

Remapping the Insulin Gene/*IDDM2* Locus in Type 1 Diabetes

Bryan J. Barratt,¹ Felicity Payne,¹ Chris E. Lowe,¹ Robert Hermann,¹ Barry C. Healy,¹ Denise Harold,² Patrick Concannon,³ Neda Gharani,⁴ Mark I. McCarthy,⁴ Mark G. Olavesen,² Rose McCormack,⁵ Cristian Guja,⁶ Constantin Ionescu-Tîrgoviște,⁶ Dag E. Undlien,⁷ Kjersti S. Rønningen,⁸ Kathleen M. Gillespie,⁹ Eva Tuomilehto-Wolf,¹⁰ Jaakko Tuomilehto,^{10,11} Simon T. Bennett,² David G. Clayton,¹ Heather J. Cordell,¹ and John A. Todd¹

Type 1 diabetes susceptibility at the *IDDM2* locus was previously mapped to a variable number tandem repeat (VNTR) 5' of the insulin gene (*INS*). However, the observation of associated markers outside a 4.1-kb interval, previously considered to define the limits of *IDDM2* association, raised the possibility that the VNTR association might result from linkage disequilibrium (LD) with an unknown polymorphism. We therefore identified a total of 177 polymorphisms and obtained genotypes for 75 of these in up to 434 pedigrees. We found that, whereas disease susceptibility did map to within the 4.1-kb region, there were two equally likely

candidates for the causal variant, $-23HphI$ and $+1140A/C$, in addition to the VNTR. Further analyses in 2,960 pedigrees did not support the difference in association between VNTR lineages that had previously enabled the exclusion of these two polymorphisms. Therefore, we were unable to rule out $-23HphI$ and $+1140A/C$ having an etiological effect. Our mapping results using robust regression methods show how precisely a variant for a common disease can be mapped, even within a region of strong LD, and specifically that *IDDM2* maps to one or more of three common variants in a ~ 2 -kb region of chromosome 11p15. *Diabetes* 53: 1884–1889, 2004

From the ¹Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, U.K.; ²Oxagen, Abingdon, Oxon, U.K.; the ³Molecular Genetics Program, Benaroya Research Center and Department of Immunology, University of Washington School of Medicine, Seattle, Washington; the ⁴Imperial College Genetics and Genomics Research Institute, Imperial College Faculty of Medicine, Hammersmith Hospital, London, U.K.; the ⁵Department of Medical Genetics, Queen's University Belfast, Belfast City Hospital, Belfast, U.K.; the ⁶Clinic of Diabetes, Institute of Diabetes, Nutrition and Metabolic Diseases "N. Paulescu" Bucharest, Romania; the ⁷Institute of Medical Genetics, Ullevål University Hospital, University of Oslo, Oslo, Norway; the ⁸Laboratory of Molecular Epidemiology, Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway; ⁹Diabetes and Metabolism, Division of Medicine, University of Bristol, U.K.; the ¹⁰Diabetes and Genetic Epidemiology Unit, National Public Health Institute, University of Helsinki, Helsinki, Finland; and the ¹¹Department of Public Health, University of Helsinki, Helsinki, Finland.

Address correspondence and reprint requests to Bryan J. Barratt, JDRF/WT Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Wellcome Trust/MRC Building, Hills Road, Cambridge, CB2 2XY, U.K. E-mail: bryan.barratt@cimr.cam.ac.uk.

Received for publication 28 January 2004 and accepted in revised form 29 March 2004.

R.H. is currently affiliated with the Department of Virology, University of Turku, Turku, Finland. D.H. is currently affiliated with the Department of Psychological Medicine, University of Wales College of Medicine, Heath Park, Cardiff, U.K. N.G. is currently affiliated with Rutgers University, Nelson Biological Laboratories, Piscataway, New Jersey. M.I.M. is currently affiliated with the Oxford Centre for Diabetes, Endocrinology and Metabolism, Oxford, U.K. M.G.O. is currently affiliated with the Wellcome Trust Centre for Human Genetics, Oxford, U.K. S.T.B. is currently located at Solexa, Chesterford Research Park, Little Chesterford, Essex, U.K. R.M. is currently located at AstraZeneca, Alderley Park, Macclesfield, U.K.

