Panning for gold: genome-wide scanning for linkage in type 1 diabetes

John A. Todd* and Martin Farrall

Wellcome Trust Centre for Human Genetics, Nuffield Departments of Surgery and Clinical Medicine, University of Oxford, Windmill Road, Headington, Oxford OX3 7BN, UK

Received June 14, 1996

Genome-wide scans for linkage of chromosome regions to type 1 diabetes in affected sib pair families have revealed that the major susceptibility locus resides within the major histocompatibility complex (MHC) on chromosome 6p21 (λs = 2.4). It is recognized that the MHC contains multiple susceptibility loci (referred to collectively as IDDM1), including the class II antigen receptor genes, which control the major pathological feature of the disease: T-lymphocyte-mediated autoimmune destruction of the insulin-producing pancreatic β cells. However, the MHC genes, and a second locus, the insulin gene minisatellite on chromosome 11p15 (IDDM2; λs = 1.25), cannot account for all of the observed clustering of disease in families (λs = 15), and the scans suggested the presence of other susceptibility loci scattered throughout the genome. There are four additional loci for which there is currently sufficient evidence from linkage and association studies to justify fine mapping experiments: IDDM4 (FGF3/11q13), IDDM5 (ESR/6q22), IDDM6 (D6S281/6q27) and IDDM12 (CTLA-4/2q33). IDDM4, 5 and 8 were detected by genome scanning, and IDDM12 by a candidate gene strategy. Seven other named loci are not discounted but remain to be replicated widely. Multiple susceptibility loci were expected as genome-wide scans of the mouse model of type 1 diabetes had shown that although the MHC is the major mouse locus, at least 13 genes unlinked to the MHC are involved in the development of disease. Genome-wide scans using 1000 affected sibpair families will be required to be confident that all genes with effects on familial clustering equivalent to the insulin gene locus (λs = 1.25) have been detected. The identification of aetiological determinants requires exclusion of hitchhiking polymorphisms in regions of linkage disequilibrium, as demonstrated for the MHC and the insulin gene loci, and functional studies implicating the disease-associated variant in pathogenesis. Ultimately, targeting of specific candidate mutations in mice by homologous recombination and replacement will be necessary to prove the primary role of any candidate mutation.

INTRODUCTION

Diabetes, the dysregulation of glucose homeostasis, affects about 6% of the general population. The most serious form, type 1 diabetes, which affects up to 0.4% of European-derived populations, is caused by an autoimmune destruction of the insulin producing β cells of the pancreas in children, with a peak age of onset of 12 years. The β-cell destruction is irreversible, and despite insulin replacement by injection patients suffer early mortality, kidney failure and blindness (1,2). The major aim, therefore, of genetic research is to identify the genes predisposing to type 1 diabetes and to use this information to understand disease mechanisms and prevent the total destruction of β cells. The mode of inheritance of type 1 diabetes does not follow a simple Mendelian pattern, and the concordance of susceptibility genotype and the occurrence of the disease is much less than 100%, as evidenced by the 30–70% concordance of identical twins (3,4).

GENOME-WIDE SCANNING

Before the first genome scans, we knew of two susceptibility loci affecting the occurrence of type 1 diabetes: the MHC on chromosome 6p21 and the insulin gene region on chromosome 11p15. Both were initially identified in small case–control association studies, which were later confirmed by other case–control studies and by association (5) and linkage studies (6,7) in families. The risk of type 1 diabetes to a randomly chosen member of the general population is 0.4% but the risk to siblings of affected individuals is 6%, giving a sibling risk/population prevalence ratio, λs, of 15 (8). The MHC (IDDM1; locus specific λs = 2.5 in 300 UK families calculated from the proportion of sibpairs sharing zero alleles identical-by-decent) and insulin gene region (INS/IDDM2; λs = 1.25 in 525 families from various populations; Bennett, J.A.T., unpublished) loci cannot explain this degree of familial clustering, suggesting that other genes exist or that there are shared environmental factors capable of causing clustering (8).

