During our search for a marker for non-insulin-dependent diabetes mellitus (NIDDM) in a large multigenerational family with a form of NIDDM termed maturity-onset diabetes of the young (MODY), we learned a great deal that may serve to streamline the search for diabetes-susceptibility genes in other families. We describe here our experience and suggest strategies that may enhance the search for markers for other diabetes susceptibility genes with genetic linkage approaches. Diabetes 41:401-407, 1992

Genetic factors play an important role in the development of non-insulin-dependent diabetes mellitus (NIDDM; 1,2). However, the search for susceptibility genes has been hampered by the complex, non-Mendelian mode of inheritance of this disorder, i.e., the segregation of NIDDM in families does not conform to simple autosomal or sex-linked dominant or recessive inheritance. Nonetheless, emboldened by the success of studies that have identified the genes responsible for monogenic disorders such as muscular dystrophy (3) and cystic fibrosis (4), we and others have begun to use genetic approaches to identify genes that contribute to the development of NIDDM. However, finding the genes responsible for NIDDM, and there are likely to be several, among the 3 billion base pairs (bp) and ~100,000 genes that comprise the human genome, is a formidable task and can be compared to looking for a “needle in a haystack.” What is the most efficient strategy for identifying diabetes-susceptibility genes?

ONE LARGE FAMILY IS BETTER THAN MANY SMALL FAMILIES

Because >60 different disorders in man are associated with glucose intolerance and in some cases with overt diabetes (2), it seems likely that NIDDM will be genetically heterogeneous and that mutations at several different loci, either individually or in interaction, may produce a similar phenotype with the features of NIDDM. Such genetic heterogeneity could confound the search for markers for the genes that contribute to the development of NIDDM. Consequently, we decided to focus our initial efforts on studying individual families rather than groups of families because there is likely to be less heterogeneity within a family than between families. However, we also recognized that a common disorder such as diabetes has an appreciable probability of being heterogeneous even within a single pedigree, especially as the pedigree is expanded to include more affected individuals.

We chose families for our study with the following criteria. First, to increase the chances of identifying a marker, we wanted to study a family in which the mode of inheritance of the diabetic phenotype was known. Second, each family had to be large enough to provide statistical support for linkage; e.g., in a fully penetrant autosomal dominant disorder, a nuclear family with 1 affected parent and 1 offspring would be sufficient if the marker and disease are tightly linked and assuming that the mating type is informative (i.e., the affected parent is heterozygous for the marker). Finally, cell lines derived from a significant number of family members had to be available to provide a ready and inexhaustible source of DNA because the number of markers that would need to be typed was unknown and we did not want to run out of DNA during the study. Also, the availability of cell lines ensured that the study would not be compromised by the death of key individuals during its course. The RW family fulfilled these requirements.
The RW pedigree is a white family that has been studied continuously since 1958 (5,6, Fig. 1) and has a form of NIDDM termed maturity-onset diabetes of the young (MODY). The segregation of NIDDM in this family is consistent with autosomal dominant inheritance of a highly penetrant diabetes-susceptibility gene. Forty-nine members of the RW family have NIDDM, and 38 of these are still alive, and lymphoblastoid cell lines or DNA are available for 39 members. Cell lines and/or DNA are also available for 111 nondiabetic members.

THE SEARCH
Most subjects with NIDDM in the RW family are characterized by a reduced and delayed insulin secretory response to glucose, particularly among the offspring of II-2 and II-5 (5,6). A striking feature of the diabetic subjects in the RW family is their remarkable response to sulfonylurea drugs for long periods of time (5,6), although about one-third became insulin requiring (not insulin dependent) within 3–25 yr. Insulin resistance is not a feature of the diabetes in this family (5,6). Insulin tolerance tests performed in seven diabetic offspring of II-5 showed normal sensitivity to insulin. Insulin requirements to achieve euglycemic control in seven diabetic patients of generations III–V were 15–35 U in six patients and 46 U in the seventh patient. The number of insulin receptors on monocytes and cultured fibroblasts were normal, and there was normal amino acid transport. With the Bergman minimal model method, preliminary studies suggest an absence of insulin resistance in nondiabetic offspring of II-5 in generations IV and V who have the at-risk adenosine deaminase (ADA) marker for MODY (W.H. Herman, F.J. Ortiz-Alonso, J.B. Halter, G.I.B., S.S.F., unpublished observations).

