Markers on Distal Chromosome 2q Linked to Insulin-Dependent Diabetes Mellitus

Grant Morahan,* Dexing Huang, Brian D. Tait, Peter G. Colman, Leonard C. Harrison

Insulin-dependent diabetes mellitus (IDDM) is a multigenic autoimmune disease. An IDDM susceptibility gene was mapped to chromosome 2q34. This gene may act early in diabetogenesis, because "preclinical" individuals also showed linkage. Human leukocyte antigen (HLA)–disparate, but not HLA-identical, sibs showed linkage, which was even stronger in families with affected females. The genes encoding insulin-like growth factor–binding proteins 2 and 5 were mapped to a 4–megabase pair interval near this locus. These results indicate the existence of a gene that acts at an early stage in IDDM development, screening for which may identify a specific subset of at-risk individuals.

IDDM is characterized by lymphocytic infiltration of the pancreatic islets and T lymphocyte–mediated destruction of insulin-producing β cells (*1*). In both humans and diabetes-prone nonobese diabetic (NOD) mice, genes that map to the major histocompatibility complex are important in the disease process (*2*). Studies of NOD mice have mapped at least 12 other susceptibility genes to specific chromosomal locations (*3, 4*). In humans, markers near the insulin and insulin-like growth factor (IGF) loci are associated with IDDM (*5*). Recently, a genomewide scan found evidence for up to 18 human chromosomal regions linked to disease loci (*6*); however, the significance of many of these was questioned by a subsequent analysis (*7*) that concluded that genome scans for linkage require maximized lod (logarithm of the odds ratio for linkage) score (MLS) values higher than those demonstrated. Other susceptibility loci have been mapped by either genomewide scans or tests of candidate genes to human chromosomes 2q, 6, 11, and 15 (*8–10*). We adopted a different strategy, restricting our search to potential human homologs of NOD susceptibility genes. Given the map location of these genes, and the known synteny relations of the mouse and human genomes, we identified chromosomal regions most likely to contain the corresponding human homologs. We now describe the search on human chromosome 2q for susceptibility genes corresponding to *Idd5*, which affects infiltration of lymphocytes into islets (*4*).

Idd5 is located on proximal mouse chromosome 1 (*4*). We estimated that the corresponding location in the human genome would be near the microsatellite markers *D2S128-D2S164-D2S126* and therefore typed multiplex families at markers in this interval as well as at flanking markers. Analysis of 98 affected sib pairs is shown in Table 1. The strongest evidence for linkage was apparent with the markers *D2S137-D2S301-D2S164*, for which MLS values (*11*) in excess of 2.5 were obtained. In linkage studies not based on genome scans, MLS values of >2.3 may be taken as significant evidence for linkage (*11, 12*). We designate the diabetes susceptibility gene linked to D2S164 as *IDDM13*.

Human distal chromosome 2q contains many genes that could play a role in IDDM, such as those encoding CTLA4, CD28, the interleukin-8 receptor (IL8R), natural resistance–associated macrophage protein (NRAMP), and insulin-responsive substrate 1 (IRS1) (*13*). To determine the relation of these genes to the microsatellites tested, we typed the multiplex families for inheritance of polymerase chain reaction–based alleles at some of these markers (Table 1). *CTLA4* and *CD28* are within 150 kb of each other, whereas *IL8R* and *NRAMP* are both physically very close to *VIL* (*13*) and the microsatellite marker *D2S1471* (*9*). Multipoint linkage analyses placed *CTLA4-CD28*, *D2S1471* (and hence *VIL*, *IL8R*, and *NRAMP*), and *IRS1* in the order shown in Table 1. MLS values obtained for these loci were less than those obtained in the *D2S137-D2S164* interval, so they were not considered further as candidates for *IDDM13*.

Because human leukocyte antigen (HLA) genes exert the single greatest influence of any IDDM susceptibility genes, the interaction of the HLA complex with *IDDM13* was analyzed by following the inheritance of markers in those sibs who shared identical HLA haplotypes and in those who differed in HLA haplotype (Table 2). Significant linkage was observed in sib pairs who had no more than one HLA haplotype in common. There was no evidence for linkage to markers near *IDDM13* in those sib pairs who shared identical HLA haplotypes, suggesting that in the presence of two copies of the HLA susceptibility haplotype there was no requirement for identity at *IDDM13* for development of IDDM.

