Section 1: In Search of Diabetes Genes
Challenges and Strategies for Investigating the Genetic Complexity of Common Human Diseases
Stephen S. Rich¹ and Patrick Concannon²

There is substantial interest in the identification of genes underlying susceptibility to complex human diseases because of the potential utility of such genes in disease prediction or therapy. Type 1 diabetes is an example of one such disorder and is presumed to arise from the effect of multiple genes and environmental factors. One identified locus has a major effect on type 1 diabetes susceptibility (IDDM1), whereas other loci have significant, yet small, individual effects (IDDM2, IDDM15). It is unclear whether susceptibility for type 1 diabetes arises because of the effects of loci acting independently or whether there are important interactions between loci. Although genetic tools are continuing to be developed to enable examination of candidate regions, the means to identify and narrow “true” susceptibility regions continues to be limited by the lack of statistical power resulting from inadequately sized collections of families. This report provides an evaluation of the approaches for identification of regions harboring type 1 diabetes genes, methods to identify the gene regions that interact to define the risk for type 1 diabetes, and efforts to fine-map the variants responsible.

Type 1 diabetes provides a useful paradigm for genetically complex human diseases. Type 1 diabetes arises from autoimmune destruction of the pancreatic β-cells, which are the sole source of insulin. This loss of β-cells ultimately results in an inability to regulate blood glucose levels without the administration of exogenous insulin. Type 1 diabetes is the third most prevalent chronic disease of childhood, affecting 0.3% of the general population by age 20 years and has a lifetime risk of nearly 1% (1). It is estimated that ∼1.4 million people in the U.S. (10–20 million people worldwide) have type 1 diabetes (2,3). In most cases, a preclinical period marked by the presence of autoantibodies to pancreatic β-cell antigens (GAD65, insulin, or tyrosine phosphatase–like protein IA-2) precedes the onset of overt hyperglycemia. This preclinical period provides a theoretical opportunity for preventive intervention in individuals recognized as “susceptible.” The availability of an effective replacement therapy (insulin) that can be used after the onset of disease means that disease prediction in this preclinical stage must be highly specific if aggressive preventive therapies are to be tested or applied. The etiology of type 1 diabetes is unknown, but it is recognized to be due to both genetic and environmental determinants (4).

GENETIC BASIS OF TYPE 1 DIABETES
The genetic basis of type 1 diabetes is complex and likely to be due to genes of both large and small effect. There have been numerous studies investigating genetic susceptibility loci, using both case-control and family study designs. Early studies of disease concordance using twin designs reported higher monoyzotic (MZ) concordance rates than dizygotic (DZ) concordance rates, with MZ rates approaching 50% (5,6). These early studies were likely biased, however, because recruitment of the twins was through advertisement and solicitation; therefore, affected concordant pairs were more likely to participate than discordant pairs. Population-based twin studies confirmed the increased concordance in MZ pairs, but with the concordance of 30–40% compared with a concordance rate in DZ pairs of 5–10% (7,8). Based on the results of these twin studies, it is clear that susceptibility to type 1 diabetes is determined, in part, by genetic risk factors but that probably <50% of the total risk can be attributed to the effects of shared genes. Studies of first- and second-degree relatives of probands also support familial aggregation of type 1 diabetes (9–11). These findings for type 1 diabetes are similar to those obtained for other common
human disorders that exhibit familial aggregation but not simple Mendelian patterns of transmission of risk. Thus, the search for genes contributing to type 1 diabetes risk likely involves multiple genes and environmental exposures and must consider the possibility of interactions between these different risk factors.