*To whom correspondence should be addressed
Studies in the mouse model of type 1 diabetes (the spontaneously diabetic strain NOD) indicated that genome-wide scanning in human multifactorial disease was feasible (9–11). NOD mouse is a good model of human type 1 diabetes because it develops an autoimmune β-cell destruction leading to insulin deficiency that bears strong similarities to the human disease (12). Genome-wide linkage analyses of progeny from crosses of NOD mice with diabetes-resistant strains have yielded mapping data on the location of at least 13 susceptibility loci, most of which have been replicated and some of which are being fine mapped (11,13). Moreover, insights concerning the mode of inheritance of disease have been gained (10,11):

(i) The MHC, on chromosome 17, is the major diabetogenic locus. It is necessary but not sufficient for the development of type 1 diabetes, and is comprised of at least two separate loci (Idd1 and Idd16). Idd1, which is itself likely to be more than one gene, resides within the class II region and corresponds to amino acid variation in exon 2 of the Iab gene, and a deletion mutation of the Hea gene, which eliminates expression of the IE molecule. Restoration of either wild-type IA and IE expression by transgenesis of NOD mice prevents disease (14,15). Simple transgenic experiments are, however, probably not the best way to test the role of an allele of variable penetrance in a complex disease because of the unpredictability of levels and tissue specificity of expression and the fact that the copies of the cognate gene in the recipient remain intact. Instead, specific mutagenesis by homologous recombination is preferred. The MHC class II genes act through peptide-binding and T-cell recognition but the diabetogenic peptides involved are unknown as are the mechanisms responsible for the protection provided by the wild-type molecules (2,15). The MHC susceptibility gene complex overall encodes a dominant trait with variable penetrance (10).

(ii) None of the non-MHC susceptibility loci are as important as the MHC locus, but collectively, they are equally as important and are necessary for disease development. Many combinations of non-MHC genes in concert with the MHC genes are sufficient to cause diabetes. The non-MHC genes encode dominant traits with variable penetrance (10).

(iii) In the NOD mouse this combination of MHC and non-MHC genes is sufficient to cause β-cell destruction because under germ-free conditions the disease is almost 100% penetrant (16).

(iv) Non-MHC genes interact, and a multiplicative genetic model fits the data well (17).

(v) The more susceptibility alleles at unlinked loci an individual carries the greater the risk of developing the disease. This is consistent with the classic polygenic threshold liability model [to which the multiplicative model is a good approximation (18)]. Such a model would be compatible with the inheritance of human type 1 diabetes as the risk for first-degree, second-degree and third-degree relatives declines rapidly (17,19).

(vi) Relatively few genes (about 10) account for the bulk of the variation in the cross with ever-larger numbers of genes contributing ever-smaller portions of the remaining variance (11).

In 1988 three efforts to collect large numbers of families with more than one case of type 1 diabetes were initiated: the Human Biological Data Interchange (HBDI) collected families from the USA (now at a total of 331) (20), the British Diabetic Association Warren repository encompassed the collections of UK families by the laboratories of Tony Barnett and John Todd (n = 320) (21), and a collection of families from France from Philippe Froguel and Mark Lathrop (n > 130) (22). So far, two genome scans using at least 250 microsatellite markers have been published in 61 (22) and 96 (23) affected sibpair families. The larger of the two studies had a 99.9% chance of detecting a gene with an effect similar to that of IDDM1/MHC (λs = 2.5) at P ≤ 0.001 (maximum LOD score, MLS ≥ 2.3), and aside from the MHC, none was found (23). It was concluded that the presence of a second gene with an effect equivalent to MHC elsewhere in the genome is unlikely (22,23). The absence of other large effects implies that future studies to detect and prove linkage will have to be executed in much larger numbers of families.

For example, none of chromosome 6q was excluded for a gene with λs ≤ 2, even with a map that was 92% informative (Fig. 1) (24). In 429 affected sibpair families, however, almost all of chromosome 6q was excluded for a gene λs ≥ 2 (24). For genes with smaller effects (λs = 1.25) of the order of that obtained for the INS IDDM2 locus none of chromosome 6q could be excluded (Fig. 1) (24). In the 96 family genome scan (about 70% of families fully informative) there was only a 15% chance of detecting a λs = 1.25 locus at P ≤ 0.02 (MLS ≥ 2) (23). We were therefore lucky to detect the INS IDDM2 locus with the marker D11S922, which was part of the genome scanning sets (P = 0.02; in these particular families the λs was 1.7, higher than the overall figure of 1.25) (23).