The recent development of the MAPMAKER/Sibs program allows multipoint linkage analysis to be performed on data collected from affected sib pairs; it calculates MLS at 1–centimorgan (cM) intervals between genetic markers, taking into account the probability of crossovers at each point (*7*). This program was used to analyze linkage over a large (~30 cM) interval of distal chromosome 2q. In this interval, the peak MLS was apparent in the region corresponding to *IDDM13*, near the markers *D2S137-D2S164* (Fig. 1A). The multipoint analysis

Table 1. Affected sib-pair analysis at polymorphic markers on chromosome 2q. Sib pairs were typed for inheritance of alleles as described (*24*) and the data analyzed with GAS programs (*25*). The total number of sib pairs informative for each marker is shown, as is the number of sib pairs sharing two, one, or zero alleles identical by descent. χ^2 and *P* values were calculated with respect to the expected sharing of two, one, or no alleles in a 1:2:1 ratio. MLS values were calculated for a total of 98 affected sib pairs, as described by Holmans (*11*) with the GAS programs (*25*).

Marker	Informative sib pairs	Shared alleles			χ^2	*P*	MLS
		2	1	0			
CTLA4	44	12	27	5	4.3	0.12	0.607
D2S155	25	9	12	4	2.0	0.36	0.021
D2S334	25	8	10	7	1.1	0.12	0.267
D2S128	82	29	35	18	4.7	0.095	1.562
D2S143	66	30	22	14	15.1	0.00053	1.904
D2S137	87	36	35	16	12.5	0.0019	2.573
D2S301	86	36	34	16	13.1	0.0015	2.580
D2S164	72	33	26	13	16.7	0.00024	3.343
VIL-D2S1471	67	27	29	13	7.4	0.024	1.664
IRS1	42	19	17	6	9.6	0.0083	1.571
D2S126	64	22	28	14	3.0	0.22	1.375

G. Morahan, D. Huang, L. C. Harrison, Walter and Eliza Hall Institute of Medical Research, Royal Melbourne Hospital, Victoria 3050, Australia.
B. D. Tait and P. G. Colman, Royal Melbourne Hospital, Victoria 3050, Australia.
*To whom correspondence should be addressed. E-mail: morahan@wehi.edu.au

was repeated, stratifying the affected sib pairs according to sharing of HLA haplotypes. The extent of linkage to *IDDM13* increased for those sib pairs who were mismatched at HLA, whereas there was no significant linkage in the HLA-identical sib pairs. These results confirm the single-point analyses presented in Tables 1 and 2.

A predictor of IDDM is the presence of circulating islet cell antibodies (ICA) in first-degree relatives of affected probands (*14*). However, not all ICA^+ individuals progress to clinical disease, suggesting that genes that control this phenotype may act early in the disease process. To determine whether the gene we detected on chromosome 2q is important in such islet autoimmunity, we performed multipoint linkage analyses, including those sibs who had either ICA or IDDM. The results, summarized in Fig. 1B, show that the evidence for linkage increases markedly near *IDDM13*, suggesting that this gene controls an early event in islet cell autoimmunity.

One of the differences between human IDDM and the NOD mouse model is the sex difference of the affected individuals: Female NOD mice develop diabetes at a much higher frequency than do males (*1*), whereas the sex balance is more even in humans. (Of the affected individuals in this study, 47% were female.) To determine whether *IDDM13* may be a human counterpart of a NOD gene that contributes to this female bias, we performed multipoint linkage analysis on the HLA-disparate sibs, because this subset showed the greatest linkage in the earlier analyses. Both IDDM and preclinical (ICA^+) individuals were examined. Pedigrees in which IDDM males were in excess (those comprising all males, or two males and one female) were removed, and the remaining families, enriched for affected females, were tested for linkage. A similar manipulation was performed for families that had a majority of affected females. Because it was not possible to determine sex bias in the many families that contained a single affected male and a single affected female, the conservative approach was taken of including these families in both analyses. Linkage to *IDDM13* increased in the set that contained families with excess affected females, and was reduced in the set having a bias to affected males (Fig. 1C). Thus, *IDDM13* appears to confer susceptibility to islet autoimmunity preferentially in females. A second peak was apparent distal to *D2S1471*; whether this represents the influence of an additional susceptibility gene remains to be determined.

The flanking markers *D2S128* and *D2S164* are linked by 4 megabase pairs of cloned DNA in the yeast artificial chromosome (YAC) contig WC-772 (*15*). YACs from this contig were tested for the presence of genes mapped to 2q33-35, or from the corresponding region of the mouse genome (proximal chromosome 1). Of these genes, those encoding IGF-binding protein 2 (IGFBP2) and IGFBP5 (*16*) were found to be present in YACs flanked by the markers *D2S137* and *D2S164* (*17*). Genes that were tested but not detected in these YACs included those encoding human epidermal growth factor receptor 4 (HER4), G protein–coupled receptor 16 (GPCR16), fibronectin (FN1), ubiquinone (UBQ), transition protein 1 (TNP1), Ku-p80, and serpin-thrombin inhibitor (PNI) (*18*). Thus, of the structural genes tested, only *IGFBP2* and *IGFBP5* were detected in the YACs between the relevant flanking markers.