**HUMAN MAJOR HISTOCOMPATIBILITY COMPLEX AND IDDM1**

An enduring and statistically significant association between a candidate gene region and risk of type 1 diabetes is the human major histocompatibility complex on human chromosome 6p. The statistical association between the presence of the HLA (B15) antigen and diabetes was first reported by Nerup et al. (12) and was rapidly confirmed (13–15). Continued examination of the HLA region identified extensive genetic complexity, with multiple genes in linkage disequilibrium clustered in a narrow (~6-Mb) physical region. There are several fundamental findings from early studies of the HLA association with type 1 diabetes: 1) HLA is not the only type 1 diabetes genetic risk factor, 2) both HLA class I and class II loci are important factors, and 3) HLA class II loci (DR, DQ, DP) appear to be stronger contributors to type 1 diabetes risk than class I loci (A, B, C).

Two loci in the HLA class II complex (HLA-DRB1 and HLA-DRQ1) have been shown to represent the major determinants of this region, although other class II (HLA-DPB1) and class I (HLA-A, HLA-B, and others) loci may contribute to susceptibility (16,17). It has been suggested that genes in the HLA region may contribute up to 50% of the total genetic risk for type 1 diabetes, based on studies in which allele sharing at markers in the HLA region was assessed in affected family members. However, the effect of HLA may be more than simply to increase risk (18,19). Although there may be multiple genes that contribute to risk of type 1 diabetes, the HLA region (IDDM1), including the HLA-DR and -DQ genes, is the only major genetic determinant. Nearly 95% of type 1 diabetic subjects have one of the DQB1*0201 or DRB1*04, DQB1*0302 haplotypes. Although the heterozygous genotype (DRB1*0301,DQB1*0201/DRB1*04,DQB1*0302) is present in only 2% of the general population, it is present in 30–40% of type 1 diabetic patients in most, but not all, populations (20,21). Much remains to be learned about the manner in which HLA contributes to risk and its interaction with other risk loci.

Despite the overwhelming evidence for a strong effect on type 1 diabetes risk mapping to the HLA region, it has proven difficult to dissect the linked region into specific component genes that confer susceptibility. Studies of diverse (African or Asian) or isolated (Sardinia or Finland) populations can aid in fine-mapping disease variants from the “trans-racial” comparison of haplotypes that are and are not associated with disease (22–24). This approach was pioneered in the study of the role of the HLA region in type 1 diabetes. The presence of a “true” etiologic variant should correlate with susceptibility (or protection) independent of the adjacent haplotype background. This feature, of course, is complicated in the HLA region, where there is extended strong linkage disequilibrium and where multiple loci, each with similar function in antigen presentation and immunoregulation, are associated with disease. Trans-racial mapping demonstrated that each of the three primary class II loci (DRB1, DQB1, and DQA1) has a role in risk (25). These approaches can be applied to non-HLA loci in the search for type 1 diabetes genes and, as an example, have been used in the study of ACE level and the ACE insertion/deletion (I/D) polymorphism (26).

Sardinia is perhaps one of the most carefully studied ethnically isolated populations (particularly with respect to the HLA region) (27,28). HLA-DR3 “marks” one of the two main type 1 diabetes susceptibility haplotypes in European and European-derived populations. However, its effect is thought to be influenced by an unknown number of non-DR/DQ disease loci both within and outside the HLA region because the DR and DQ alleles present on the DR3 haplotype also occur on other haplotypes not associated with disease. In the Sardinian population, the HLA-DQB1, -DQA1, and -DRB1 loci do not account for the entire association of HLA-DR3 (DRB1*0301-DQB1*0201) with type 1 diabetes risk (24,30). Using a stratification/conditional analysis approach, three regions outside DQ-DR, but still within the major histocompatibility complex, were considered “risk modifiers.” The combined impact of these risk modifiers was substantial and highly significant ($P = 6.2 \times 10^{-7}$). Haplotypes defined by single nucleotide polymorphisms (SNPs) within the DMB and DOB genes, and a tumor necrosis factor (TNF) microsatellite locus identified 40% of Sardinian DR3+ haplotypes as non-pre-disposing (31). This unique haplotype distribution and the high population frequency of DR3 found in Sardinia provide unique information that is needed to identify the sources of non-DR-DQ heterogeneity on DR3 haplotypes. Inclusion of selected diverse yet isolated populations (Sardinia, Finland, Japan, and China) with larger admixed populations (Caucasians from Europe, the U.S., and Australia; U.S. African-Americans and Mexican-Americans) should provide further insight and facilitate type 1 diabetes gene identification.

