Review

Genetic analysis of type 1 diabetes using whole genome approaches

John A. Todd

Wellcome Trust Centre for Human Genetics (Nuffield Department of Surgery, University of Oxford), Windmill Road, Headington, Oxford OX3 7BN, United Kingdom

ABSTRACT Whole genome linkage analysis of type 1 diabetes using affected sib pair families and semi-automated genotyping and data capture procedures has shown how type 1 diabetes is inherited. A major proportion of clustering of the disease in families can be accounted for by sharing of alleles at susceptibility loci in the major histocompatibility complex on chromosome 6 (*IDDM1*) and at a minimum of 11 other loci on nine chromosomes. Primary etiological components of *IDDM1*, the *HLA-DQB1* and *-DRB1* class II immune response genes, and of *IDDM2*, the minisatellite repeat sequence in the 5' regulatory region of the insulin gene on chromosome 11p15, have been identified. Identification of the other loci will involve linkage disequilibrium mapping and sequencing of candidate genes in regions of linkage.

The most serious form of diabetes, insulin-dependent or type 1 diabetes, is an incurable autoimmune disease affecting 1 in 300–400 children. Insulin replacement prolongs life but does not prevent the complications of blindness, kidney disease, and neuropathy. The process of T-lymphocyte-dependent destruction of the insulin-producing β cells in the pancreas is overwhelmingly complex (1, 2). In the long term, however, we must understand disease mechanisms so that autoimmunity can be prevented, while maintaining a functioning immune system. Progress in these goals will have enormous benefits for autoimmune disease in general, which affects about 4% of the population, and will also facilitate transplantation research, which faces a similar problem of generalized versus selective/specific immunosuppression. This goal can be achieved by identifying the mutations that predispose to autoimmunity. The genetics of type 1 diabetes is accessible experimentally because of the existence of two excellent animal models, the nonobese diabetic (NOD) mouse and the Biobreeding (BB) rat, and because the disease is strongly inherited in human families with a clear-cut diagnosis and an early age of onset (3).

The inheritance of type 1 diabetes (and non-insulin-dependent diabetes, a more common, but less severe, defect in glucose homeostasis) has remained a mystery (4)

until recent technological advances and large collections of families with multiple cases allowed linkage analysis of the entire genome. Previously, the candidate gene approach led to the identification of two susceptibility gene regions: the major histocompatibility complex (MHC) on chromosome 6p21 (*IDDM1*) (5–7) and the insulin gene region (*IDDM2*) on chromosome 11p15 (8), but how many other loci and their relative importance could not be determined until a whole genome survey had been completed. Recent success (9, 10) reveals that the MHC region, *IDDM1*, is the major locus with upwards of 11 other chromosome regions involved. The interaction (epistasis) (11) or lack of interaction (heterogeneity) (12) of these genes and identification of the etiological mutations will now follow by positional cloning and candidate gene approaches in linked chromosome regions. There is, however, a major hurdle in the positional cloning of diabetes genes: diabetes mutations encode susceptibility; they do not, even in the presence of a full complement of susceptibility alleles at multiple loci, as in type 1 diabetic genetically identical twins (13, 14), always result in the overt disease. The concordance of mutation–phenotype is incomplete, in contrast to “simple” Mendelian diseases such as cystic fibrosis and muscular dystrophy. Therefore, the identification of causal or etiological mutations in multifactorial diseases is difficult, but not impossible (15, 16).

Common Disease Genes

Very serious diseases such as phenylketonuria and hemophilia tend to be highly clustered in families (that is, the sibling risk of developing the disease greatly exceeds the population frequency, the ratio λ_s) because they are caused by rare, highly penetrant mutations (Fig. 1). Such mutations follow a simple mode of inheritance because the presence of the mutation always corresponds with the occurrence of the disease. Any child who inherits one (dominant) or two (recessive) copies of the mutation will develop the condition (Fig. 1). The low population frequency of the disease mutation is due to negative selection of the mutation in the population (unless there is positive selection). On the other hand, common, multifactorial diseases (the so-called “polygenic” dis-

eases; polygenic gene action implies the involvement of many genes with small and additive effects, although the term polygenic has become popularized in the general description of common diseases with an inherited component) such as diabetes or schizophrenia affect millions of individuals but the degree of familial clustering is much less, with λ_s values of 3–20. Type 1 diabetes has $\lambda_s = 15$ (lifetime sib risk = 6%/population frequency = 0.4%) (17), whereas type 2 diabetes, being 10 times more common, only has $\lambda_s = 3.5$ (18). In general, the smaller the λ_s , the more difficult it will be to clone the genes involved (unless a candidate gene is positive). It will also be more difficult to prove beyond reasonable doubt which of the many polymorphisms in a disease-associated region is the etiological one. Common disease genes, therefore, encode susceptibility; they may not be necessary for the development of disease, and, by definition, they are not sufficient to cause the disease.

It is possible that the clustering of type 1 diabetes in families is due to shared environmental factors. Dietary habits could be an example, particularly since nutrition during pregnancy or early in life may influence susceptibility to type 1 diabetes (19, 20). Environmental factors are notoriously difficult to identify, however, so a genome-wide linkage study was a more realistic option to address the question directly: can the $\lambda_s = 15$ be accounted for by shared alleles at different loci? Furthermore, the screen must cover all or most of the genome. In monogenic disease a genome search can be halted when linkage is discovered but this is not the case with multifactorial diseases. Recent genetic analyses of asthma (17 marker loci reported) (21), psoriasis (69 marker loci) (22), and schizophrenia (35 marker loci) (23) have not yet encompassed the whole genome. Positive evidence of linkage was obtained after only a fraction of genome was analyzed. Without data from the rest of the genome it is not possible to estimate how important in relative terms the first linkage is or assess what the overall mode of inheritance might be.