Additional information for this article can be found in an online appendix at <http://diabetes.diabetesjournals.org>. Further information on JDRF/WT Diabetes and Inflammation Laboratory research, including gene annotations and polymorphisms, is available at http://dil-gbrowse.cimr.cam.ac.uk/cgi-bin/DIL_GenomeView.cgi.

LD, linkage disequilibrium; PH, protective haplotype; SNP, single nucleotide polymorphism; VNTR, variable number tandem repeat; VPH, very protective haplotype.

© 2004 by the American Diabetes Association.

The insulin/IGF-2 variable number tandem repeat (*INS-IGF2* VNTR) is situated ~ 600 bp 5' of the *INS* transcription start site and is comprised of 14–15 bp tandem repeating sequences with the consensus ACAGGGGTSYGGGG (1). In Caucasian populations, this VNTR has two main size classes. The shorter class I alleles have between 26 and 63 repeat units, and the class III alleles have between 141 and 209 repeat units. Intermediately sized class II alleles are rare in white European populations (1,2). Early studies (1,3,4) of *INS* association with type 1 diabetes reported that the class I VNTR homozygous genotype was present at a higher frequency in case subjects compared with control subjects. Subsequent studies of flanking polymorphisms reported that this association was restricted to markers within a 19-kb interval (5) and, later, to a 4.1-kb interval spanning *INS* (6). Within this interval there were 10 candidate causal common variants, the VNTR and nine single nucleotide polymorphism (SNPs) (6). Of these 10 candidates, the strongest disease associations were with the VNTR and the $-23HphI$ and $+1140A/C$ SNPs (relative risk [RR] for homozygous genotype = 4.5, for all three polymorphisms). A number of additional studies (7–10) excluded the SNPs in the 4.1-kb region and led to the VNTR being proposed as the etiological variant (11). This proposal was principally based on the differential association of class III VNTR-bearing haplotypes, the protective haplotype (PH) and very protective haplotype (VPH) (9), a phenomenon most easily explained by sequence variation

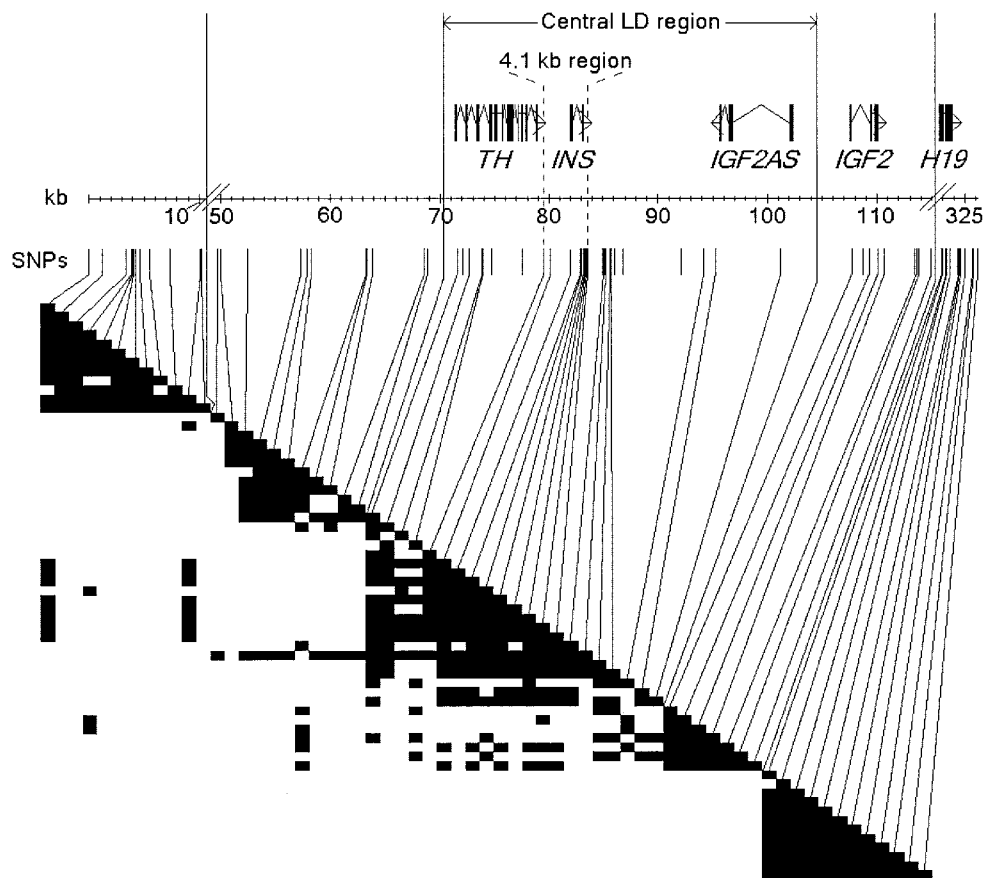


FIG. 1. Intermarker LD (pairwise D') estimates in the *IDDM2/INS* region, revealing five major LD blocks ($D' > 0.7$, shaded black). The average D' between markers within blocks is 0.84, whereas the average D' between markers in different blocks is 0.21.