**INS LOCUS AND IDDM2**

A series of early studies (e.g., the study by Zavattari et al. [31]) identified and confirmed an association of the INS “class I” allele of a variable number of tandem repeats (VNTR) polymorphism and type 1 diabetes, with an increase in the homozygote frequency in diabetic subjects. Re-sequencing of a 4.3-kb region around the insulin gene in multiple individuals revealed a number of common polymorphisms. However, none of these markers had alleles more strongly associated with type 1 diabetes than the INS VNTR (33). Genetic studies of allele and haplotype transmission in ~700 combined European and U.S. families (34) revealed that the INS VNTR-associated trait in type 1 diabetes was not recessive (with homozygosity of VNTR class I alleles providing risk). Rather, risk appeared to be largely a dominant trait with the VNTR class III alleles encoding protection from disease. The biological mechanism underlying the genetic risk conferred by the VNTR has yet to be resolved. However, an intriguing clue is provided by the finding that different alleles at the VNTR are associated with increases or decreases in the steady-state levels of insulin mRNA, both in the pancreas and in the thymus. The expression of insulin in the thymus could
GENETICS OF COMPLEX HUMAN DISEASE

potentially influence immune tolerance to insulin and its precursors. The genetic findings are consistent with a model in which various expression levels of insulin in the thymus affect the ability to induce antigen-specific tolerance. Specifically, the class III VNTR alleles are associated with higher levels of INS mRNA in the thymus, which may account for its associated reduction of risk (or protection) for type 1 diabetes (35–38).

GENOME SCREENS TO IDENTIFY TYPE 1 DIABETES SUSCEPTIBILITY REGIONS

One of the first genome scans to identify regions that could harbor major type 1 diabetes susceptibility loci (i.e., those with effects comparable to or larger than HLA) was published in 1994 (39). The first stage screened 96 families with origins in the U.K. using 289 genetic markers spanning the genome at roughly 10-cM spacing. For regions yielding evidence suggestive of linkage (maximum logarithm of odds score [MLS] >1.0), a second sample consisting of 102 families from the U.K. and 84 families from the U.S. were screened with specific subsets of the markers. The first-stage genome screen identified IDDM1 (MLS = 7.3) with the marker TNFα. The evidence for linkage was strong, with an estimated locus-specific genetic relative risk of λS = 2.5, covering 40 cM of chromosome 6p. A second locus with near-significant evidence for linkage was IDDM2 (MLS = 2.1), using the TH/INS haplotype. This region also had a reasonably strong locus-specific effect, estimated to be λS = 1.7 in the sample of families. It should be noted, however, that the power to detect this region may have derived from prior knowledge and the selective genotyping of the INS VNTR. Eighteen other regions throughout the genome had MLS >1.0 (P < 0.05), suggesting that other regions in the genome could contribute to the genetic basis of type 1 diabetes. This result was consistent with the results of linkage studies in the NOD mouse that suggest the existence of a large number of loci, most with modest effects. However, with the exception of IDDM1, and a region on the long arm of human chromosome 2 (IDDM7), there was little overlap between regions identified in the human by genome scanning and those reported in the mouse. In the second phase of the study, the evidence for linkage of type 1 diabetes to IDDM1 remained strong (MLS = 5.5 and 6.7, respectively). However, the evidence for linkage of type 1 diabetes to IDDM2 (TH/INS) weakened significantly, dropping to MLS = 0.4 and 0.1, respectively.