Only recently have highly informative microsatellite maps (24) and semi-automated genotyping technology (25, 26) been available, allowing decisive whole genome screens (9, 10). The power of a

FIG. 1. Inheritance of rare and common diseases. (Figure is adapted from Newcombe, Second International Conference on Congenital Malformations, New York, NY, July 14–19, 1964, The International Medical Congress, Ltd., New York.)

study depends not only on the map and on the technology but, probably most critically, also on the number of families studied, the quality of clinical phenotyping, and the selection criteria (27). Genome screens are of little value if the sample size is small because the false-positive rate will be unacceptably high. The availability of large pedigrees and better opportunities to define phenotypes accurately meant that the first successful whole genome screens occurred in inbred strains of mice and rats and were published some 4 years before the first human genome screens (28–31). Currently, even the most notorious human multifactorial traits such as schizophrenia are yielding interesting linkage data in the face of adequate family numbers and accurate genotyping technology (23). Furthermore, some of the pitfalls of the statistical methods are now better defined and more widely recognized, providing more cautious interpretation of linkage data (32, 33).

One of the main issues is the statistical significance required in a genome screen to declare linkage. With a map of infinitely dense markers, $P \leq 2 \times 10^{-5}$ is required to be confident that the linkage is true at the 5% level (34). Unfortunately, the “waiting time” (35) for $P \leq 2 \times 10^{-5}$ in a multifactorial disease may be long. In contrast, Thomson (36) has suggested that 1×10^{-3} might be sufficient to claim linkage, putting the emphasis on sensitive detection, with a higher false-positive rate. But

we know from the previous genetic analyses of psychiatric disorders including Alzheimer disease that $P < 0.001$ in one data set is not sufficient to claim linkage (32). I think the Thomson criterion is not stringent enough. The strategy we have adopted is a compromise. If $P < 0.001$ is obtained in one data set (again emphasizing that this data set must be in the order of at least 100 sib pair families or equivalent) and, by specifically testing a second data set, further evidence of linkage is obtained at $P < 0.05$, then we consider the linkage is worth reporting and, more importantly in an experimental sense, merits further genotyping.

Once the evidence for linkage of a region to type 1 diabetes reaches this level of support, we then embark on saturation mapping using all available microsatellites in the region and searching for allelic association (or linkage disequilibrium) between microsatellite alleles and alleles at the disease locus (9, 37). Outside *IDDM1* and *IDDM2*, we have now found replicated evidence of allelic association between microsatellites and disease in four different chromosome regions (unpublished), including chromosome 2q31 (*IDDM7*) (37). Since linkage disequilibrium rarely extends beyond 2 centimorgans (cM) (38), a positive, replicated result at this stage helps confirm the primary linkage data but, more interestingly, places the disease locus within a 2-cM or 2-Mb interval.

In our genome screen (map of 289 markers with 11-cM average spacing on 96 sib pair families) we found some positive evidence of linkage at 18 loci, outside *IDDM1* and *IDDM2*. These 18 “sniffs” had maximum logarithm of odds score (lod score; MLS) ≥ 1.0 , which corresponds to $P = 0.02$. This “sniff threshold” was chosen because we wanted to detect loci that gave evidence at $P \leq 0.05$ or $MLS = 0.7$, but we had subdivided the genome screen data by *IDDM1* sharing two alleles versus one or zero alleles identical-by-descent (IBD), so $\log t$ or 0.3 was added, where t = number of tests (39), which in this case was 2. This gives an experimental threshold value of $MLS = 1$. Obviously, even with 96 sib pair families and a fairly good map, many, if not all, could be false positives. But the sniffs provide starting points for further typing.

Furthermore, we (9) and Hashimoto *et al.* (10) reported the results for marker loci that showed less sharing of alleles IBD than expected, which is biologically implausible. These “negative sharing” data are very important because they give an empirical readout on the random fluctuations that are observed during the testing of almost 300 marker loci on 96 families (9). By comparison of these random fluctuations with increased sharing data it was clear that affected United Kingdom sibs tend to share chromosome regions more often than observed by chance: there were 7 increased sharing marker loci with $P < 0.005$ and only 1 locus, *D4S430*, at less than this P value that showed negative sharing. Surprisingly, *D4S430* gave the largest distortion we observed in the entire screen with quasi-MLS = 4.1 ($P = 1 \times 10^{-5}$)! As expected for a false positive, we failed to replicate the data in additional data sets, in contrast to 2 of the positive sharing loci with initial $P < 0.005$ on chromosomes 11q13 and 6q25–q27 (9). This result highlights the essential requirement for independent replication before declaring linkage, even with $P = 1 \times 10^{-5}$. Hamer *et al.* (40) reported evidence of linkage ($P = 1 \times 10^{-5}$) between human chromosome Xq28 and male sexual orientation, but they only studied 40 pairs. The study would have been much more convincing if they had obtained evidence of replication in additional sib pairs.

The replication-based strategy has led to positive evidence for allelic association in our laboratory (37), and in other laboratories (41), which is unlikely to occur in the case of a false-positive linkage. To date, additional evidence for at least 7 of the 18 sniffs has been obtained (unpublished; refs. 9, 10, 37, and 42–44).

Type 1 Diabetes Results: June 1995

Table 1 shows that type 1 diabetes is indeed a multigenic trait, with *IDDM1* being the major locus and at least 11 other

Table 1. Inheritance of type 1 diabetes in United Kingdom multiplex families

Chromosome	Locus	λ_s	λ_s 95% CI	Source or ref(s).
6p21	<i>IDDM1</i>	2.60	1.7–4.2	9
11p15	<i>IDDM2</i>	1.29	1.0–1.7	9, 16
15q	<i>IDDM3</i>			45
11q13	<i>IDDM4</i>	1.07	1.0–1.5	9, 10
6q25 (ESR)	<i>IDDM5</i>	1.16	1.0–1.7	9
18q	<i>IDDM6</i>	1.10	1.0–1.5	T. Merriman and J.A.T.* 9
2q31	<i>IDDM7</i>	1.13	1.0–1.6	9, 37, 44
6q27 (D6S264)	<i>IDDM8</i>	1.42	1.0–2.2	J. Davies and J.A.T.* 42
3q21-q25	<i>IDDM9</i>	1.26	1.0–1.7	S. Gough and J.A.T.*
Xq	<i>DXS1068</i>	1.21		46
10p11.2-q11.2	<i>IDDM10</i>	1.45	1.0–2.2	P. Reed and J.A.T.* 9, 10
7p	<i>GCK</i>			41