The extent to which an individual gene contributes to the familial clustering of a disease may be calculated from the relative proportion of affected sibs who share zero alleles identical by descent. Such a λs value has been calculated for HLA as 3.1 (6). For both *IDDM2* (near the insulin locus) (6) and *IDDM7* on 2q31-33 (9) λs is 1.3. The overall λs for IDDM is 15. From the data in Table 1, λs for *IDDM13* is 1.6, indicating that the contribution of this gene to the relative risk of developing IDDM is at least as great as that of other non-HLA susceptibility genes identified so far. Two, one, or zero alleles identical by descent were shared by 33, 26, and 13 sib pairs, respectively, at *D2S164*. These numbers may be used to estimate disease allele frequency and genetic heterogeneity (*19*), and this analysis suggests that the *IDDM13* disease allele has a frequency of ~0.1 and that it contributes to 30% of IDDM cases. The *IGFBP2* locus has two known alleles, one of which occurs with a frequency of 0.1 (*16*). Our mapping data showed that, of the structural genes exam-

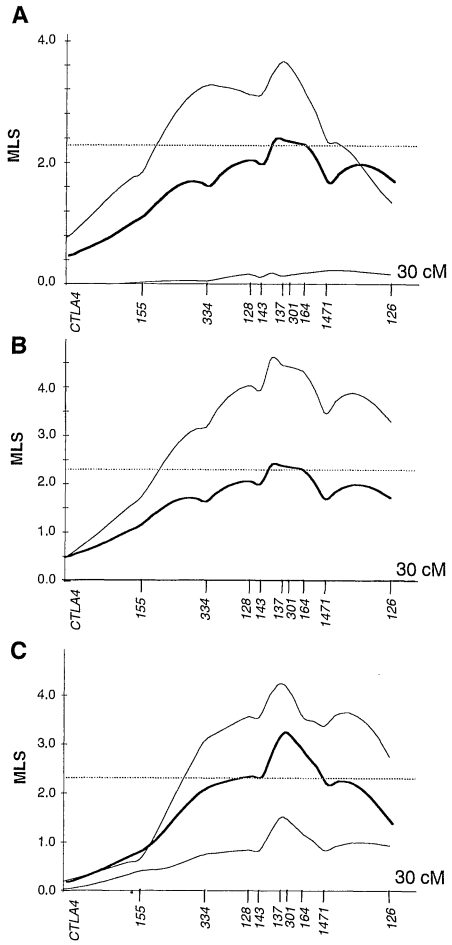

Fig. 1. Multipoint linkage analyses over ~30 cM of human chromosome 2q. (**A**) Variation in MLS according to HLA status. Bold line, total affected sib pairs; lower line, HLA-identical IDDM sib pairs; upper line, HLA-nonidentical IDDM sib pairs. (**B**) Linkage in sib pairs with IDDM or ICA. Lower line (bold), all IDDM sib pairs; upper line, sib pairs with IDDM or ICA. Total number of affected sib pairs was 114. (**C**) Effect of sex on linkage analysis. Bold line, all HLA-nonidentical sib pairs (IDDM and ICA^+); lower and upper lines, all such sib pairs after removal of families in which IDDM females or IDDM males, respectively, were in excess. The MAPMAKER/Sibs program version 0.9 (*7*) was used to calculate MLS at 1-cM intervals between the markers *CTLA4* and *D2S126*. Numbers on the abscissa refer to microsatellite markers, with the *D2S* prefix omitted for clarity. Distances between markers were taken from (*23*) or were calculated from the data. This region includes the structural genes *CTLA4-CD28* and *VIL-NRAMP-IL8R* (near *D2S1471*). The dotted line indicates MLS = 2.3, taken as significant evidence for linkage.

Table 2. Effect of HLA genes on inheritance of *IDDM13*. Results from Table 1 were reanalyzed according to whether the affected sib pairs shared no more than one HLA haplotype or were identical by descent (share both parental haplotypes) at HLA. HLA haplotypes were determined for the Australian families; the HLA genotypes for the BDA families are reported in the BDA Warren catalog.