within the VNTR rather than *cis* effects of SNP alleles on these haplotypes. In addition, expression studies (12,13) suggested that the VNTR could directly alter the transcription of *INS* and *IGF2*. However, the fine mapping of *IDDM2* susceptibility to the VNTR was heavily reliant on the premise that the association was restricted to the 4.1-kb interval. Thus, when Doria et al. (14) reported disease association with a SNP in the 5' region of the tyrosine hydroxylase gene (*TH*) and reinterpreted the findings of Lucassen et al. (6), it became possible that the *IDDM2* causal variant mapped outside the 4.1-kb region. Here we have performed a more detailed genetic analysis of the *IDDM2* region to address the uncertainties surrounding the location of the causal variant.

We identified novel polymorphisms in the region by extensive sequencing and resequencing (online appendix Tables 1 and 2 [available from <http://diabetes.diabetesjournals.org>]) and, together with additional polymorphisms from published literature and public databases, compiled a much more extensive and dense variant map than had been previously used (in total, 177 polymorphisms) (online appendix Table 3). By genotyping polymorphisms in up to 434 affected sibpair families from the U.K. and combining data with those analyzed by Bennett et al. (9), we assembled genotypes for a total of 75 markers across the region. Analysis of parental chromosomes revealed five main regions of strong LD in the *TH/INS/IGF2* region and at a neighboring gene, *H19* (Fig. 1). In the central region of 28.5 kb, 21-marker diversity is limited to 16 haplotypes with a frequency $>1\%$, accounting for 84% of all chromosomes. Within this central region, the ~ 4.1 -kb region (6) is

identifiable, with diversity limited to four haplotypes with frequency $>1\%$, accounting for 98% of all chromosomes (Fig. 2).

Logistic regression analysis, conditioned on parental genotype (15), of 74 markers (VNTR subclasses excluded) indicated that the strongest association with disease was at $-23HphI$ ($\chi^2 = 29.58$, 1 df [degree of freedom], $P = 5.4 \times 10^{-8}$) (Fig. 3A and online appendix Table 4), an often-used surrogate marker for the VNTR class I and class III alleles. These two polymorphisms reside in the central ~ 4.1 kb within 600 bp of each other and show near complete concordance (99.77%) in European populations (9,16). Other than $-23HphI$, a number of markers within the same LD block also had highly significant disease associations. However, it is not apparent from this single-locus analysis which of the associated variants is likely to be the main influence on disease risk, which are associated due to LD with the primary causal variant (hitchhiking) or which may be associated with disease independently of the primary causal variant. To address this question, we applied the stepwise conditional logistic regression approach (15). We included the effect of $-23HphI$ in the regression model and tested the contribution to the model of each locus in turn (Fig. 3B and online appendix Table 5). Of all markers ($n = 73$) only two (*TH/199.1* and *TH/227.3*) achieved a significance of $P < 0.05$ ($P = 0.034$ and 0.014 , respectively). Once the number of tests performed ($n = 73$) is taken into consideration, these P values suggest that LD with $-23HphI$ or the VNTR is sufficient to explain the association of the other markers tested.

-2733A/C	-2221MspI	-23HphI	+805DraIII	C5347T(novel)	+1127PstI	+1140A/C	+1355C/T	+1404Fnu4HI	+1428FokI	Frequency %
A	C	A	C	C	C	C	C	G	G	76.0*
C	T	T	G	C	T	A	T	T	A	14.8†
C	C	T	C	T	C	A	C	G	G	4.4‡
C	C	A	C	C	C	C	C	G	G	3.3*

FIG. 2. Limited SNP haplotype diversity across the insulin gene. *Susceptible haplotypes (bearing class I VNTR). †PH and ‡VPH, based on the nomenclature described in Bennett et al. (9), both of which bear class III VNTR.