Despite the difficulty in replicating the linkage signal at IDDM2, other regions initially identified through the genome scan with lesser evidence were replicated in the second phase, particularly that on chromosome 11q (FGF3, MLS = 1.3), with evidence strengthening when absence of HLA sharing among affected sib-pairs was considered. In addition, the region on 6q (ESR, MLS = 2.0) was replicated, but in the presence of increased HLA sharing in affected sib-pairs. At the same time, Lathrop, Julier and colleagues carried out a genome scan in a subpanel of 61 mixed French and U.S. affected sib-pairs (40) and also reported linkage to 11q13 (IDDM4), although consideration of HLA sharing resulted in opposite effects in this study. In later combined analyses, evidence of linkage to IDDM4, IDDM5 (6q24), and IDDM8 (6q27) each exceeded genome-wide levels of significance (41,42) in some studies but could not be replicated in others.

A collaboration of investigators used families collected in the U.S. (mainly from the Human Biological Data Interchange [HBDI] collection) to perform a similar genome-wide screen for type 1 diabetes, but with nearly twice the sample size (43). At the same time, the U.K. group studied more families, raising the total of number of U.K. affected sib-pair families who were genome scanned to 356 (44). In the U.S. study, a total of 212 affected sib-pairs were used for the initial screen, and 467 were used for follow-up of regions where there was suggestive evidence of linkage. In addition to the evidence of linkage for IDDM1 (HLA), the only strong linkage observed was for a region on chromosome 1q (near D1S1644 and AGT, logarithm of odds [LOD] >3.0), with no evidence of other loci being significant, including IDDM4, IDDM5, and IDDM8. Although the sample size of this second major effort was twice that of the original, mounting evidence suggested that the non-IDDM1 locus-specific genetic risk ratios were likely to be less than λS = 1.5. This result would require many more families to provide adequate power to detect susceptibility regions. The U.K. study reported evidence of linkage to INS/IDDM2, 10p11 (IDDM10), and 16q, all with P < 7.4 × 10−4.

Simultaneously with the joint analysis of the U.K. and U.S. data, an independent collaborative group of investigators in Scandinavia performed a genome scan to detect genes for type 1 diabetes (45). A total of 424 families (464 affected pairs) with at least two affected relative pairs were ascertained from Denmark, Sweden, and Norway to provide a sample with standardized collection and relatively homogeneous ethnic background. In addition to HLA (IDDM1), the IDDM15 region was detected with a LOD score of 4.8 (D6S300-D6S283). LOD scores >2 were identified for chromosome 2 (D2S113), chromosome 5 (D5S407), and chromosome 16p (D16S407-D16S287). Linkage in the presence of association (using the transmission disequilibrium test) was confirmed for IDDM2 (INS) with an excess transmission of INS alleles from heterozygous parent to diabetic child. After partitioning families on the basis of HLA or INS, there was significant evidence of heterogeneity in the region on 16p (D16S405 with HLA, D16S3113 with INS). In these data, there was also evidence of heterogeneity based on stratification with HLA (IDDM1) for chromosome 2p (D2S113), INS (IDDM2), and regions on chromosomes 4 (D4S142) and 13 (D13S153). Although these data are of interest, the most compelling evidence in favor of genetic susceptibility to type 1 diabetes occurs in IDDM1, IDDM2, and IDDM15. Further exploration of other regions and interactions suggested by inter-region stratification will require significantly larger sample sizes.