Although clinical studies of diabetic subjects in the RW family suggested that the gene responsible for MODY may be expressed in the insulin-secreting β-cell, they did not provide any clues as to its identity. Thus, we resorted to a strategy for cloning a gene whose product is unknown, which was originally called reverse genetics and is now termed positional cloning. With this approach, one uses genetics to identify a marker that segregates with the disease gene. The marker can be a protein blood group polymorphism or a DNA polymorphism. Having identified a linked marker, you then identify others in the vicinity and generate a genetic map in the region of the disease gene. Overlapping DNA fragments that span the likely location of the disease gene are isolated, and the genes in this cloned DNA region are characterized to determine which may be responsible for the
disease. The identification of the mutation that impairs the function of this gene and causes the disease phenotype completes the story. The strategy of positional cloning is time consuming and fraught with many false leads. However, if there is a pedigree of sufficient size and adequate resources, it is only a matter of time until 1 of the >2500 restriction fragment–length polymorphisms (RFLPs) that have been described is found to be linked to the disorder. We tested 79 markers before we identified a DNA polymorphism in the ADA gene on chromosome 20 that was linked to MODY (7); we were relatively lucky in our search for a DNA marker. Because the human genetic map was only fragmentary when we started, our search was not systematic. A useful genetic map of highly polymorphic markers is available for many chromosomes (8), and this information can and should be used as a guide to select markers for testing. Theoretically, 150 markers spaced about 20 cM (~20 million bp) apart should be adequate to cover the human genome (9), and a collection of highly polymorphic probes suitable for screening studies have been described for many chromosomes (8).

HARDLY ANYTHING IS AS SIMPLE AS IT SEEMS

The typing of family members for different polymorphisms is only part of a linkage study. These data must be analyzed to identify those markers that cosegregate with the disease gene. The analysis is carried out with computer programs such as LIPED (10) or LINKAGE (11), the latter of which is relatively user friendly and requires only a basic understanding of human genetics and familiarity with personal computers. These programs require that several variables be specified, including mode of inheritance of the disorder and frequency and penetrance of the disease gene. Our analyses of linkage in the RW family also included an age of onset correction because some younger individuals who were phenotypically normal could be at risk but unaffected because of their young age.

MODY has been defined as NIDDM in the young and autosomal dominant inheritance (5,6). However, it was apparent at the start of the study that it would be necessary to allow for the possibility of incomplete penetrance, because there was one "unaffected" female member among the descendants of II-5 (subject 8072) who was 34 yr of age and had an affected parent and an affected child. However, at ages 16–17 yr, she had one glucose tolerance test characteristic of diabetes and one characteristic of impaired glucose tolerance (IGT). A second unaffected female, 55 yr of age (patient 8364), had an affected parent and a child with IGT. Careful examination of the whole pedigree also revealed additional features consistent with the possibility of a more complex mode of transmission than otherwise expected. Although the overall proportion of affected individuals in those at risk of MODY did not differ significantly from 0.5, which is consistent with an autosomal dominant mode of transmission, the proportion of affected females in those at risk was significantly less than 0.5 (15 affected of 44 at-risk members; \( \chi^2 = 4.4; P < 0.05, 1[df] \)). In addition, the risk to offspring of affected females was significantly less than the risk to offspring of affected males.

Although our analyses of the RW family suggested that the MODY gene may not be completely penetrant (i.e., not all individuals who inherited this gene had NIDDM), we tested our markers for linkage to the MODY gene with two models. In the first, our highly penetrant model, the penetrance of the MODY gene was assumed to be 0.95 and 1 in women and men, respectively. In the second model, a model of reduced penetrance for the MODY gene, the penetrance of the MODY gene was assumed to be 0.60 and 0.80, respectively; these values were estimated from data on the younger generations because they were the best characterized (12).