Conversely, we then tested whether LD between any of the other markers and -23HphI were capable of explaining the association of -23HphI with disease, i.e., that -23HphI was possibly associated with disease due to LD

with an alternative primary variant. We included the effect of each marker ($n = 73$) in turn in the regression model and tested the effect of -23HphI. -23HphI added significantly to the model in all cases, with the exception of

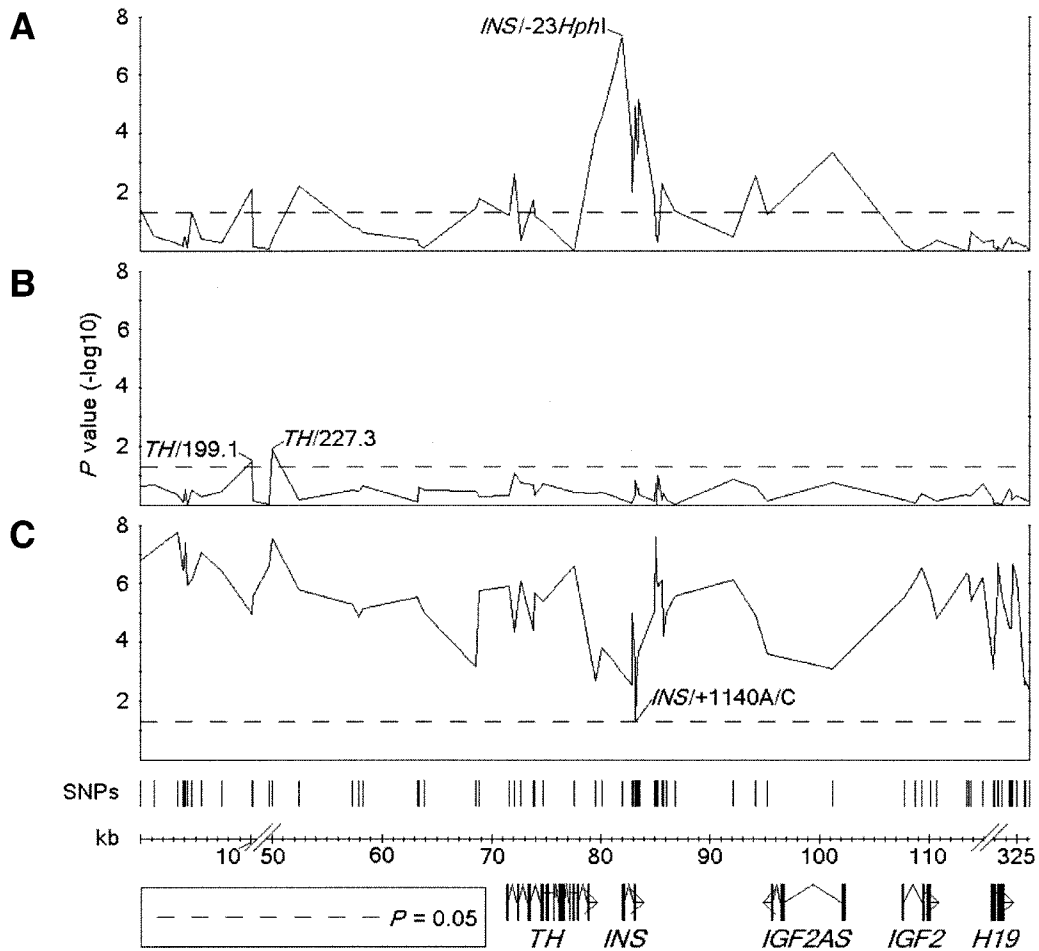


FIG. 3. Conditional logistic regression analysis at *IDDM2*. A: Single-point association with type 1 diabetes. B: The effect of adding markers to a model that already accounts for the association of -23HphI. Significance may indicate that the test marker has a modifying effect on, or an effect independent of, -23HphI-conferred risk. Nonsignificance implies that the single-point association of the test marker may be accounted for by LD with -23HphI. C: The effect of adding the -23HphI marker into models that already account for each of the other markers in turn. This indicates whether LD with any of the other markers may explain the association of -23HphI.