JOINT ANALYSIS OF U.S./U.K. GENOME SCREEN DATA

Cox et al. (46) has recently reported on an analysis of a combined data set that includes the previously genome-scanned U.S. and U.K. samples along with 225 newly scanned U.S. samples. In this analysis, there were seven sites at which the multipoint evidence for linkage was suggestive (LOD score ≥2.2, P ≤ 7.4 × 10−4) or significant (LOD score ≥3.6, P ≤ 2.2 × 10−5) by the criteria of Lander and Kruglyak (47). A summary of these results is provided...
TABLE 1
Summary of significant and suggestive linkage in the U.S./U.K. data

<table>
<thead>
<tr>
<th>Region/locus</th>
<th>Chromosome</th>
<th>LOD score</th>
<th>IDDM loci</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA</td>
<td>6p</td>
<td>65.8</td>
<td>IDDM1</td>
</tr>
<tr>
<td>INS</td>
<td>11p</td>
<td>4.28</td>
<td>IDDM2</td>
</tr>
<tr>
<td>D10S565</td>
<td>10p</td>
<td>2.80</td>
<td>IDDM10</td>
</tr>
<tr>
<td>D16S3098</td>
<td>16q</td>
<td>4.13</td>
<td></td>
</tr>
<tr>
<td>D1S2709</td>
<td>1q</td>
<td>2.36</td>
<td>IDDM15</td>
</tr>
<tr>
<td>TNF</td>
<td>1q</td>
<td>2.20</td>
<td></td>
</tr>
</tbody>
</table>

Adapted from Cox et al. (46).

in Table 1. A stratified analysis was performed by calculating correlations between family-specific nonparametric LOD (NPL) scores for pairs of loci that emerged from the initial analysis. Only one potential interaction (1q42 and IDDM2) approached marginal significance. Conditioning on HLA genotypes (families in which all affected individuals were given an HLA genotype were given positive weights) identified an interaction of IDDM1 with 3p12-q13 (LOD score increased from 0.41 to 1.95, \( P = 0.005 \)) when the type 1 diabetic subjects in a family were DR4 negative. A second possible interaction was identified on 10p13 (LOD score increased from 1.38 to 3.66, \( P = 0.0007 \)) when all affected family members had DR3/4. When early age at onset was used as a weight, a region on 6q27 suggested significant heterogeneity (LOD score increased from 0.94 to 3.69, \( P = 0.003 \)), a region that may contain IDDM5.

Another approach to detecting epistasis or heterogeneity uses a conditional logistic regression framework (NPL regression). Using the same original genome scan markers, the combined U.S./U.K. data were analyzed using single-locus and multiple-locus models. Single-locus models produced similar results to that of Cox et al. (46), as shown in Table 2 (LOD scores similar, but not equivalent, to those in Table 1). Also provided in Table 2 are the MLS results performed under Holman’s possible triangle (MLS) and the estimated locus-specific \( \lambda s \), using the ratio of 0.25/\( \lambda s \) as the asymptotic estimate with multipoint identity by descent (IBD) estimation. Non-HLA regions have small locus-specific \( \lambda s \) (1.18–1.30).

Using the NPL regression approach, the genetic marker locations were entered (and removed) in a stepwise fashion. Restricting the loci to only those with significance in the first step allows use of the entire data while testing for significant interactions. Significant evidence for interactions between markers/regions in the genome were observed (Table 3). The sign of the coefficient is important because it reflects the type of interaction present in the data. A negative coefficient represents significantly increased sharing at one locus and no increase (or decreased) sharing at another locus. A positive coefficient represents increased sharing at one locus and increased sharing at a second locus, consistent with an epistatic gene effect.

These results suggest that there may be numerous interactions between previously identified and novel loci. The interaction identified by Cox et al. (46) using the pairwise correlation of NPL scores (1q and 11p) is not observed in this analysis. Because this analysis did not stratify by HLA genotype or early age at onset, the comparisons with the conditional results cannot be made. However, conditional on age at onset, an interaction between 6p (IDDM1) and 6q (IDDM15) was observed. Differences in these two approaches are to be expected because the focus of each approach differs (conditioning on specific important family characteristics or stratification by risk genotype, and modeling NPL statistics in a multivariate sense).