We selected the RW family in part because we assumed that the NIDDM seen in a single family was more likely to have a common etiology. In our initial tests for linkage, we assumed that NIDDM was due to the presence of the MODY gene. Moreover, because IGT is a feature of the natural history of MODY, we believed that a significant fraction of individuals with IGT were in the process of developing diabetes. Initially, because of this optimistic view, we did not fully appreciate that the NIDDM in the RW pedigree was genetically heterogeneous. We know that there are several different causes of glucose intolerance in this family, one of which is the MODY gene. When we came to this realization, we redefined the MODY phenotype for linkage analysis to take into account other features of the MODY phenotype in addition to glucose intolerance. Specifically, because prospective studies indicated that most susceptible individuals could be identified by 25 yr of age and many between 9–14 yr of age, on the basis of a diabetic oral glucose tolerance test or the presence of fasting hyperglycemia, we specified that only those individuals who had NIDDM diagnosed before age 25 yr had MODY. Individuals with NIDDM diagnosed after age 25 yr could be affected because they had the MODY gene but were not tested before age 25 yr or they had inherited some other susceptibility gene; the affection status of these individuals was coded as unknown with regard to MODY. The major factor influencing the probability that an individual with unknown affection status carries the MODY susceptibility gene is whether they appear to transmit it to any offspring. With these strict diagnostic criteria, MODY is present in the offspring of II-2 and II-5. In contrast, the six offspring of II-3 (Fig. 1) who are diabetic were diagnosed after 25 yr of age and among them they have 27 tested offspring (9 with fasting plasma glucose levels only), of which none are diabetic and 21 were ≥25 yr of age when tested. Thus, the diabetic members of this branch of the family probably do not have MODY.

Insulin values were also available for all subjects who had glucose tolerance tests at least since 1962. This additional information may have been useful in distinguishing subjects with MODY from those with NIDDM due to other causes because most individuals with MODY were hypoinsulinemic. For example, two of the affected individuals among the descendants of II-3 (subjects 8091 and 8278) had exaggerated insulin responses to glucose consistent with the idea that they did not have
MODY but rather another form of NIDDM. Thus, had we used the information on insulin levels in diabetic patients in the RW pedigree more effectively, we may have recognized the heterogeneous etiology of the NIDDM in this family much sooner. Note that once the heterogeneity of etiology of NIDDM in the RW pedigree was recognized, the genetic complexity was reduced; e.g., in the branches of the family segregating for MODY (II-2 and II-5), there were no differences in penetrance between males and females or risk to offspring according to sex of affected parent.

DIAGNOSTIC CONUNDRUMS

What is the effect of misdiagnosis in a linkage study? Misdiagnosis always leads to a decrease in the evidence for a true linkage and an increase in the estimate of the recombination fraction. In linkage analysis of a fully penetrant autosomal dominant disorder (normal homozygote penetrance of 0 and heterozygote and susceptible homozygote penetrance of 1), the magnitude of that effect will be exactly the same whether the misclassification involves misdiagnosing an affected individual as unaffected or an unaffected individual as affected. However, asymmetries in the magnitude of this effect can arise when penetrances are relaxed. For example, if the penetrance of the normal homozygote remains 0 but the penetrance of the heterozygote and susceptible homozygote is reduced (i.e., <1.0), misdiagnosing an affected individual as unaffected has less consequence than misdiagnosing an unaffected individual as affected. Conversely, if the penetrance of the normal homozygote is allowed to be nonzero but the penetrance of the susceptible genotypes is assumed to be 1, misdiagnosing an affected individual as unaffected is more serious than misdiagnosing an unaffected individual as affected. Although relaxing the penetrance assumptions can mitigate the problems of misdiagnosis, the associated cost is a loss of information and therefore a loss of power in the ability to detect linkage. An alternative approach would be to consider all unaffected individuals as having unknown phenotype—thus, there can be no misdiagnosis of unaffected individuals. However, because unaffected individuals can carry as much information as affected individuals in a fully penetrant dominant disorder, the study of only affected individuals reduces considerably the information and power to detect linkage, without addressing the problem of possible misdiagnosis among affected individuals.

In our study of MODY in the RW pedigree, it was clear from the beginning that the disease was highly but not completely penetrant. Therefore, in our linkage studies, models allowing for incomplete but high penetrance of susceptible genotypes seemed most reasonable. However, whether to allow for phenocopies (i.e., to assign a penetrance to normal homozygotes of >0) was problematic. Eighteen of the diabetic members in the RW pedigree were diagnosed after age 25 yr, and of these, 11 had no offspring with diabetes diagnosed before age 25 yr. Thus, the potential proportion of phenocopies was quite high. Allowing for a high proportion of phenocopies not only reduces the power to detect linkage but also leads to genetic models that can seem inappropriate. For example, assuming that MODY was rare, with a susceptibility allele frequency of 0.0001, and allowing for a normal heterozygote penetrance of 0.1 and heterozygote and susceptible homozygote penetrance of 0.95, yields a predicted population prevalence of 10 and 99.9% of the affected individuals who would not have the MODY susceptibility allele. Thus, not only would the power to detect a linkage be considerably reduced by allowing for this level of phenocopies, but also the genetic model for MODY becomes clearly inappropriate. In our search for markers for the MODY gene in the RW family, the strategy of using strict diagnostic criteria based on age at diagnosis, and considering diabetic individuals who do not meet that criteria as being of unknown affection status minimizes the probability of misdiagnosing unaffected individuals as affected without severely decreasing the power to detect linkage.

How should individuals with IGT be classified? Numerous individuals in the RW family had IGT. We considered these individuals as having unknown affection status. However, because we allowed for incomplete penetrance, we could have considered individuals with IGT as being unaffected, especially those who were noted with IGT on the basis of a single test. In future studies, we will consider only those individuals showing IGT on multiple tests as having unknown affection status. Two individuals in the pedigree had insulin-dependent diabetes mellitus; it is clearly appropriate to consider them as having unknown affection status.

Figure 1 summarizes our understanding of the probable causes of NIDDM in the RW family. There are two unaffected individuals >25 yr old (1 woman [subject 8072] and 1 man [subject 10083A]) whose ages are 34 and 42 yr, respectively, who have an extended at-risk haplotype, which includes the ADA gene and other chromosome 20 markers that flank it. As noted above, 8072 members had one diabetic glucose tolerance test and an IGT test at ages 16-17 yr. However, the reasons why she and subject 10083A do not have diabetes are unknown. Preliminary tests of subject 8072 suggest that she may have increased insulin sensitivity; however, she also developed gross hyperglycemia during a prolonged glucose infusion. There is one individual whose age is <25 yr (patient 10018) who has the at-risk DNA marker for the MODY gene but normal glucose tolerance when the linkage study was initiated; he is now 20 yr old and has IGT (he also had IGT at age 10 yr). One newly typed individual (patient 11090, age 10 yr) has the DNA marker that is tightly linked to the MODY gene and is at high risk of developing NIDDM. Because of the tight linkage between the ADA and MODY genes, we use DNA typing to identify individuals in the RW family who are at risk of developing NIDDM because they inherited the MODY gene. Prospective studies of these subjects may indicate the nongenetic factors that influence the onset of glucose intolerance and suggest new approaches for its treatment.
ONE MICROSATELLITE-TYPE DNA POLYMORPHISM IS WORTH A THOUSAND TWO-ALLELE RFLPS