TABLE 1
Parental frequencies of *INS* haplotypes in data included in the PH/VPH association analysis

Dataset	No. of families*	Class I bearing	PH	VPH
U.K. Warren	396	1,279 (80.7)	235 (14.8)	70 (4.4)
U.K. simplex†	458	1,458 (79.6)	270 (14.7)	104 (5.7)
Finland	1,053	3,616 (85.9)	368 (8.7)	228 (5.4)
Norway	178	564 (79.2)	98 (13.8)	50 (7.0)
Romania	287	984 (85.7)	119 (10.4)	45 (3.9)
U.S.	201	686 (85.3)	78 (9.7)	40 (5.0)
Bart's Oxford study	387	1,235 (80.0)	234 (15.1)	79 (5.1)
Combined	2,960	9,822 (83.0)	1,402 (11.8)	616 (5.2)

Data are *n* (%). *In which phase was determinable; †Belfast, Bristol, Yorkshire, and Southwest U.K. combined.

+1140A/C ($P = 0.051$) (Fig. 3C and online appendix Table 5), indicating that +1140A/C could perhaps be just as effective in explaining the association of the region. This result was not unexpected because it has been previously reported (9) that +1140A/C is in very strong LD with -23HphI. We therefore tested +1140A/C as if it were the primary causal variant. By including +1140A/C in the model and testing the contribution of each other marker in turn, four markers (*TH/227.3*, allele Z-16 of the *TH* microsatellite, *TH/139.1*, and *IGF2/A20531C*) achieved a significance of $P < 0.05$ ($P = 0.024, 0.037, 0.0024$, and 0.039 , respectively) (online appendix Table 6). Although +1140A/C does not, on its own, explain the association of the other markers quite as well as -23HphI, even the contribution of *TH/139.1* to the model would not withstand correction for the 73 tests performed. As such, on the basis of the above analyses, +1140A/C must also be considered as a candidate causal variant, along with -23HphI and the VNTR.

Previously, both -23HphI and +1140A/C SNPs have been excluded as causal variants (9), based on the observation that these SNPs had the same alleles on both the PH and VPH, and the association of these haplotypes with type 1 diabetes had been shown to differ ($P = 0.048$). However, because this P value was marginal and no correction for the relatedness of affected individuals was made, we sought to replicate this finding in a larger dataset. Further families from Finland, Romania, Norway, the U.S., Bart's Oxford study, and additional simplex families from the U.K. were genotyped at -23HphI and +1428FokI, since the haplotypes of these two SNPs distinguish between the susceptible haplotype (A allele at -23HphI), PH (T allele at -23HphI and A allele at +1428FokI), and VPH (T allele at -23HphI and G allele at +1428FokI). In total, 3,722 fully genotyped affected offspring (distributed among 3,056 pedigrees) also had both parents fully genotyped at both loci.

To investigate the PH and VPH association, case and pseudo-control sets were generated in which the phase of the transmitted -23HphI and +1428FokI alleles was also determined. Of the 3,722 cases in 3,056 pedigrees, phase was unambiguously determinable for 3,585 cases (in 2,960 pedigrees). The resulting haplotypes were then assigned as either PH or VPH, as previously defined (9), or as class I VNTR-bearing (principally the susceptible haplotype). Parental haplotype frequencies and the numbers of families for which phase was determinable in each population are shown in Table 1. Haplotype risks for the PH and VPH, relative to the class I-bearing haplotypes, were estimated

by conditional logistic regression and found to be nearly identical. For the PH, the haplotype risk was 0.46 ($P = 2.0 \times 10^{-40}$, 95% CI 0.41–0.52), whereas for the VPH, the haplotype risk was 0.43 ($P = 3.0 \times 10^{-22}$, 95% CI 0.37–0.51). There was no evidence for population heterogeneity for these haplotype risks using either a seven-population categorization (populations as in Table 2, 12 df, $P = 0.36$) or a five-population categorization (U.K. Warren, U.K. simplex, and Bart's Oxford study grouped, 8 df, $P = 0.44$). The haplotype risks in each population individually are shown in Table 2. Further tests were constructed to estimate the risks of the six possible haplotype combinations, but no significant evidence for a difference in association between PH homozygotes and VPH homozygotes or between class I/PH and class I/VPH heterozygotes was observed (data not shown). Given all of the above results, it can be concluded that no significant evidence for a difference in association between the PH and VPH is found in these data.