MOLECULAR GENETICS

It has become apparent that to detect genes that, while of biological significance to type 1 diabetes pathogenesis, make relatively small individual contributions to genetic relative risk (such as INS/IDDM2), much larger samples of affected sib-pair families will need to be genome scanned. Under the assumption that support intervals for any localization obtained in such scans will be wide (e.g., 5 cM) and that approximately five non-HLA regions will be identified in such scans, we estimate that further fine-mapping will require genotyping ~2,500 SNPs per family per region. This crude estimate is based on needing to genotype SNP markers for linkage disequilibrium mapping at a density of approximately one marker for every 10 kb and that 1 cM corresponds to ~1,000 kb. It is anticipated that for a complex disease such as type 1 diabetes, not all families will be supportive of linkage in a given region. One may estimate that ~500 families would be supportive of linkage in a given region. These families would have been identified using linkage and family-based association methods in small intervals.

Strategies and methods for fine-mapping and identification of common disease genes continue to evolve. The size of a region will be determined by the curve covering the area of 1 LOD score less than the peak obtained from multipoint linkage (support interval). It is crucial at the initial linkage stage of the effort that a sufficient number of families be studied that provide sufficient confidence that the gene or genes being sought is clearly within the support interval. With sufficiently large sample sizes, it is
Table 3
NPL regression interaction analysis of U.S./U.K. data

<table>
<thead>
<tr>
<th>Chromosome 1</th>
<th>Chromosome 2</th>
<th>β</th>
<th>LRT</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>2q (189 cM)</td>
<td>17q (123 cM)</td>
<td>−0.188</td>
<td>3.387</td>
<td>0.0328</td>
</tr>
<tr>
<td>6p (45 cM)</td>
<td>1q (132 cM)</td>
<td>−0.485</td>
<td>12.348</td>
<td>0.0002</td>
</tr>
<tr>
<td>6p (45 cM)</td>
<td>6p (85 cM)</td>
<td>1.397</td>
<td>71.750</td>
<td>0.0000</td>
</tr>
<tr>
<td>6p (45 cM)</td>
<td>11p (6 cM)</td>
<td>0.635</td>
<td>20.423</td>
<td>0.0000</td>
</tr>
<tr>
<td>6p (45 cM)</td>
<td>16p (41 cM)</td>
<td>−0.403</td>
<td>7.197</td>
<td>0.0036</td>
</tr>
<tr>
<td>6p (45 cM)</td>
<td>16q (100 cM)</td>
<td>−0.304</td>
<td>4.921</td>
<td>0.0133</td>
</tr>
<tr>
<td>11p (6 cM)</td>
<td>1q (80 cM)</td>
<td>−0.214</td>
<td>4.905</td>
<td>0.0169</td>
</tr>
<tr>
<td>11p (6 cM)</td>
<td>16p (145 cM)</td>
<td>0.211</td>
<td>4.392</td>
<td>0.0181</td>
</tr>
<tr>
<td>11p (8 cM)</td>
<td>17q (123 cM)</td>
<td>−0.169</td>
<td>2.428</td>
<td>0.0500</td>
</tr>
<tr>
<td>16p (100 cM)</td>
<td>16q (41 cM)</td>
<td>0.417</td>
<td>13.391</td>
<td>0.0001</td>
</tr>
<tr>
<td>16q (100 cM)</td>
<td>17q (123 cM)</td>
<td>0.229</td>
<td>4.362</td>
<td>0.0183</td>
</tr>
</tbody>
</table>

β: Conditional logistic regression coefficient; LRT, likelihood ratio test.

It is reasonable that regions of 5 cM can be obtained. With the development of detailed haplotype maps of human chromosomes, it may be possible to avoid the systematic screening of tightly and regularly spaced SNP markers within these regions, as proposed above, in favor of using a selected set of SNPs that tag specific defined haplotypes. An approach using such haplotype-tagged SNPs to cover all common haplotypes across the entire support interval in a given region might significantly reduce the SNP typing load necessary to identify linkage disequilibrium within that region (48). Even before this SNP typing pass, functional candidate genes can be selected for high priority typing. If a functional candidate gene under a linkage peak turned out to be positively associated, focus would turn exclusively to the functional candidate only if the association accounted for all the observed linkage.

