE MUTATIONS

When a mutation has been identified, it is necessary to show that it is causing the disease in a patient or pedigree. Many silent mutations or polymorphisms occur, and linkage between that mutation and diabetes does not exclude the possibility that the identified mutation is silent and that it co-segregates with a neighboring pathogenic mutation. Three complementary approaches are feasible.

1. Expression of the mutation in an in vitro system. The demonstration that a mutation is not expressed efficiently is the classical and ultimate method. Thus, Permutt et al. (79) have shown that an insulin promoter mutation reduced expression in BTC cell lines, and Gidds-Jain et al. (80) demonstrated low activity of glucokinase mutations.

2. Linkage in pedigrees with diabetes. Once the mutation has been found in a subject, first-degree relatives should be screened for the same mutation by an allele-specific polymerase chain reaction method. This will determine whether the mutation co-segregates with the relevant pathophysiological abnormality. In view of the likelihood that diabetes is a polygenic disease, a defined mutation may not necessarily be clinically expressed as diabetes in every family member.

3. Mutation frequency in diabetic and non-diabetic subjects. A statistically increased prevalence of the mutation in diabetic subjects suggests that it has a role. This method is directly comparable to a population association study and the same caveats apply.

ANIMAL MODELS

In polygenic disorders, animal models offer an excellent approach to genetic analysis. For type I diabetes, the availability of the NOD mouse has indicated at least three separate genes, and the search for homologies in humans is a most useful resource. For type II diabetes, there is no definitely applicable animal model, although the GK rat may be suitable (81). The latter was obtained by repeatedly back-crossing offspring of Wistar rats who had glucose intolerance. After 32 generations, a stable nonobese type II diabetes model was achieved. That it took so many generations to achieve in an inbred strain indicates the polygenic nature of diabetes in this model. The search for mutations that induce obesity in rodents is likely to be relevant to this important contributor to type II diabetes (82).

In summary, type II diabetes is still a geneticist's nightmare. The major problem is the identification of suitable pedigrees, sib-pairs, and populations for study. Clinicians and physiologists need to identify better methods of assessing the phenotype and then identify populations or pedigrees with a specific subtype of diabetes.

ACKNOWLEDGMENTS

These studies have been funded by the Medical Research Council, the Wellcome Trust, the Alan and Babette Sainsbury Trust, and the Juvenile Diabetes Foundation. We are grateful for discussions with many colleagues.

REFERENCES

21. Hailes CN, Barker DP: Type 2 (non-insulin-dependent) diabetes mellitus:


