Perspectives in Diabetes

Positional Cloning Works!

Identification of Genes That Cause IDDM

Stephen S. Rich

The strategy of positional cloning has been highly successful in identifying a number of single gene disorders that exhibit clear Mendelian patterns of inheritance. Positional cloning of insulin-dependent diabetes mellitus (IDDM) has been assisted by the expansion of molecular genetic tools and highly informative markers, so that new IDDM susceptibility genes are being uncovered. The ultimate characterization of IDDM susceptibility, including the manner in which the individual genes interact to determine the genetic component of susceptibility, may be hindered by the complexity of the disease itself. Recent developments in analytic and experimental genetics have renewed enthusiasm in the use of identity by descent (state) methods that use affected relatives (sib pairs) rather than pedigrees as a fundamental tool of gene mapping. Given the relative position of IDDM susceptibility genes, major hurdles in understanding the roles of the identified genes in defining genetic susceptibility as well as their function lie ahead. Diabetes 44:139-140, 1995

Significant progress has been made recently that advances our basic understanding of the genetic factors that contribute to insulin-dependent diabetes mellitus (IDDM) susceptibility. The use of the positional cloning strategy to uncover IDDM susceptibility genes points to the power of the approach and provides hope that similar progress can be made in the identification of non-insulin-dependent diabetes mellitus susceptibility genes.

Although population data and studies of familial aggregation have supported the existence of genetic factors underlying IDDM susceptibility, no clear underlying model to explain genetic susceptibility has been demonstrated. Most genetic modeling efforts have suggested the presence of a single major locus with some effects on modifying genes that govern the inheritance of IDDM and with some evidence suggesting unlinked modifying genes (1-5). For over a decade, only one major candidate region, the major histocompatibility complex human leukocyte antigen (HLA) on 6p, and, more recently, one contributing candidate region (the insulin gene) have been clearly identified as important in the genetic risk for developing IDDM (6-11). Approaches to identify IDDM susceptibility genes by the model-dependent approach, therefore, appeared to be limited by the uncertainty of the model and the availability of suitably large and informative families.

Whether anonymous DNA markers or candidate genes are used, the purpose of the IDDM search is to determine the optimal number of “experimental units” (families, sibships, and their composition) and the saturation level of the genetic map to identify IDDM genes. Since pedigree material is limited and the mode of inheritance of IDDM is not known with certainty, the mapping methods that appear most applicable are those based on affected relative (usually sibling) pairs. These methods rely on the proportion of alleles at a genetic marker locus that are shared by the affected relatives (12-13). Using affected relatives also avoids some of the problems of model-dependent analysis (lod scores) in that no genetic model needs to be specified and information from unaffected relatives need not be used in the test statistic. In a complex disease such as IDDM, family members may be unaffected because they are not at risk genetically (no disease genes) or because they are at risk genetically but are at low environmental risk.

The increase in interest in the affected relative pair methods has accompanied the discovery of multitudes of highly polymorphic marker loci based on simple tandem repeat polymorphisms (14-15). In this fashion, affected relatives are genotyped with highly polymorphic markers and scored for the number of marker alleles that they share in common (either none, one, or two). This distribution is then compared with that expected under the null hypothesis of “no linkage,” dependent upon the probabilities that pairs of relatives share zero, one, or two alleles identical by state. Previous work has demonstrated that, for studies of affected sib pairs, the critical statistic that guides the number of pairs needed for reasonable power is the ratio of risk of IDDM in siblings to the risk in the general population (16). Estimates made from previous studies have suggested that this ratio may be around 15, depending on the population (17). This ratio suggests that if IDDM is caused by several genes that act in additive fashion or in some interaction with HLA and insulin that contributes to 50% of the total genetic variation, there may be success in identifying several with ratios around 2.0; however, if IDDM is caused by genes that act in a complex multiplicative fashion or have many genes with small effects, the search may be quite difficult.

From the Department of Public Health Sciences, Bowman Gray School of Medicine, Winston-Salem, North Carolina.

Address correspondence and reprint requests to Dr. Stephen S. Rich, Department of Public Health Sciences, Bowman Gray School of Medicine, Medical Center Blvd., Winston-Salem, NC 27157-1063.

Received and accepted for publication 3 November 1994.

IDDM, insulin-dependent diabetes mellitus; HLA, human leukocyte antigen.

DIABETES, VOL. 44, FEBRUARY 1995
Recent experience has caused significant optimism to be generated in the search for other IDDM genes. Following a genome search using highly polymorphic markers in a series of affected sibling pair families, several IDDM susceptibility genes of moderate effect have now been identified using the sib pair strategy—linkage to chromosome 11q (IDDM4) near the FGF3 region (18–20), linkage to chromosome 15q (IDDM3, 20), and linkage to chromosome 6q (IDDM5) near the ESR region (18). In the January issue of *Diabetes*, yet another IDDM susceptibility locus has been mapped to another chromosomal region, 2q31, near the HOXD8 locus (18a) in a region previously noted of interest by other investigators (18). Again, this 2q31-linked locus suggests interaction with HLA (IDDM1) and the insulin locus (IDDM2) in a fashion reminiscent of IDDM3, IDDM4, and IDDM5. These examples of heterogeneity with respect to HLA genotype suggest complex inheritance and population-specific genetic architecture.

The accumulated evidence now suggests that genetic susceptibility to IDDM is determined by the action of several genes, some of which (IDDM1) have a significant effect and others (IDDM2-IDDM5) that have a smaller effect. The current strategy to isolate and characterize IDDM susceptibility genes by positional cloning, also known as “reverse genetics,” represents a strategy used to identify a disease gene based on its chromosomal location without knowing the biochemical function of its product (21). This work with IDDM represents the first clear demonstration of mapping complex human disease loci and, therefore, serves as a reason for optimism. Note, however, that the search for the individual genes and how they interact to determine susceptibility remains a distant goal.

Mapping represents the first component of a long process to “find” the IDDM genes. Once a chromosomal assignment has been made for a disease locus, the next step is to reduce the size of the region to one that can be managed by physical mapping methods. It is estimated that there are at least 50,000–100,000 genes in the human genome. This means that there are, on average, 16–34 genes in a 1-Mb (megabase) area, assuming equal distribution of genes. In diseases without evidence of interactive effects of genes being important for susceptibility (breast/ovarian cancer, for example), the search for “the susceptibility gene” appears to consist of a standardized, yet arduous, combination of physical mapping, candidate identification, and mutation analysis. Thus, once a disease locus is positioned within a YAC contig, several methods can be used to search for candidate genes and implement the screening of the candidates.

A gene for IDDM should pass a series of requirements. For a single gene disorder, 1) the gene should be expressed in the tissues affected by the disease, 2) sequence changes in the gene should be detected in affected subjects but not in unaffected subjects, and 3) expression of a “wild type” gene in an affected cell line should be able to complement the defect. Unfortunately, the transmission of IDDM susceptibility is complex and represents the joint contribution of genetic and environmental risk. There is, at this stage, little guidance to determine the path for uncovering an interacting series of susceptibility genes.

**ACKNOWLEDGMENTS**

This work was supported, in part, by a grant from the American Diabetes Association.

**REFERENCES**