No. of shared HLA haplotypes	Marker	Informative sib pairs	Shared alleles			χ^2	*P*	MLS
			2	1	0			
1 or 0	*D2S137*	44	22	15	7	14.7	0.0007	2.753
	D2S301	42	21	14	7	14.0	0.0009	2.747
	D2S164	37	19	12	6	13.7	0.0011	2.672
2	*D2S137*	33	10	14	9	0.8	0.66	0.072
	D2S301	31	10	12	9	1.6	0.44	0.171
	D2S164	25	10	9	6	3.2	0.2	0.512

ined, only *IGFBP2* and *IGFBP5* are present in the region of highest MLS values. Furthermore, IDDM subjects have reduced or undetectable amounts of IGFBP2 protein (*20*). Together, these results suggest that *IGFBP2* may be a candidate for the *IDDM13* diabetes susceptibility gene.

Other human IDDM susceptibility genes may reside more proximal (*IDDM7*) (*9*, *10*) and more distal (*9*) on chromosome 2q. *IDDM7* was reported to lie within 1 cM of *D2S152* (*9*) which we estimate is near *CTLA4*. We did not find significant linkage to this region nor did another study (*21*). Therefore, only the distal region (for which MLS = 1.3 was reported) (*9*) was tested further. No significant linkage was found (MLS = 0.77) with *D2S427*, a marker from this region. Studies mapping *Idd5* found linkage to multiple markers on mouse chromosome 1, which suggests that more than one gene may be involved (*4*). Thus, the presence of perhaps two genes (*IDDM7* and *-13*) on chromosome 2q may parallel the results from the NOD mouse.

Stratification of the data set according to biologically meaningful criteria may be a powerful tool in linkage analysis not only of IDDM but also of other complex diseases. Furthermore, inclusion of sibs who share preclinical phenotypes may increase the support for linkage, and also provide the opportunity to assess the stage in the disease process in which susceptibility is conferred. Although we obtained evidence for linkage taking into account all affected sib pairs, stratification according to HLA status was particularly informative. As shown for *IDDM13*, not only did the overall MLS increase, but the peak of highest linkage was narrowed and more clearly defined. Such stratification may incur a penalty (generally taken by subtracting from the lod score the log of the number of tests). In the case of *IDDM13*, the two additional tests (HLA status and sex) provide evidence for linkage even after adjustment with this penalty.

As to why *IDDM13* was detected here but not in a previous study (*9*), it is becoming apparent that there are differences in linkage between different data sets (*21*). Such discrepancies may be attributable to differences in disease gene frequencies or may be due to other factors; from the results presented here, such factors may include the relative proportion of HLA-disparate sib pairs and the sex bias in the particular data set. The number of affected sib pairs required to attain confirmation of linkage is much higher than that required for demonstration of linkage (*22*). Confirmation of linkage of a diabetes susceptibility gene to *D2S301* will require analysis of more multiplex families, but will be assisted by including ICA^+ nondiabetic sibs and by testing those families in which HLA-nonidentical females are preferentially affected.