Moreover, both studies detected a third locus, on chromosome 11q13 linked to FGF3 (IDDM4) (Table 1) (22,23), which, subsequently, has been confirmed (25). At FGF3 Davies et al. obtained P = 0.003 in the 96 families, but only in those sibpairs who shared 1 or 0 alleles IBD at MHC, and HASHimoto et al. reported P < 0.05 in 61 sibpairs (22,23). After follow-up studies, for the FGF3 region Davies et al. obtained P = 6 × 10−5 in 282 affected sibpairs (overall λs = 1.24; Kawaguchi, Nakagawa, J.A.T., unpublished, 23) and HASHimoto et al. obtained P = 0.0008 in 251 affected sibpairs conditioned on sharing of alleles at the MHC/IDDM1 locus (22). Combined data from 596 families provides substantial support (P = 1.5 × 10−9) for IDDM4 (λs = 1.39) without conditioning on MHC (25).

Further positive evidence for linkage of chromosome 6q, detected initially in the UK genome scan (23), has been obtained in additional studies using much larger numbers of families with evidence for linkage at the P < 1 × 10−5 level for both IDDM5 (λs = 1.42 using the marker ESR) and IDDM8 (λs = 1.43 at D6S446) in 418 and 361 families, respectively (25), and P = 0.004 for IDDM5 (multipoint λs = 1.1 and P = 0.0006 for IDDM8

### Table 1. Published IDDM loci: June 1996

<table>
<thead>
<tr>
<th>Most linked markers</th>
<th>Locus name</th>
<th>Chromosome</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA-DQB1,-DRB1</td>
<td>IDDM1*</td>
<td>6p21</td>
<td>(37)</td>
</tr>
<tr>
<td>Insulin minisatellite/VNTR</td>
<td>IDDM2*</td>
<td>11p15</td>
<td>(45)</td>
</tr>
<tr>
<td>D15S107</td>
<td>IDDM3</td>
<td>15q26</td>
<td>(26)</td>
</tr>
<tr>
<td>FGF3. D11S137</td>
<td>IDDM4*</td>
<td>11q13</td>
<td>(22,23,25,26)</td>
</tr>
<tr>
<td>D6S476-ESR-D6S448</td>
<td>IDDM5*</td>
<td>6q25</td>
<td>(23,25)</td>
</tr>
<tr>
<td>D2S5152</td>
<td>IDDM7</td>
<td>2q31</td>
<td>(27)</td>
</tr>
<tr>
<td>D6S281</td>
<td>IDDM8*</td>
<td>6q27</td>
<td>(23–25,52)</td>
</tr>
<tr>
<td>D14S677</td>
<td>IDDM11</td>
<td>14q24.3–q31</td>
<td>(29)</td>
</tr>
<tr>
<td>CTLA-4</td>
<td>IDDM12*</td>
<td>2q33</td>
<td>(31)</td>
</tr>
<tr>
<td>IGFBP2-2,5</td>
<td>IDDM13</td>
<td>2q34</td>
<td>(30)</td>
</tr>
<tr>
<td>GCK</td>
<td></td>
<td>7q</td>
<td>(32)</td>
</tr>
</tbody>
</table>

*Confirmed; other loci require additional studies.
(multipoint $\lambda_s = 1.17$) in 429 families (24) (note that at least 300 families are common to both studies). Based on these current published data we have little doubt that IDDM4, 5 and 8 are true susceptibility loci. It appears that the $\lambda_s$ values for IDDM4, 5 and 8 are unlikely to exceed 1.4, and hence, the detection of these loci in the primary genome scans was indeed fortuitous. Nonetheless, in the absence of high levels of power, genome scans with replication in additional data sets can yield true loci (Table 1).

Field and colleagues were even luckier to detect IDDM4 on chromosome 11q13 because they scanned 20 Canadian affected sibpair families with 50 markers, one of which was FGF3 ($P = 0.07$) (26). Their decision to follow-up this $P$ value was prompted by the presence of a locus within 2 cM of FGF3 that had been linked previously to atopy. Atopy is an immunological disorder underlying asthma and the defects responsible for atopy may overlap with those involved in autoimmunity. They also obtained support for linkage ($P = 0.02$) in 25 Canadian sibpair families for D15S107 on chromosome 15q26 (26). This was followed up by typing D15S87 and D15S107 (6–10 cM proximal to D15S87) in 100 UK families ($P = 0.01$ at D15S107; almost the same families as the 96 families used by Davies et al. (23)) and 125 USA HBDI families ($P = 0.05$), giving evidence of linkage in 250 families ($P = 0.001$) (26). The authors proposed that the D15S107-linked locus be designated IDDM3. However, a more recent combined analysis of 265 families, which includes at least 200 of the 250 families studied by Field et al. (26), provides no significant support for linkage of D15S107 to type 1 diabetes ($P > 0.05$) (25). Results from typing of the 96 UK families for D15S107, which include most of the 100 UK families studied by Field et al., gave 1, 0 IBD values of 59, 55 (Cucca and J.A.T., unpublished). The total currently available 1, 0 IBD data from non-overlapping families is, therefore, 259, 235 ($P > 0.05$). Any gene effect at D15S107 appears to be small. Formal exclusion of a locus with $\lambda_s \leq 1.25$ at D15S107 will be demanding, requiring analysis of very large numbers of families.