λ_s values were calculated, as described (9, 12), from IBD values based on 96–238 United Kingdom families with one affected sibling diagnosed under age 17 years and one under 29 years. Data were not available for this data set for the chromosome 15 and 7 loci, *D15S107* and *GCK*, respectively. Confidence intervals (CI) were calculated by Cordell and Todd (47). Where the lower confidence interval fell below 1.0, the value 1.0 is given (1.0 is the minimum value for λ_s). Assuming a pure multiplicative model, which may not be the case (48), the product of all loci is 16.2, which approximates to the observed familial clustering of type 1 diabetes ($\lambda_s = 15$). In this model, *IDDM1*, for example, accounts for about 35% of this clustering ($2.6 = 15^x$, leading to $x = \log 2.6 / \log 15 = 0.35$), with 95% confidence intervals of 20% and 53% in this particular collection of families from the United Kingdom. *IDDM* symbols are the official nomenclature in the Genome Data Base.

*Unpublished data.

loci contributing (unpublished; refs. 9, 10, 37, and 42–44). For each of these loci there is convincing statistical support, based largely on replication of linkage, and allelic association in multiple independent data sets. This picture is remarkably similar to the NOD mouse in which at least 15 loci have been mapped (2, 49). Under a multiplicative model the products of the NOD *Idd* loci interact epistatically, indicating that *Idd* proteins function in related pathways (50). It is likely that the major pathophysiological pathway in NOD type 1 diabetes is autoimmune destruction of β cells. Breeding experiments and the development of congenic strains have shown that the MHC provides organ specificity for non-MHC gene-encoded defects in immune tolerance (2). One would therefore expect susceptibility genes for different mouse models of autoimmunity to colocalize and this, indeed, may be the case (51). Formal proof of this awaits identification of non-MHC mutations—so far none of the mouse polygenes has been identified, and none of the human ones, except *IDDM2*, which is the minisatellite (or variable number of tandem repeat, VNTR) locus embedded in the 5' regulatory region of the insulin gene on chromosome 11p15 (16).

In humans, analysis of the interactions between genes, or lack of interaction, has just begun (48, 52). Such genetic analysis is problematical because of the very large family data sets required to distinguish between one genetic model and another (in which the predicted sharing probabilities do not differ markedly). However, once marker alleles in strong linkage disequilibrium with the disease locus alleles, or the mutations themselves, are discovered, it will be possible to decipher disease

gene interrelationships much more cost effectively in population-based analyses. Nevertheless, it appears that *IDDM1* and *IDDM2* function epistatically (48, 52), whereas *IDDM1* and *IDDM4* may act independently (48). These results will probably depend on which population is studied as there are well-established interpopulation differences in the inheritance of loci. For example, the relative risk of *IDDM2* is dependent on the presence of the HLA-DR4 phenotype in France (53) but not in all other populations studied so far, including the United States (54, 55), Belgium (56), Finland (57), and the United Kingdom (16, 58).

Table 1 indicates that most familial clustering of type 1 diabetes in families is probably caused by shared alleles at several unlinked loci across the genome. Even if the overall λ_s value of 15 is an underestimate, there are probably more susceptibility loci to be located (9). Taking the lower 95% confidence interval for λ_s , the worst case is that in these United Kingdom families (one sibling diagnosed under age 17 years, the other under 29 years), we have explained so far [this estimate does not include data for *IDDM3* on chromosome 15 (45) or for the *GCK* locus (41)] 24% of familial clustering (including *IDDM1* contributing between 20% and 53%). These results do not imply that the environment is unimportant. It is, because monozygotic twins are only 36% concordant (note, however, twins carrying highly susceptible genotypes are likely to be much more than 36% concordant). Therefore, even if all the susceptibility alleles are present in an individual the penetrance of the alleles is dependent on exposure (or lack of exposure) to environmental factors (59). One way in which the

environment could influence the immune system would be in the establishment of the T- and B-cell antigen receptor repertoires.

Similar conclusions were reached previously by whole genome screening of the mouse genome (28, 60), but in mice, by development of congenic strains, it is possible to delineate the relative contribution of MHC versus non-MHC genes: both sets of genes are necessary but not sufficient (2). Individually non-MHC *Idd* loci are neither necessary nor sufficient for the development of disease. It appears that a combination of defects contributes to a threshold of β -cell dysfunction: this process, particularly in humans, takes a long time, on average at least 5 years. More than 90% of β -cell function must be lost before clinical insulin deficiency is diagnosed. It is highly likely that the pathways that constitute disease initiation, progression, and end stages are conserved in mouse and human.

Gene Identification

Linkage studies are robust but, unless thousands of families are available, the ability to fine map a locus is severely limited (61). In the families available currently ($n = 400$) minimal regions can be defined to about 5- to 20-cM regions that might contain 250–1000 genes (9). Studies in monogenic disease (38, 62–64), between markers in certain chromosome regions such as the T-cell receptor β -chain locus (65), in anonymous regions (66), at *IDDM1* (67), and at *IDDM2* (16) indicate that allelic association extends, in general, not much beyond 2 cM or 2 Mb of DNA. Therefore, if reproducible allelic association of a marker locus is observed, then the disease mutation probably lies within 1–2 Mb of the marker locus.