REFERENCES AND NOTES

1. J.-F. Bach, *Endocrine Rev.* **15**, 516 (1994); M. C. Honeyman and L. C. Harrison, *Semin. Immunopathol.* **14**, 253 (1993).
2. J. A. Todd, *Immunol. Today* **11**, 122 (1990).
3. M. Prochazka, E. H. Leiter, D. V. Serreze, D. L. Coleman, *Science* **237**, 286 (1987); J. A. Todd *et al.*, *Nature* **351**, 542 (1991); B. De Gouyon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1877 (1993); G. Morahan, P. McClive, D. Huang, P. Little, A. Baxter, *ibid.* **91**, 5898 (1994); D. Serreze, M. Prochazka, P. Reifsnyder, M. Bridgett, E. Leiter, *J. Exp. Med.* **180**, 1553 (1994).
4. R. J. Cornall *et al.*, *Nature* **353**, 262 (1991); H. J. Garchon, P. Bedossa, L. Eloy, J.-F. Bach, *ibid.*, p. 260.
5. G. Bell, S. Horita, J. Karam, *Diabetes* **33**, 176 (1984).
6. J. L. Davies *et al.*, *Nature* **371**, 130 (1994).
7. L. Kruglyak and E. Lander, *Am. J. Hum. Genet.* **57**, 439 (1995).
8. L. Field, R. Tobias, T. Magnus, *Nature Genet.* **8**, 189 (1994); L. Hashimoto *et al.*, *Nature* **371**, 161 (1994); D. F. Luo *et al.*, *Am J. Hum. Genet.* **57**, 911 (1995).
9. J. Copeman *et al.*, *Nature Genet.* **9**, 80 (1995).
10. F. Pociot *et al.*, *Autoimmunity* **19**, 169 (1994); D. Owerbach and K. Gabbay, *Diabetes* **44**, 132 (1995).
11. P. Holmans, *Am. J. Hum. Genet.* **52**, 362 (1993).
12. G. Thomson, *Nature Genet.* **8**, 108 (1994); L. Kruglyak and E. Lander, *ibid.* **11**, 241 (1995).
13. N. Buonavista, C. Balzano, P. Pontarotti, D. Le Paslier, P. Golstein, *Genomics* **13**, 856 (1992); J. White *et al.*, *ibid.* **24**, 295 (1994); M. Stoffel *et al.*, *Diabetologia* **36**, 335 (1993).
14. E. Bonifacio *et al.*, *Lancet* **335**, 147 (1990).
15. Whitehead Institute–MIT Center for Genome Research, Human Physical Mapping Project, data release 9 (December 1995).
16. E. Ehrenborg *et al.*, *Biochem. Biophys. Res. Commun.* **176**, 1250 (1991); S. Allander *et al.*, *J. Biol. Chem.* **269**, 10891 (1994).
17. YACs containing the markers *D2S128, D2S137, D2S301*, and *D2S164* were assembled into a contig (WC-772) described in the Whitehead–MIT Center for Genome Research database. To test for genes previously assigned to distal chromosome 2q, or proximal mouse chromosome 1, we screened YACs by Southern (DNA) hybridization on nylon membranes (Amersham) with probes for *IGFBP2* (*16*), *HER4*, *GPCR16*, and *FN1* (*18*). Alternatively, YACs were screened by the polymerase chain reaction for the loci *IGFBP2*, *IGFBP5* (*16*), *UBQ*, *TNP1*, *Ku-p80*, and *PNI* [A. Duncan, W. Chow, B. Robinson, *Cytogenet. Cell Genet.* **60**, 212 (1992); H. Luerssen *et al.*, *Genomics* **8**, 324 (1990); G. Taccioli *et al.*, *Science* **265**, 1442 (1994); R. Carter *et al.*, *Genomics* **27**, 196 (1995)]. We showed that the loci *UBQ*, *TNP1*, *Ku-p80*, *FN1*, and *HER4* were absent from the YACs of the WC-772 contig, as was *PNI* (B. Citron, personal communication).
18. D. Zimonjic, M. Alimandi, T. Miki, N. Popescu, M. H. Kraus, *Oncogene* **10**, 1235 (1995); T. M. Wilkie *et al.*, *Genomics* **18**, 175 (1993); B. Wu *et al.*, *Cytogenet. Cell Genet.* **63**, 29 (1993).
19. G. Morahan, in preparation.
20. W. F. Blum *et al.*, *Growth Regul.* **3**, 100 (1993); S. Giannini *et al.*, *J. Clin. Endocrinol. Metab.* **79**, 1824 (1994).
21. D. F. Luo *et al.*, *Hum. Mol. Genet.*, in press.
22. B. Suarez, C. Hampe, P. van Eerdewegh, in *Genetic Approaches to Mental Disorders*, E. Gershon and C. Cloninger, Eds. (American Psychiatric Press, Washington, DC, 1994), pp. 23–46.
23. J. Weissenbach, *Nature* **359**, 794 (1992).
24. DNA was collected from 25 multiplex families from the Melbourne Prediabetes Family Study. IDDM subjects fulfilled World Health Organization criteria and were dependent on daily insulin injections. Additional samples from 15 families were provided by N. Okech, N. Howard, and J. Knight, and DNA from 22 families was obtained from the British Diabetes Association Warren Repository. Each family member was typed at the indicated chromosome 2 marker. Microsatellite markers were analyzed as described (*23*) with oligonucleotide primers obtained from Research Genetics (Huntsville, AL). Polymorphisms associated with structural genes were detected as described: *CTLA4* (*13*), *VIL-D2S1471* (*9*), and *IRS1* (*13*). *CD28* is within 150 kb of *CTLA4*, whereas *NRAMP* and *IL8R* have been mapped close to *VIL* (*13*). In some instances, alleles at *D2S137* and *D2S301* were determined by haplotyping.
25. A. Young, GAS Manual User Guide, version 1.2 (Oxford University, 1995).

26. We thank C. Steele for obtaining samples; M. C. Honeyman for assembling HLA data; S. Foote for helpful discussions and for providing YACs; J. Knight, N. Okech, and N. Howard for supplying DNA samples and HLA data from multiplex IDDM families; and those investigators who provided probes. Supported by the National Health & Medical Research Council of Australia, the Rebecca Cooper Foundation, the Victorian Health Promotion Foundation, the Ian Potter Foundation, and Hoechst Australia and Diabetes Australia.

7 December 1995; accepted 19 April 1996