Similarly, there is little support from affected sibpair linkage data for IDDM7 on chromosome 2q31 (27): $P = 0.05$ in 435 families (25), although linkage of this region of chromosome 2q may be dependent on the MHC and INS genotype of the affected siblings (28) and on their age-of-onset (27). There is additional support for IDDM7 from family-based association studies (see below). Data for IDDM6, 9 and 10 on chromosomes 18, 3 and 10, respectively, remain unpublished. IDDM11 has recently been assigned to chromosome 14q24.3–q31 using marker D14S67 with $P = 1 \times 10^{-5}$ in 313 sibpairs (29). IDDM11 and 13, newly assigned to chromosome 2q34 (30), both require replication in further studies. The localization of IDDM13 some 20 cM telomeric of CTLA-4/IDDM12 (31) was based on data from only 98 sibpairs and involved conditioning of families not only by MHC but also by the gender of the affected siblings, thereby reducing the number of informative affected sibpairs to less than 50 (30).

Evidence for D2S152/IDDM7 (27), CTLA-4/IDDM12 (31) and for a locus linked to the glucokinase gene (GCK) on chromosome 7q (32) include not only data from affected sibpair linkage analysis but also data from testing for allelic association or linkage disequilibrium.

**LINKAGE DISEQUILIBRIUM MAPPING**

For a disease such as cystic fibrosis in which there is a complete concordance between the presence of the mutation and the diagnosis of disease simple recombination mapping can be used to fine map the location of the gene to intervals that can be cloned,
sequenced and analysed for gene content and mutations. Given the number of families currently available with multiple cases of type 1 diabetes, linkage analysis cannot hope to narrow disease intervals much below 5 cM as thousands of families are required to map genes of typical effect to 1 cM resolution (33,34). In multifactorial disease, therefore, as demonstrated at the MHC and at INS, linkage disequilibrium mapping is the only practical way to fine map susceptibility genes.

The decay of linkage disequilibrium is proportional to the genetic and physical distance along the chromosome; with sufficient time, alleles at linked loci will reach equilibrium and show no allelic associations. The pattern of disequilibrium between alleles of a disease locus and alleles of neighbouring marker loci can, however, be unpredictable owing to confounding population genetic forces including migration and mutation. In general, results in monogenic disease have proved the utility of linkage disequilibrium mapping (35). In populations with turbulent gene pools such as those of the UK or of the USA, linkage disequilibrium is unlikely to extend beyond 2 cM of the aetiologic mutation, and is probably restricted to distances no greater than 1 cM. Hence, saturation of a 5–20 cM region, which was identified by linkage analysis, with microsatellite markers to search for linkage disequilibrium should be productive, provided the populations are not so diverse genetically that multiple founder chromosomes exist. This strategy remains unproved, but published data from fine mapping of both the HLA and insulin gene regions indicate that it should yield decisive data (36,37), albeit these were candidate genes. Regardless of the unpredictable nature of linkage disequilibrium, the most associated allele with disease in a region should be that of the aetiologic mutation.

The transmission disequilibrium test (TDT) (38) is one of several family-based tests of association (and linkage) (39). TDT assesses the deviation from 50% of the transmission of alleles of a marker locus from parents to affected children. Significant TDT results are not obtained unless the alleles being tested are in linkage disequilibrium with an allele of the aetiologic mutation. TDT is a powerful test of association and its application does not require prior data implicating a polymorphism with the disease. It provides an alternative method for testing for linkage disequilibrium to the comparison of gene frequencies in patients and matched controls because it avoids the possibility that the cases and the controls have been drawn inadvertently from different gene pools. If TDTs are carried out in separate populations of different ethnic origins and the results mixed then false-negative results are likely. The ideal data sets for evaluating TDT of marker alleles with disease will be collections of families from separate countries and some effort will have to be expended to ensure that the ancestry of the families within each country is as homogeneous as possible.