To our surprise we discovered allelic association between a polymorphic microsatellite on chromosome 2q31, *D2S152* and type 1 diabetes (37). This Génethon microsatellite (heterozygosity = 0.77) has four common alleles, 228, 230, 232, and 240 mobility units [almost equivalent to base pairs as measured on the Applied Biosystems 373A DNA sequencing machine (26)]. The most common allele, 228, was transmitted to diabetic siblings in American families more often than 50% (37). This result was replicated in two other independent family data sets from Sardinia and Italy (37). In Sardinia, all four alleles vary in their transmission, with 228 and 240 being transmitted more often than 50% and 230 and 232 less often than 50%. In French families the same four alleles display the same trend of transmission, although it did not reach statistical significance (M. Delépine and C. Julier, personal communication). Recently, She and colleagues (42) have shown that the 228 allele is also associated with type 1

diabetes in 47 independent multiplex families from the southeastern United States (47 transmissions versus 26 nontransmissions; $P = 0.014$). Nevertheless, in the largest data set from the United Kingdom there is no evidence for allelic association between *D2S152* and disease (37). This sort of result is, however, typical of microsatellites (or diallelic polymorphisms) in linkage disequilibrium with type 1 diabetes. For example, the tyrosine hydroxylase (TH) gene tetranucleotide repeat microsatellite is only 9 kb telomeric of the *IDDM2-VNTR* locus, and yet, despite its proximity to the etiological mutation, alleles of the TH locus are associated (that is, more frequently transmitted to affected siblings in families) in some family data sets but not others (16) (Y. Kawaguchi and J.A.T., unpublished). Different alleles show transmission disequilibrium in different populations. This can be accounted for by the linkage disequilibrium between TH alleles and VNTR alleles and the frequency of TH minisatellite haplotypes (16). Hence, it will not be surprising to observe, in the future, microsatellite alleles close to the other IDDM loci in which alleles show different associations in different populations. Note that recently the impression has been given that linkage disequilibrium can only be carried in genetically isolated populations such as the Finns (61). This is not the case since linkage disequilibrium analyses have been carried out in the *IDDM1* and *IDDM2* regions since 1974 in non-genetically isolated populations. In fact, a number of populations founded at different times (compare the Finns at about 2000 years old with the Sardinians at >10,000 years old) and of different ethnic origin (68) is desirable for linkage disequilibrium mapping in order to, first, detect association, and then, to fine map the most associated region and ultimately the etiological mutation.

A second feature of linkage disequilibrium mapping is that, at least in the first instance, it is much safer to carry out association studies within families, thereby avoiding population stratification problems (69). One example of a number of similar family-based association tests (70) is the transmission/disequilibrium test (71), which we used to detect allelic association between *D2S152* and type 1 diabetes (37) and also to analyze the association of the glucagon receptor and type 1 diabetes (72). However, several authors have given the impression that this test can only be applied "in the presence of a known association." This is not true; we (37) and others (41) have used the test to analyze the association of marker alleles with disease in regions of linkage. The test was first used by Thomson and colleagues (73), and then by Spielman and colleagues (71), to prove beyond reasonable doubt that the association of the insulin gene

region with type 1 diabetes was a true case of linkage disequilibrium and not population stratification.

Once linkage disequilibrium is detected in a region of linkage, microsatellite disequilibrium curves will be drawn across regions of 2-Mb scale by isolating new microsatellites from yeast artificial chromosome, P1, and cosmid clones to fine map the new IDDM loci. Furthermore detailed mapping will require characterization of diallelic polymorphisms (or point mutations), which occur in DNA at about a 200-fold increased frequency of microsatellites. This should allow, as found for *IDDM2*, delineation of the most associated "minimal region" (74) in which candidate genes can be characterized, and haplotypes (16) and genotypes (16, 57, 75), both predisposing and protective, can be defined to permit exclusion of hitchhiking polymorphisms and identification of the etiological mutations [referred to as "cross-match analysis" (16)].

Whereas NOD mice provided essential "proof of practice" that polygenes could be mapped by linkage (28, 60), linkage disequilibrium mapping is not an option in mice. Because *Idd* alleles have low and variable penetrance with a dominant mode of inheritance, the *Idd* loci cannot be mapped using standard outcrosses. Instead, congenic strains in which specific chromosome regions have been introgressed using a microsatellite marker-directed backcross breeding approach must be constructed (2, 76). The disease frequency of the congenic stock indicates whether or not the *Idd* locus or candidate gene is in the introgressed chromosome segment. This strategy has proved highly successful in fine mapping of *Idd3* and *Idd10* in the NOD mouse (2, 76). Ultimately, the minimal intervals will be small enough to embark on positional cloning and candidate gene sequencing. In addition to biological analysis of candidate gene products, it will be necessary to apply embryonic stem cell technology to introduce specific mutations into candidate genes to prove, unequivocally, that a mutation can change the course of disease. Fortunately, NOD mice provide an extremely sensitive model for type 1 diabetes with many characteristics in common with the human disease—sensitive in the sense that almost any manipulation, whether it be by transgenesis or by injecting antibodies or cytokines, perturbs the course of the disease. The NOD congenic mouse gene-targeting approach will also be used to identify human mutations.

The Future

As type 1 diabetes gene mutations are identified, they, and polymorphisms in linkage disequilibrium with them, can be used to assess genetic risk (77). One of the most striking features of the association of

IDMM1 with type 1 diabetes is the very significant protection conferred by the *DQB1*0602* allele—an allele that occurs in >20% of the general population but in <1% of type 1 diabetic patients. Similarly, a genotype at *IDDM2* has been defined that is associated with a 70% reduction in type 1 diabetes risk—a very significant effect for a gene that might only account for about 10% of familial clustering (16). Typing for combinations of the non-MHC protective alleles may help identify children who are at low risk of the disease. Autoantibodies can be extremely predictive of type 1 diabetes (78), particularly combinations of antibodies (78), even in the general population (79). They are thought to occur in most cases before age 5 years (80). Which genes encode susceptibility to early immune damage? There is evidence of heterogeneity in *IDDM1* susceptibility according to age of onset (81). Which non-MHC genes influence the development of type 1 diabetes before age 5 years, which constitutes 25% of all cases?