Several different types of RFLPs have been described. These include nucleotide sequence substitutions that change the cleavage site for a restriction endonuclease (13) and the variable number of tandem repeat (VNTR)-type (14) DNA polymorphism. These two types of DNA polymorphisms are characterized with Southern-blot techniques and it takes a considerable amount of time to type 50-100 individuals. In addition, with VNTR-type DNA polymorphisms, it is often necessary to run additional gels to fully exploit the high heterozygosity of this class of DNA polymorphisms in linkage studies. Although multiallele polymorphisms such as VNTRs are more useful for linkage studies than two-allele systems, VNTRs do not seem to be randomly distributed but rather are concentrated near the ends of chromosomes. Several other types of DNA polymorphisms have been characterized including variable number of dinucleotide repeat or microsatellite DNA polymorphisms (15,16) and Alu variable poly(A) DNA (AluVPA) polymorphisms (17,18). These DNA polymorphisms appear to be randomly distributed. They are typed with the polymerase chain reaction (PCR; 19), and PCR typing of 100 samples may be completed in several days. As with VNTRs, the microsatellite and AluVPA DNA polymorphisms are often highly polymorphic. This increases their usefulness for linkage analysis and allows for more-rapid exclusion mapping (i.e., fewer markers have to be typed per distance excluded) in the absence of linkage and easier detection of linkage when present. Another advantage of polymorphisms that can be typed with PCR is that they do not rely on the willingness of investigators to provide probes because all that is required are oligonucleotide primers, which can often be easily prepared by core facilities in most universities. It is also possible to include several sets of primers in the PCR reaction and thereby type several loci at the same time—multiplex PCR. It seems likely that a series of primer pairs that are evenly distributed throughout the human genome will be available in the future and that these will speed up the identification of markers for disease genes. Although there may not be a 1000-fold advantage of microsatellite-type polymorphisms over other forms, they have many advantages and will likely replace other DNA polymorphisms in the first stages of mapping disease genes.

IT IS EASIER TO MISS A LINKAGE THAN TO GET A FALSE POSITIVE RESULT

This is not a new lesson and is well known to geneticists but deserves some discussion in this context. If we had assumed that all individuals with NIDDM had the MODY gene and that the MODY gene was highly penetrant, typing our initial pedigree sample of 50 individuals with the two-allele PstI RFLP at the ADA locus (there are no recombinants between ADA and the MODY gene in the RW family) would have given a maximum log of the odds (LOD) score of 1.1 at a recombination fraction of 0.2. We would have investigated a LOD score of this magnitude and possibly made the discovery of linkage with further probing. However, a more likely scenario is raised by the results we would have obtained with a marker such as D20S14. With D20S14, the maximum LOD score is 0.36 at a recombination fraction of 0.35 and linkage can be rejected (i.e., the LOD score is < -2) for all recombination fractions <0.1. Because the sex-averaged recombination fraction between D20S14 and ADA is 0.12, it is unlikely that we would have devoted additional attention to this region of chromosome 20. Thus, testing for linkage with different genetic models and diagnostic schemes for MODY and the high heterozygosity of the AluVPA DNA polymorphism in the ADA gene were crucial in facilitating the identification of a marker for MODY.

We tested 79 markers for linkage with MODY with several different genetic models for linkage with MODY. Although a LOD score of 3 is usually taken as support for linkage, whether this value is appropriate when testing many markers with several different genetic models is unknown. Simulations can be carried out to assess the significance of a particular LOD score when analyses have been carried out with several different parameters. Moreover, simulations can also be used to estimate the likelihood of detecting linkage in a pedigree of interest at different recombination fractions and should be performed with every pedigree being studied to determine how useful it will be for linkage studies.

We conducted prediction studies as suggested by Weeks et al. (20) to confirm the significance of our linkage between ADA and MODY in the RW pedigree. In these analyses, a marker with the same allele frequencies as the ADA marker but unlinked to the MODY gene was simulated to segregate through the original pedigree structure, and linkage analyses were conducted under each combination of model and diagnostic scheme. In 10,000 simulations, the maximum LOD score obtained was 3.16, with the highly penetrant model and strict diagnostic scheme, which suggested that our original results (i.e., a LOD score of 5.25) were highly significant even considering the four ways in which the data were examined. In addition, only 2 of 10,000 simulations gave a LOD score >2.33, which was the maximum LOD score obtained between ADA and MODY with the less-stringent diagnostic criteria for definition of MODY and the incompletely penetrant model. However, the critical test of our original linkage discovery (7) came through increasing the information from the RW pedigree. The maximum LOD score on the original data set was 5.25 at a recombination fraction of 0 between ADA and MODY with strict diagnostic criteria for the MODY phenotype (i.e., that only those individuals diagnosed with NIDDM before age 25 yr had MODY, and those with NIDDM diagnosed after age 25 yr were considered as affection status unknown) and penetrance of the MODY gene in males and females of 1 and 0.95, respectively. By collecting additional family members and using other markers so that all key individuals were fully informative for linkage, the LOD score increased to 10.2 (this represents odds in favor of linkage of >10 billion to 1) at a recombination fraction of 0 with the strict diagnostic criteria and assuming a penetrance of the MODY gene of 0.95 in both men and women. It was gratifying to find that
all of the additional data were consistent with our original findings.