For loci outside the MHC and INS regions, evidence for association by TDT has been reported for three loci to date: D2S152/IDDM7 on chromosome 2q31 (27), GCK on chromosome 7q (32) and CTLA-4/IDDM12 on chromosome 2q33 (31). In affected sibling families selected by age-of-onset (one sibling diagnosed under age 17 years, the other under age 29 years) (27) there is evidence for increased transmission of the 228 allele of the D2S152 microsatellite marker (191 transmissions versus 151 non-transmissions; \( P = 0.03 \); note that in table 4 of Luo et al. families were not selected by age-of-onset as was the case in the original report of the association of D2S152 with type 1 diabetes) (25). Selection of families in which siblings have a defined and childhood age-of-onset (27,40) should improve the chances of obtaining positive linkage and TDT results because it is expected that the familial clustering of childhood-onset type 1 diabetes would be greater than later onset disease; the sibling risk of type 1 diabetes through 30 years of age was reported at 6.3% compared with a sibling risk of disease at 31–45 years of 2.4% (41). The analysis by Luo et al. (25), overlooked the value of obtaining significant TDT results in several independent datasets: \( P = 0.003 \) (85 transmissions versus 50 non-transmissions) in USA/HBDI families [which include at least 20 of the 53 USA families from the south-east of the country analysed by Luo et al. (42)] (27), \( P = 0.008 \) in Sardinian families (78 transmissions versus 48 non-transmissions; Hill, Cucca and J.A.T., unpublished) and \( P = 0.03 \) in Italian families (29 transmissions versus 15 non-transmissions) (27). It is reasonable to expect negative data sets using TDT owing to the sensitivity of the detection of linkage disequilibrium to history and genetic diversity of the populations studied, and the likelihood that in ethnically distinct groups different founder chromosomes may predominate carrying different alleles of the microsatellite marker. Confirmation of IDDM7 will require analysis of further polymorphic markers flanking D2S152 to determine the region most associated with type 1 diabetes, and the analysis of additional populations.

Some positive evidence of linkage of GCK on chromosome 7q to type 1 diabetes was observed in two non-overlapping sets of families (mainly from the USA/HBDI and UK/BDA-Warren collections): \( P = 0.03 \) in 186 families (set 1) and \( P = 0.02 \) in 153 families (set 2), giving a combined \( P \) value of 0.003 (32). GCK was considered as a candidate gene and alleles of a microsatellite in the gene, GCK3, were analysed by TDT (32), providing evidence of association in set 1 (\( P = 0.0009 \)), but replication was not achieved in set 2 \( [P = 0.6] \); note that a review of these results states that replication of TDT data was obtained; table 4 of Owerbach and Gabbay (28), but this was not the case]. Combined analysis of GCK3 provided a TDT \( P = 0.001 \) (32). Analysis of other family data sets will be required to extend these findings. The locus has not yet been assigned an IDDM number.

CTLA-4/IDDM12 has been assigned to chromosome 2q33 (31), about 10 cM telomeric from the putative location of D2S152/IDDM7. Chromosome 2q has been a focus of mapping efforts because it probably contains the homologue of the NOD mouse type 1 diabetes susceptibility gene Idi5, which was mapped by linkage analysis to mouse chromosome 1 (43). In addition, there are several strong candidate genes in the region, including the homologous and physically linked pair of T-cell regulatory genes CTLA-4, which is a negative regulator of T-cell proliferation, and CD28, the product of which stimulates T cells to divide on engagement of the antigen T-cell receptor with peptide bound to HLA class II and class I molecules. Linkage to CTLA-4 was reported (\( P = 6.6 \times 10^{-5} \)) in 48 Italian families (31), followed up by a TDT \( P = 0.0001 \) for an A/G polymorphism in exon 1 of the gene in a mixture of 231 multiplex and simplex Italian and Spanish families (31). Significant TDT data were not obtained in the UK and USA multiplex families, which is perhaps not surprising given the ethnic diversity of these populations, and given the high probability that the exon 1 A/G polymorphism is not the aetiologic mutation (it encodes a Thr/Ala substitution in the leader peptide). Strong additional support for the association of CTLA-4 gene region with type 1 diabetes did come from analysis of a large case–control study in Belgium in which the G allele give an odds ratio = 1.6 (\( P = 0.0002 \)) (31). This localization is of special interest because three groups have found evidence that the CTLA-4 gene is associated with susceptibility to another organ-specific autoimmune disease, Graves’ disease (antithyroid):
in a USA case–control study, odds ratio = 2.8, \( P = 0.01 \) (44); in a Chinese case–control study (\( P = 0.04 \)) (31); and in a case-control study from Germany (odds ratio = 1.96, \( P = 0.0001 \); K. Badenhoop, unpublished data). It is possible that there is a susceptibility gene in the region that predisposes to both diseases via an autoimmune mechanism, with \( CTLA-4 \) and \( CD28 \) being prime candidates. Fine mapping by defining the region of linkage disequilibrium around the \( CTLA-4/CD28 \) genes is now required.