Most authors agree that position 57 of the HLA-DQ β chain plays a direct role in anti-islet autoimmunity—namely, in peptide antigen presentation (82, 83)—and is a major component of *IDDM1* (3). Polymorphism in the β chain of HLA-DR also plays a role in type 1 diabetes, most likely residue 74 (84). For example, the *DRB1*0403* allele is capable of completely overriding the susceptibility encoded by the *DQB1*0302* allele (84). As such, it is probably worth typing for this allele to identify individuals less likely to develop disease. Note that binding of peptides of insulin B chain, which is an important autoantigen in type 1 diabetes (85, 86), to HLA-DR is, in part, determined by residue 74 of the β chain (87). In contrast, residue 52 of the DQ α chain, although useful for risk assessment (88), is probably only implicated owing to its linkage disequilibrium with the etiological mutations at DQ β chain 57 and perhaps at DR β -chain residue 74. There must be other MHC genes, because of the existence of disease-associated extended MHC haplotypes. These non-class II genes will be difficult to identify owing to the strong linkage disequilibrium between alleles at certain loci in the MHC. In the NOD mouse, congenic strains indicate that another susceptibility locus, in addition to the MHC class II locus, is encoded in the region between *K* and *Lmp2*, but not including the *Tap* genes, affecting both insulinitis and diabetes (89, 90).

It has been reported that HLA-*DRB1*04* is transmitted to diabetic offspring more often from fathers than from mothers. By analysis of a very large United Kingdom family data set and comparison of previously published ones, it is evident that the magnitude of this effect, if operative at all, is small (91). A more reproducible and substantial phenomenon

comes from the observation that the number of type 1 diabetic offspring from type 1 diabetic mothers is significantly less than that from type 1 diabetic fathers (92). This has been attributed to maternal imprinting of *IDDM2* (53, 93, 94). This, however, cannot be the explanation because there is no evidence for maternal imprinting in United Kingdom families (16, 58), and yet in 384 United Kingdom type 1 diabetic multiplex families there are 31 fathers with type 1 diabetes and only 9 affected mothers. Taking all analyses into account, there does appear to be convincing evidence for a maternal imprinting effect at *IDDM2/INS-VNTR* in families from France (53), the United States (16, 94), and Canada (93), in that transmission of VNTR susceptibility alleles is seen more often from fathers than from mothers. In the United Kingdom the converse is true: transmission of susceptibility alleles of the VNTR to diabetic offspring is more pronounced from mothers than from fathers (58), and this result has been replicated (16). A comparison of the combined results from France and the United States with those from the United Kingdom reveals a significant difference between paternal transmission of VNTR susceptibility alleles ($P = 0.004$) but not between maternal transmission ($P > 0.05$). Interpopulation differences in this parent-of-origin effect at *IDDM2* therefore appear to exist but remain to be explained.

Although it is clear that *IDDM2* is the VNTR 5' to the insulin gene (16), current results implicating insulin gene expression with allelic variation are difficult to reconcile (refs. 96 and 97; D. Owerbach and K. Gabbay, personal communication). It is likely that the VNTR does influence insulin expression, but the critical question is whether or not this is relevant to the effect of the VNTR on susceptibility to type 1 diabetes. Perhaps the VNTR influences expression of other genes in the region, such as insulin-like growth factor II, which not only contains an immunological epitope with similarity to the major insulin B-chain epitope (98, 99) but also is a major factor in control of fetal growth (100).

The genetics of type 1 diabetes is obviously complex, and as we venture into the biology of the susceptibility genes the situation will not improve. The challenges for the resolution of the genetics of late-onset type 2 diabetes are, therefore, formidable. It is generally agreed that type 2 diabetes is likely to be etiologically heterogeneous, in contrast to HLA-associated autoimmune type 1 diabetes, which is probably the result of one major pathophysiological pathway (T-cell-mediated β -cell destruction). The parents of most type 2 diabetic sib pairs are deceased. It is possible that in many type 2 diabetic sib pair families the parents might also have been affected, implying homozygosity at

susceptibility loci. Such families provide very little power to detect linkage. Furthermore, the high frequency of the disease (5–50%, depending on the population) means that a significant proportion of affected individuals will be phenocopies—that is, affected for reasons other than possession of the susceptibility alleles under analysis. Nevertheless, the genetics of type 1 diabetes may yet yield clues to the genetics of the type 2 disease. Zimmet and colleagues (101) have shown that a significant proportion (up to 20%) of type 2 diabetics possess antibodies to GAD and have higher frequencies of HLA-DR3 and -DR4. These patients tend to be insulin-treated and may be late-onset type 1 diabetics. Perhaps certain non-MHC type 1 diabetes genes also predispose to insulin-treated type 2 diabetes.

I thank my colleagues, both past and present, and my collaborators, Linda Wicker, Larry Peterson, Steve Bain, Tony Barnett, and Beth Rowe for their hard work and enthusiasm, without whom Table 1 would not have been possible; Neil Risch for valued discussions and direction; and, for financial support, The Wellcome Trust, the Medical Research Council (United Kingdom), the British Diabetic Association, the Juvenile Diabetes Foundation International, and Lilly Industries (U.K.).