FUTURE DIRECTIONS
Because ADA and MODY are tightly linked, it is possible to identify young individuals with normal carbohydrate metabolism in the RW family who are at high risk of developing NIDDM because they inherited the MODY gene. Prospective studies of these individuals may reveal the nongenetic factors that influence the age at onset of MODY. Moreover, given the variable length of time between age at diagnosis of diabetic glucose tolerance, or even glucose intolerance, and the appearance of fasting hyperglycemia, an understanding of the factors that influence the development of diabetic glucose tolerance may suggest therapeutic strategies that may delay the progression of the disease.

We are using the linkage information to isolate the MODY gene and identify the mutation that impairs its function. Once we have isolated this gene, we will be able to assess its overall contribution to the development of NIDDM in the population as a whole. In this regard, in collaboration with numerous laboratories that have characterized families with MODY-type NIDDM, we have not identified a second family in which NIDDM and ADA are tightly linked. These results underscore the heterogeneous nature of the MODY phenotype (5,6,21) and suggest that the gene responsible for MODY in the RW pedigree should be termed MODY1. We have also determined that NIDDM and ADA are not tightly linked in Pima Indians.

Based on our success of identifying a DNA marker linked to MODY in the RW family, we have begun to study other diabetes-prone families. Because the studies of other families will not have a base >30 yr of continuous clinical characterization on which they can rely, their success will depend on how well they are designed. In this respect, characterization of the diabetic phenotype is particularly important. In addition to information on age, weight, height, and age at diagnosis of NIDDM, we believe that all individuals available for study in the family should be tested with a standard oral glucose tolerance test with measurements of fasting and 30-, 60-, 90-, and 120-min glucose and insulin concentrations. The results of these tests will reveal affected and asymptomatic individuals and will provide other information about the diabetic phenotype in the family, including the nature of the insulin secretory response and inference as to the presence of insulin resistance. Lymphoblastoid cell lines should be prepared from all individuals to provide a source of DNA. The search for markers that segregate with NIDDM might begin with candidate genes such as those for insulin, insulin receptor, islet amyloid polypeptide, glucokinase and the glucose transporters. The evolving genetic map of the human genome, including the map of microsatellite DNA polymorphisms, should be used to select loci for testing. For many chromosomes, it is possible to select a series of loci whose typing would allow that chromosome to either be included or excluded as the site of the diabetes-susceptibility gene. The linkage analysis for MODY-type forms of NIDDM is relatively straightforward. This form of NIDDM has an autosomal dominant mode of inheritance. Because there is little loss in power of detecting a linkage or for exclusion mapping, the MODY gene should be assumed to be incompletely penetrant, and a value of 0.85 in men and women seems both conservative and appropriate based on our analyses in the RW family. In large families, the possibility of genetic heterogeneity should be considered, and individuals whose affection status is not clear should be classified as unknown. Individuals with a single IGT test can be considered as unaffected because the analyses allow for incomplete penetrance of the diabetes-susceptibility gene. The linkage studies in the RW family examined the cosegregation of markers with NIDDM. However, it is also possible to examine cosegregation of markers with a quantitative trait. In fact, in some families, it may be more appropriate to study the genetics of a quantitative trait such as fasting or stimulated glucose or insulin concentrations than of NIDDM. Although it may be easier to identify diabetes-susceptibility genes in MODY than in the more common form or forms of NIDDM, our ability to analyze the genetics of complex disorders that result from the interactions between two or more genes is rapidly evolving and families with non-MODY forms of NIDDM should not be neglected. The time is opportune for studying the genetics of NIDDM.

ACKNOWLEDGMENTS
This study was supported by the Howard Hughes Medical Institute and National Institutes of Health Grants AM-20527 and DK-20595.

REFERENCES