Once chromosomal regions with the best evidence for linkage and linkage disequilibrium have been identified through fine-mapping, positional candidate gene approaches can be applied to identify candidate susceptibility loci. Using data from the Human Genome Project and related efforts, candidate 3’-expressed sequence tags and human genes are being identified continuously; therefore, it is likely that many of our candidate regions will have numerous options for investigation. From a physiological point of view, the candidate genes will be chosen based on the known autoimmune process, including genes involved in attributes of the β-cell, cellular immunity, inflammation, or glucose metabolism. Genes involved in any of these or related processes and located in the defined regions will be potential candidates for detailed analysis. Candidate genes to be studied further will be identified by extensive bioinformatics research, using public databases containing mapping and sequence information. SNPs within the candidate genes will be used for intensive evaluation by haplotype-tagged SNPs analysis to maximize efficiency and decrease expense.

The method for testing a candidate gene for association with disease based on the haplotype-tagged approach requires genotyping approximately five polymorphisms spaced throughout the gene. When the genomic sequence of a gene is available, including exon/intron structure and some sequence from the 5’-promoter and 3’-untranslated regions, locating polymorphisms along the candidate becomes straightforward. Polymorphisms are often annotated in the GenBank sequence file whenever the sequencing laboratory encounters variation from multiple individuals. Detailed information will be available for most, if not all, candidates for testing. If multiple entries for sequence tagged sites and 3’-expressed sequence tags for a candidate gene exist in their respective databases, all known sequences can be compared using sequence analysis software to identify new SNPs from the differences between two sequences. In all cases, however, the key to establishing the haplotype-tagged SNPs will be the resequencing of a number of individuals to determine which SNPs contain the majority of the observed variation in the candidate gene. Initial identification of a candidate gene for type 1 diabetes will be tested by the information obtained from linkage, family-based association analyses using the linked families, family-based association analyses using trans-ethnic mapping families (African-American, Hispanic-American, Asian trios), and case and control subjects.

Our major goal was to define type 1 diabetes susceptibility genes. In essence, the question is, “When is a gene our gene?” In the case of complex human disease, this is not an easy question to answer. Certain results are consistent with variation in a candidate gene accounting for susceptibility. First, there is a higher risk for specific variants (as defined by linkage and association). Second, the candidate accounts for the evidence for linkage in a subset (if not all) of families. Third, the gene allows partitioning of families by presence of the variant that alters evidence for linkage and association at other sites, suggesting gene–gene interaction. Other issues related to gene identification include measuring correlated phenotypes (related to autoimmunity or glucose metabolism or inflammation) that demonstrate pleiotropic effects of the variant. Ultimately, the variant will need to be tested in a careful epidemiological and functional setting to evaluate the effectiveness of the candidate on predicting risk and early detection of clinical manifestation of autoimmunity and disease.

### Issues and Outcomes

Obviously one cannot predict how many effects will occur within one linkage peak or the size of each effect if there are multiple, clustered genes, not the allele frequencies of the etiological variants. The strategy of gene discovery for complex human disease is evolving from linkage to fine-mapping to identification. Although still not completely resolved, the strategies underlying linkage are better defined than those for fine-mapping. The important issue for linkage is to have exceptional power to allow the support interval to be defined so that one is “certain” that the gene is within the interval, not outside the interval. For the present stage of gene identification, it is necessary to have DNA re-sequenced, especially the DNA containing the peak of the association, to harvest all detectable SNPs to allow systematic exclusion of nonassociated markers so that every variant that could be the disease variant would be identified.

### Acknowledgments

This work was supported in part by a grant from the Juvenile Diabetes Research Foundation (research grant...
REFERENCES


23. S.S. RICH AND P. CONCANNON