- Drell, D. W. & Notkins, A. L. (1987) *Diabetologia* **30**, 132–143.
- Wicker, L. S., Todd, J. A. & Peterson, L. B. (1995) *Annu. Rev. Immunol.* **13**, 179–200.
- Todd, J. A. (1994) *Diabetic Med.* **11**, 6–16.
- Rotter, J. I. (1981) *Am. J. Hum. Genet.* **33**, 835–851.
- Singal, D. P. & Blajchman, M. A. (1973) *Diabetes* **22**, 429–432.
- Nerup, J., Platz, P., Andersson, O. O., Christy, M., Lyngsoe, J., Poulson, J. E., Ryder, L. P., Thomsen, M., Staub, N. K. & Svegaard, A. (1974) *Lancet* **II**, 864–866.
- Cudworth, A. & Woodrow, J. (1974) *Diabetes* **24**, 345–349.
- Bell, G. I., Horita, S. & Karam, J. H. (1984) *Diabetes* **33**, 176–183.
- Davies, J. L., Kawaguchi, Y., Bennett, S. T., Copeman, J. B., Cordell, H. J., Pritchard, L. E., Reed, P. W., Gough, S. C. L., Jenkins, S. C., Palmer, S. M., Balfour, K. M., Rowe, B., Farrall, M., Barnett, A. H., Bain, S. C. & Todd, J. A. (1994) *Nature (London)* **371**, 130–136.
- Hashimoto, L., Habita, C., Beressi, J., Delepine, M., Besse, C., Cambon-Thomsen, A., Deschamps, I., Rotter, J., Djoulah, S., James, M., Froguel, P., Weissenbach, J., Lathrop, G. & Julier, C. (1994) *Nature (London)* **371**, 161–164.
- Hodge, S. E. (1980) *Am. J. Hum. Genet.* **32**, 381–393.
- Risch, N. (1990) *Am. J. Hum. Genet.* **46**, 222–228.
- Olmos, P., A'Hern, R., Heaton, D. A., Millward, B. A., Risley, D., Pyke, D. A. & Leslie, R. D. G. (1988) *Diabetologia* **31**, 747–750.
- Kaprio, J., Tuomilehto, J., Koshenvuo, M., Romanov, K., Reunanen, A., Eriksson, J., Stengård, J. & Kesaniemi, Y. (1992) *Diabetologia* **35**, 1060–1067.
- Todd, J. A., Bell, J. I. & McDevitt, H. O. (1987) *Nature (London)* **329**, 599–604.
- Bennett, S. T., Lucassen, A. M., Gough, S. C. L., Powell, E. E., Undlien, D. E., Pritchard, L. E., Merriman, M. E., Kawaguchi, Y., Dronsfield, M., Pociot, F., Nerup, J., Bouzekri, N., Cambon-Thomsen, A., Rønningen, K. S., Barnett, A. H., Bain, S. C. & Todd, J. A. (1995) *Nat. Genet.* **9**, 284–292.
- Risch, N. (1987) *Am. J. Hum. Genet.* **40**, 1–14.
- Rich, S. S. (1990) *Diabetes* **39**, 1315–1319.
- Dörner, G., Thielke, H., Mohnike, A. & Schneider, H. (1985) *Exp. Clin. Endocrinol.* **85**, 1–6.
- Johansson, C., Samuelsson, U. & Ludvigsson, J. (1994) *Diabetologia* **37**, 91–94.
- Cookson, W. O. C. M., Sharp, P. A., Faux, J. A. & Hopkin, J. M. (1989) *Lancet* **i**, 1292–1295.
- Tomfohrde, J., Silverman, A., Barnes, R., Fernandez-Vina, M. A., Young, M., Lory, D., Morris, L., Wuepper, K. D., Stastny, P., Menter, A. & Bowcock, A. (1994) *Science* **264**, 1141–1145.
- Wang, S., Sun, C., Walczak, C. A., Ziegler, J. S., Kipps, B. R., Goldin, L. R. & Diehl, S. R. (1995) *Nat. Genet.* **10**, 41–46.
- Gyapay, G., Morissette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, G. & Weissenbach, J. (1994) *Nat. Genet.* **7**, 246–339.
- Reed, P. W., Davies, J. L., Copeman, J. B., Bennett, S. T., Palmer, S. M., Pritchard, L. E., Gough, S. C. L., Kawaguchi, Y., Cordell, H. J., Balfour, K. M., Jenkins, S. C., Powell, E. E., Vignal, A. & Todd, J. A. (1994) *Nat. Genet.* **7**, 390–395.
- Pritchard, L. E., Kawaguchi, Y., Reed, P. W., Copeman, J. B., Davies, J. L., Barnett, A. H., Bain, S. C. & Todd, J. A. (1995) *Hum. Mol. Genet.* **4**, 197–202.
- Risch, N. & Zhang, H. (1995) *Science* **268**, 1584–1589.
- Todd, J. A., Aitman, T. J., Cornall, R. J., Ghosh, S., Hall, J. R. S., Hearne, C. M., Knight, A., Love, J. M., McAleer, M. A., Prins, J.-B., Rodrigues, N., Lathrop, G. M., Pressey, A., DeLarato, N., Peterson, L. M. & Wicker, L. S. (1991) *Nature (London)* **351**, 542–547.
- Rise, M. L., Frankel, W. N., Coffin, J. M. & Seyfried, T. N. (1991) *Science* **253**, 669–673.
- Jacob, H. J., Lindpaintner, K., Lincoln, S. E., Kusumi, K., Bunker, R. K., Mao, Y.-P., Ganten, D., Dzau, V. J. & Lander, E. S. (1991) *Cell* **67**, 213–224.
- Hilbert, P., Lindpaintner, K., Beckmann, J. S., Serikawa, T., Soubrier, F., Dubay, C., Cartwright, P., De Gouyon, B., Julier, C., Takahashi, S., Vincent, M., Ganten, D., Georges, M. & Lathrop, G. M. (1991) *Nature (London)* **353**, 521–529.
- Risch, N. (1992) *Science* **255**, 803–804.
- Lander, E. S. & Shork, N. J. (1994) *Science* **265**, 2037–2048.