We can conclude that type 1 diabetes susceptibility in this region does map to one (or perhaps a combination) of three common polymorphisms in a ~2-kb region at *INS*, but cannot be mapped precisely to the VNTR. In the absence of a VPH effect, it is unlikely that the resolution of VNTR, -23HphI, and +1140A/C can be achieved by association studies in European-derived populations owing to the strength of LD between these markers. A study (2) of *INS* haplotypes indicates that the three polymorphisms are also in very strong LD in subjects from diverse populations. However, if class II VNTR risk were to differ from class III VNTR risk, resolution of these three polymorphisms may be achievable in African populations if sufficient sample sizes were available, perhaps using SNPs on

TABLE 2
PH/VPH association analysis by population and overall

Dataset	PH	VPH
U.K. Warren*	0.58 (0.46–0.74)	0.50 (0.33–0.76)
U.K. simplex*†	0.39 (0.30–0.52)	0.45 (0.30–0.69)
Finland	0.51 (0.40–0.64)	0.39 (0.29–0.53)
Norway	0.33 (0.21–0.52)	0.48 (0.26–0.88)
Romania	0.33 (0.22–0.50)	0.34 (0.17–0.65)
U.S.	0.46 (0.33–0.65)	0.39 (0.22–0.70)
Bart's Oxford study*	0.42 (0.31–0.59)	0.50 (0.30–0.82)
Combined	0.46 (0.41–0.52)	0.43 (0.37–0.51)

Data are risk ratios (95% CI). *Risk ratios for all U.K. datasets grouped together (U.K. Warren, U.K. simplex, and Bart's Oxford study) were 0.48 (0.41–0.57) for PH and 0.48 (0.38–0.62) for VPH; †Belfast, Bristol, Yorkshire, and Southwest U.K. combined.

class II-bearing haplotypes as surrogate markers (2). Despite the current lack of genetic evidence that would enable the resolution of the three remaining candidates for *IDDM2*, it is noted that the VNTR remains the best candidate. Functionally it contains multiple binding sites for transcription factors such as Pur-1 (13,17), and the type 1 diabetes susceptibility at *INS* has been proposed to arise from different levels of thymic expression (18–20), whereas there is no obvious functional role for either of the candidate SNPs. We also note that our previous conclusion, which was based on the existence of a difference in risk between the PH and VPH, that the *IDDM2* locus was a dominant protective trait (11) is no longer valid.

More generally, in the context of fine mapping susceptibility loci in common multifactorial diseases, our results confirm, as we found for the CTLA-4 gene in Graves' disease (21), that, despite strong LD, small discrete regions can be pinpointed, providing that sufficient sample sizes are used. This level of mapping resolution greatly reduces the number of polymorphisms that have to be analyzed for functional effects.

RESEARCH DESIGN AND METHODS

All families were Caucasian of European descent, with two parents and at least one affected child. The families were comprised of Diabetes U.K. Warren multiplex from the U.K., U.S. multiplex from Human Biological Data Interchange, Yorkshire simplex from the U.K. and Southwest U.K. simplex, Belfast multiplex/simplex from the U.K., Finnish multiplex/simplex (for references see 21), Norwegian simplex (22), Romanian simplex (23), and Bart's Oxford simplex/multiplex from the U.K. (24). All DNA samples were collected with informed consent.

Sequence data. Sequence data for PCR primer design were obtained from GenBank (accession nos. L15440, M32053, AC004556, AF087017, M23597, AC006408, and AH010044) and from shotgun sequencing performed by Incyte Genomics of RPC11 BAC clone number "ba"94F12 (supplied by the Wellcome Trust Sanger Institute, Cambridge, U.K.). Contigs generated from the shotgun sequencing were joined by the generation and sequencing of PCR products spanning the gaps between contigs, and unresolved contig positions and orientations were then resolved by comparison with sequence data obtained using the Celera Discovery System. Data for the region covered by the novel sequence has since been submitted independently by researchers from the Whitehead Institute/MIT Center for Genome Research as accession no. AC132217.

SNP identification. SNPs were identified either by denaturing high-performance liquid chromatography as previously described (25) or by sequencing of 32 individuals using BigDye Terminator chemistry and ABI 3700 instrumentation (Applied Biosystems, Foster City, CA). The primers used for SNP identification are shown in online appendix Tables 1 and 2. Sequence data were processed in the Staden package software (www.mrc-lmb.cam.ac.uk/pubseq). One hundred sixty-eight polymorphisms were identified, 24 of which had previously been reported in the literature. Combined with seven other SNPs identified from the literature, the VNTR minisatellite and *TH* microsatellite, a total of 177 polymorphisms were mapped to the *TH-INS-IGF2-H19* region (online appendix Table 3).