**FINE MAPPING**

The most difficult task in genetic mapping is to eliminate very closely linked polymorphisms as candidates for the aetiological mutation in the presence of strong disequilibrium. This requires knowledge of all polymorphisms in the region and the identification of rare haplotypes that do not possess the normal arrangement of alleles so that the association or transmission of candidate alleles can be tested in different haplotype backgrounds. This procedure, which we referred to as ‘cross-match analysis’ allowed exclusion of several polymorphisms-flanking the insulin minisatellite (45) and also helped locate aetiological determinants within the MHC class II region (46). The rarity of these informative haplotypes necessitates very large numbers of families (simplex families can be used or case–control data sets) or analysis of different ethnic groups, for example Afro-Caribbeans or Orientals, in which the population frequencies of certain haplotypes are higher than in European populations (37,46).

In the MHC, sequential disequilibrium mapping led researchers from the locus \( HLA-A \) to the \( HLA-DRB1 \) locus and then to the \( HLA-DQB1 \) locus (37). DNA sequencing and functional studies implicate the HLA-DQ molecule, which is the homologue of the mouse IA molecule, in the aetiology of type 1 diabetes (37). However, HLA-DQ haplotypes cannot explain all MHC associations with type 1 diabetes (47), and it is now evident that both the \( HLA-DQB1 \) and \( -DRB1 \) genes are involved in disease (37). Taking determinants at HLA-DR and -DQ into account, are all the associations of the MHC with type 1 diabetes explained? Probably not. Subdivision of certain DQ-DR haplotypes by alleles at other loci outside the class II region, at the \( LMP, TAP \) and \( DP \) loci, centromeric of the class II region, and, telomERICally, at loci in the class III region (e.g. TNF) or in the class I region, such as the \( HLA-A \) locus may reveal further determinants (37). Needless to say, to yield decisive data these experiments will require very large patient resources because the DQ and DR aetiological polymorphisms will have to be fixed.

**CURRENT GOALS**

At present affected sibpair data sets in the order of 400 families are available for genome scanning, and this is an urgent goal for type 1 diabetes genetics research. The first genome scans were pioneering affairs with modest power. The opportunities for a 400 family genome scan are better. In a genome scan of 400 affected individuals, TDT in multiplex and simplex families and case–control studies (31). As more genes are identified and mapped, the candidate gene approach will become more powerful. A combined consortium-style analysis of 1000 families will be practical within 2 years. By the end of 1996 we anticipate that 1500 affected sibpair families will be under analysis (including the USA/HBDI, UK/BDA-Warren, French, Dutch, German, Norwegian, Italian and Danish collections). In the future a combination of data from the various ongoing genome scans should provide a fairly good picture of the genetics of type 1 diabetes.

These considerations apply to genetic analysis of any common, multifactorial disease. The first-generation genome scans for schizophrenia (49), type 2 diabetes (50), asthma (Cookson, W., unpublished) and multiple sclerosis (51) have just been completed. Our colleagues will face the same challenges and dilemmas as encountered in type 1 diabetes, but, on balance, the results from type 1 diabetes are encouraging and provide useful guidelines for the analysis of other common polygenic diseases.

**ACKNOWLEDGEMENTS**

We thank the Wellcome Trust, the British Diabetic Association, the Medical Research Council and the Juvenile Diabetes Foundation for financial support, our colleagues for discussions and sharing results before publication, and Jin-Xiong She for provision of allele sharing data. J.A.T. is a Wellcome Trust Principal Fellow.

**REFERENCES**