34. Lander, E. & Botstein, D. (1989) *Genetics* **121**, 185–199.
35. Suarez, B., Hampe, C. & Van Eerdewegh, P. (1994) in *Genetic Approaches to Mental Disorders*, eds. Gershon, E. & Cloninger, C. (Am. Psychiatric Press, Washington, DC), pp. 23–46.
36. Thomson, G. (1994) *Nat. Genet.* **8**, 108–110.
37. Copeman, J. B., Cucca, F., Hearne, C. M., Cornall, R. J., Reed, P. W., Rønningen, K. S., Undlien, D. E., Nistico, L., Buzzetti, R., Tosi, R., Pociot, F., Nerup, J., Cornélis, F., Barnett, A. H., Bain, S. C. & Todd, J. A. (1995) *Nat. Genet.* **9**, 80–85.
38. Wijsman, E. M. (1995) in *Handbook of Psychoneurogenetics*, eds. Blum, K. & Noble, E. P. (CRC, Boca Raton, FL), in press.
39. Risch, N. (1991) *Am. J. Hum. Genet.* **48**, 1058–1064.
40. Hamer, D. H., Hu, S. H., Magnuson, V. L., Hu, N. & Pattatucci, A. M. L. (1993) *Science* **261**, 321–327.
41. Rowe, R. E., Wapelhorst, B., Bell, G. I., Risch, N., Spielman, R. S. & Concannon, P. (1995) *Nat. Genet.* **10**, 240–242.
42. Luo, D.-F., Bui, M. M., Muir, A., MacLaren, N. K., Thomson, G. & She, J.-X. (1995) *Am. J. Hum. Genet.*, in press.
43. Nistico, L., Buzzetti, R., Pozzilli, P., Contu, L., Serrano-Rios, M., Larráz, M. T., Tosi, R. & Giovannini, C. (1995) *Diabetes* **44**, 17A (abstr.).
44. Owerbach, D. & Gabbay, K. H. (1995) *Diabetes* **44**, 132–136.
45. Field, L., Tobias, R. & Magnus, T. (1994) *Nat. Genet.* **8**, 189–194.
46. Cordell, H. J., Kawaguchi, Y., Todd, J. A. & Farrall, M. (1995) *Ann. Hum. Genet.*, in press.
47. Cordell, H. J. & Todd, J. A. (1995) *Trends Genet.*, in press.
48. Cordell, H. J., Todd, J. A., Bennett, S. T., Kawaguchi, Y. & Farrall, M. (1995) *Am. J. Hum. Genet.*, in press.
49. McAleer, M. A., Reifsnnyder, P., Palmer, S. M., Prochazka, M., Love, J. M., Copeman, J. B., Powell, E. E., Rodrigues, N. R., Prins, J.-P., Serreze, D. V., DeLarato, N. H., Wicker, L. S., Peterson, L. B., Schork, N. J., Todd, J. A. & Leiter, E. H. (1995) *Diabetes*, in press.
50. Risch, N., Ghosh, S. & Todd, J. A. (1993) *Am. J. Hum. Genet.* **53**, 702–714.
51. Theofilopoulos, A. N. (1995) *Immunol. Today* **16**, 150–159.
52. Dizier, M.-H., Babron, M.-C. & Clerget-Darpoux, F. (1994) *Am. J. Hum. Genet.* **55**, 1042–1049.
53. Julier, C., Hyer, R. N., Davies, J., Merlin, F., Soularu, P., Briant, L., Cathelineau, G., Deschamps, I., Rotter, J. I., Froguel, P., Boitard, C., Bell, J. I. & Lathrop, G. M. (1991) *Nature (London)* **354**, 155–159.
54. Raffel, L. J., Vadheim, C. M., Roth, M.-P., Klein, R., Moss, S. E. & Rotter, J. I. (1991) *Diabetologia* **34**, 680–683.
55. Field, L. L. (1991) *Bailliere's Clin. Endocrinol. Metab.* **5**, 413–438.
56. van der Auwera, B. J., Heimberg, H., Schrevers, A. F., van Waeyenberge, C., Flament, J. & Schuit, F. C. (1993) *Diabetes* **42**, 851–854.
57. Undlien, D. E., Bennett, S. T., Todd, J. A., Akselsen, H. E., Ikaheimo, I., Reilonen, H., Knip, M., Thorsby, E. & Rønningen, K. S. (1995) *Diabetes* **44**, 620–625.
58. Bain, S. C., Prins, J. B., Hearne, C. M., Rodrigues, N. R., Rowe, B. R., Pritchard, L. E., Ritchie, R. J., Hall, J. R. S., Undlien, D. E., Rønningen, K. S., Dunger, D. B., Barnett, A. H. & Todd, J. A. (1992) *Nat. Genet.* **2**, 212–215.
59. Todd, J. A. (1991) *Diabetic Med.* **8**, 906–910.
60. Ghosh, S., Palmer, S. M., Rodrigues, N. R., Cordell, H. J., Hearne, C. M., Cornall, R., Prins, J.-B., McShane, P., Lathrop, G. M., Peterson, L. B., Wicker, L. S. & Todd, J. A. (1993) *Nat. Genet.* **4**, 404–409.
61. Kruglyak, L. & Lander, E. S. (1995) *Am. J. Hum. Genet.* **56**, 1212–1223.
62. Risch, N., deLeon, D., Ozelius, L., Kramer, P., Almes, L., Singer, B., Fahn, S., Breakfield, X. & Bressman, S. (1995) *Nat. Genet.* **9**, 152–159.
63. Jorde, L. B. (1995) *Am. J. Hum. Genet.* **56**, 11–14.
64. Histo-backa, J., de la Chapelle, A., Kaitila, I., Sistonen, P., Weaver, A. & Lander, E. S. (1992) *Nat. Genet.* **2**, 204–211.
65. Charnley, P. & Concannon, P. (1995) *Immunogenetics*, in press.
66. Peterson, A. C., Di Rienzo, A., Lehesjoki, A.-E., de la Chapelle, A., Slatkin, M. & Freimer, N. B. (1995) *Hum. Mol. Genet.* **4**, 887–894.
67. Todd, J. A. (1990) *Immunol. Today* **11**, 122–129.