Genotyping. Genotype data were obtained from Invader assays (Third Wave Technologies, Madison, WI), TaqMan chemistry (Applied Biosystems), Pyrosequencing chemistry (Pyrosequencing, Uppsala, Sweden), or PCR restriction fragment-length polymorphism digests. The estimated error rate for these technologies was ~1% (26,27). Additional *TH* microsatellite genotypes were generated from fluorescently labeled PCR products and sized using ABI 3700 instrumentation and software. Data were combined with those that were previously published (9,16).

Data analysis. Data were examined for misinheritances using PedCheck (Jeff O'Connell, 1997, 1999, University of Pittsburgh, Pittsburgh, PA) and recombination using GAS (Genetic Analysis System [<http://users.ox.ac.uk/~ayoung/gas.html>]) and potential genotyping errors removed. Intermarker pairwise estimates of LD (D') were estimated with Stata 7 (Stata, San Mateo, CA) using `pwld`, which is available as part of the Genassoc package available at www.mrc-bsu.cam.ac.uk/pub/methodology/genetics. SNPs with a parental allele frequency <5% and multiallelic markers with rare alleles (*TH* microsat-

ellite and *INS-VNTR*) were excluded from D' estimates to prevent inaccurate estimates due to sparse data. LD blocks were assigned by visual inspection of the matrix of pairwise D' estimates. Pseudo-controls were generated and conditional logistic regression analyses performed in Stata 7 using routines from the Genassoc package, according to the method described by Cordell and Clayton (15). For each affected subject, the corresponding pseudo-controls are assigned all of the other possible genotypes of offspring that could have been generated from the parents. In the subsequent conditional logistic regression analysis, case subjects and pseudo-controls are matched according to the parent-case set that they were generated from. In analyses of SNP haplotypes representing the PH and VPH, phase was determined using the phase option of the `pseudoc` routine of the Genassoc package. In all analyses, association was evaluated by fitting a conditional logistic regression model (as used in regular matched case-control studies), in which the RR of disease is given by $\beta_1 x_1 \dots \beta_j x_j$, where x_i is an indicator variable for the genotypes (or combinations of phased haplotypes) at locus i of j loci included in the test and $\beta_1 \dots \beta_j$ are the parameters to be maximized. The fitted model is compared with the appropriate null hypothesis model in which all $\beta_i = 0$, correcting for nonindependence of sibs by use of robust variance estimation. These analyses were carried out using the `rclogit` command (from the Genassoc package) within Stata 7. In the case of the *TH* microsatellite, the values were plotted for the most associated allele (Z-16/106) were uncorrected for the number of alleles tested. P values for association, exon, and SNP positions were plotted for figures using the Generic Genome Browser available at www.gmod.org.

ACKNOWLEDGMENTS

We thank the Wellcome Trust, the Juvenile Diabetes Research Foundation International, Novo Nordisk, the Novo Nordisk Foundation, the Academy of Finland, the Sigrid Juselius Foundation, and Diabetes U.K. for financial support. B.J.B. was funded by the Medical Research Council, U.K., and Oxagen, U.K.

We thank the members of the DNA resource team and Neil Walker of the JDRF/WT Diabetes and Inflammation Laboratory for sample and data services. We thank Diabetes U.K., the Human Biological Data Interchange, and the Norwegian Study Group for Childhood Diabetes for collection of the U.K., U.S., and Norwegian families, respectively.