68. Todd, J. A., Mijovic, C., Fletcher, J., Jenkins, D., Bradwell, A. R. & Barnett, A. H. (1989) *Nature (London)* **338**, 587–589.
69. Falk, C. & Rubinstein, P. (1987) *Ann. Hum. Genet.* **51**, 227–233.
70. Schaid, D. J. & Sommer, S. S. (1994) *Am. J. Hum. Genet.* **55**, 402–409.
71. Spielman, R., McGinnis, R. & Ewens, W. (1993) *Am. J. Hum. Genet.* **52**, 506–516.
72. Gough, S. C. L., Saker, P. J., Pritchard, L. E., Merriman, T. R., Merriman, M. E., Rowe, B. R., Kumar, S., Aitman, T. J., Barnett, A. H., Turner, R. C., Bain, S. C. & Todd, J. A. (1995) *Hum. Mol. Genet.*, in press.
73. Thomson, G., Robinson, W., Kuhner, M., Joe, S. & Klitz, W. (1989) *Genet. Epidemiol.* **6**, 155–160.
74. Lucassen, A., Julier, C., Beressi, J.-P., Boitard, C., Froguel, P., Lathrop, G. & Bell, J. I. (1993) *Nat. Genet.* **4**, 305–310.
75. Julier, C., Lucassen, A., Villedieu, P., Delepine, M., Levy-Marchal, C., Danzé, P. M., Bianchi, F., Boitard, C., Froguel, P., Bell, J. I. & Lathrop, G. M. (1994) *Am. J. Hum. Genet.* **55**, 1247–1254.
76. Lord, C. J., Bohlander, S. K., Hopes, E. A., Montague, C. T., Hill, N. J., Prins, J.-B., Peterson, L. B., Wicker, L. S., Todd, J. A. & Denny, P. (1995) *Mamm. Genome*, in press.
77. Todd, J. A. & Bain, S. C. (1992) *Diabetes* **41**, 1029–1034.
78. Bingley, P., Bonifacio, E. & Gale, E. (1993) *Diabetes* **42**, 213–220.
79. Aanstoot, H. J., Sigurdsson, E., Jaffe, M., Shi, Y., Christgau, S., Grobbee, D., Bruining, G. J., Molenaar, J. L., Hofman, A. & Baekkeskov, S. (1994) *Diabetologia* **37**, 917–924.
80. Leslie, R. D. G. & Elliot, R. B. (1994) *Diabetes* **43**, 843–850.
81. Vandewalle, C. L., Decraene, T., Schuit, F. C., De Leeuw, I. H., Pipeleers, D. G., Gorus, F. K. & Registry, B. D. (1993) *Diabetologia* **36**, 1155–1162.
82. Wucherpfennig, K. W. & Strominger, J. L. (1995) *J. Exp. Med.* **181**, 1597–1601.
83. Reich, E.-P., von Grafenstein, H., Barlow, A., Swenson, K. E., Williams, K. & Janeway, C. A. (1994) *J. Immunol.* **152**, 2279–2288.
84. Cucca, F., Lampis, R., Frau, F., Macis, M., Angius, E., Masile, P., Chessa, M., Frongia, P., Silvetti, M., Cao, A., De Virgiliis, S. & Congia, M. (1995) *Hum. Immunol.*, in press.
85. Muir, A., Peck, A., Clare-Salzler, M., Song, Y.-H., Cornelius, J., Luchetta, R., Krischer, J. & Maclaren, N. (1995) *J. Clin. Invest.* **95**, 628–634.
86. Eisenbarth, G. S. (1994) *Diabetes Care* **17**, 605–607.
87. Matsushita, S., Takahashi, K., Motoki, M., Komoriya, K., Ikagawa, S. & Nishimura, Y. (1994) *J. Exp. Med.* **180**, 873–883.
88. Khalil, I., d'Auriol, L., Gobet, M., Morin, L., Lepage, V., Deschamps, I., Park, M. S., Degos, L., Galibert, F. & Hors, J. (1990) *J. Clin. Invest.* **85**, 1315–1319.
89. Ikegami, H., Kawaguchi, Y., Ueda, H., Fukuda, M., Takakawa, K., Fujioka, Y., Fujisawa, T., Uchida, K. & Ogiwara, T. (1993) *Biochem. Biophys. Res. Commun.* **192**, 677–682.
90. Yamoto, E., Itoh, N., Matsumoto, E., Lund, T., Yoshino, M., Moriwaki, K., Shiroishi, T. & Hattori, M. (1995) *Diabetes* **44**, 18A (abstr.).
91. Bain, S. C., Rowe, B. R., Barnett, A. H. & Todd, J. A. (1994) *Diabetes* **43**, 1462–1468.
92. Warram, J. H., Krolewski, A. S., Gottlieb, M. S. & Kahn, C. R. (1984) *N. Engl. J. Med.* **311**, 149–152.
93. Anonymous (1995) *Hum. Mol. Genet.* **4**, 499–500.
94. Polychonakos, C., Kukuviitis, A., Giannoukakis, N. & Colle, E. (1995) *Diabetologia* **38**, 715–719.
95. Pugliese, A., Awdeh, Z., Alper, C., Jackson, R. & Eisenbarth, G. (1994) *J. Autoimmun.* **7**, 687–694.
96. Kennedy, G. C., Gorman, M. S. & Rutter, W. J. (1995) *Nat. Genet.* **9**, 293–298.
97. Lucassen, A., Srean, G., Julier, C., Elliot, T., Lathrop, G. M. & Bell, J. I. (1995) *Hum. Mol. Genet.* **4**, 501–506.
98. Geenen, V., Achour, I., Robert, F., Vandersmissen, E., Sodoyez, J. C., Defresne, M. P., Boniver, J., Lefebvre, P. J. & Franchimont, P. (1993) *Thymus* **21**, 115–127.
99. Shell, J. M., Shepherd, S. E., Klimo, G. F. & Paterson, Y. (1992) *J. Exp. Med.* **175**, 542–552.
100. DeChiara, T. M., Efstratiadis, A. & Roberston, E. J. (1990) *Nature (London)* **345**, 78–80.
101. Horton, V., Stratton, I., Zimmet, P., Mackay, I., Bottazzo, F., Lo, D., Manley, S. & Turner, R. (1995) *Diabetes* **44**, 236A (abstr.).