REFERENCES

- Bell GI, Horita S, Karam JH: A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* 33:176–183, 1984
- Stead JD, Hurler ME, Jeffreys AJ: Global haplotype diversity in the human insulin gene region. *Genome Res* 13:2101–2111, 2003
- Hitman GA, Tam AC, Winter RM, Drummond V, Williams LG, Jowett NI, Bottazzo GF, Galton DJ: Type 1 (insulin-dependent) diabetes and a highly variable locus close to the insulin gene on chromosome 11. *Diabetologia* 28:218–222, 1985
- Thomson G, Robinson WP, Kuhner MK, Joe S, Klitz W: HLA and insulin gene associations with IDDM. *Genet Epidemiol* 6:155–160, 1989
- Julier C, Hyer RN, Davies J, Merlin F, Soularue P, Briant L, Cathelineau G, Deschamps I, Rotter JI, Froguel P, Boitard C, Bell JI, Lathrop GM: Insulin-IGF2 region on chromosome 11p encodes a gene implicated in HLA-DR4-dependent diabetes susceptibility. *Nature* 354:155–159, 1991
- Lucassen AM, Julier C, Beressi JP, Boitard C, Froguel P, Lathrop M, Bell JI: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment of DNA spanning the insulin gene and associated VNTR. *Nat Genet* 4:305–310, 1993
- Julier C, Lucassen A, Villedieu P, Delepine M, Levy-Marchal C, Danze PM, Bianchi F, Boitard C, Froguel P, Bell J, Lathrop GM: Multiple DNA variant association analysis: application to the insulin gene region in type I diabetes. *Am J Hum Genet* 55:1247–1254, 1994
- Owerbach D, Gabbay KH: Localization of a type I diabetes susceptibility locus to the variable tandem repeat region flanking the insulin gene. *Diabetes* 42:1708–1714, 1993
- Bennett ST, Lucassen AM, Gough SC, Powell EE, Undlien DE, Pritchard LE, Merriman ME, Kawaguchi Y, Dronsfield MJ, Pociot F, Nerup J, Bouzekri N, Cambon-Thomsen A, Ronning KS, Barnett AH, Bain SC, Todd JA: Susceptibility to human type 1 diabetes at *IDDM2* is determined

- by tandem repeat variation at the insulin gene minisatellite locus. *Nat Genet* 9:284–292, 1995
10. Undlien DE, Bennett ST, Todd JA, Akselsen HE, Ikaheimo I, Reijonen H, Knip M, Thorsby E, Ronningen KS: Insulin gene region–encoded susceptibility to IDDM maps upstream of the insulin gene. *Diabetes* 44:620–625, 1995
 11. Bennett ST, Todd JA: Human type 1 diabetes and the insulin gene: principles of mapping polygenes. *Annu Rev Genet* 30:343–370, 1996
 12. Paquette J, Giannoukakis N, Polychronakos C, Vafiadis P, Deal C: The INS 5' variable number of tandem repeats is associated with IGF2 expression in humans. *J Biol Chem* 273:14158–14164, 1998
 13. Kennedy GC, German MS, Rutter WJ: The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. *Nat Genet* 9:293–298, 1995
 14. Doria A, Lee J, Warram JH, Krolewski AS: Diabetes susceptibility at IDDM2 cannot be positively mapped to the VNTR locus of the insulin gene. *Diabetologia* 39:594–599, 1996
 15. Cordell HJ, Clayton DG: A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet* 70:124–141, 2002
 16. Bennett ST, Wilson AJ, Esposito L, Bouzekri N, Undlien DE, Cucca F, Nistico L, Buzzetti R, Bosi E, Pociot F, Nerup J, Cambon-Thomsen A, Pugliese A, Shield JP, McKinney PA, Bain SC, Polychronakos C, Todd JA: Insulin VNTR allele-specific effect in type 1 diabetes depends on identity of untransmitted paternal allele. *Nat Genet* 17:350–352, 1997
 17. Lew A, Rutter WJ, Kennedy GC: Unusual DNA structure of the diabetes susceptibility locus IDDM2 and its effect on transcription by the insulin promoter factor Pur-1/MAZ. *Proc Natl Acad Sci U S A* 97:12508–12512, 2000
 18. Vafiadis P, Bennett ST, Todd JA, Nadeau J, Grabs R, Goodyer CG, Wickramasinghe S, Colle E, Polychronakos C: Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. *Nat Genet* 15:289–292, 1997
 19. Liston A, Lesage S, Wilson J, Peltonen L, Goodnow CC: Aire regulates negative selection of organ-specific T cells. *Nat Immunol* 4:350–354, 2003
 20. Pugliese A, Zeller M, Fernandez A Jr, Zalcberg LJ, Bartlett RJ, Ricordi C, Pietropaolo M, Eisenbarth GS, Bennett ST, Patel DD: The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. *Nat Genet* 15:293–297, 1997
 21. Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G, Rainbow DB, Hunter KM, Smith AN, Di Genova G, Herr MH, Dahlman I, Payne F, Smyth D, Lowe C, Twells RC, Howlett S, Healy B, Nutland S, Rance HE, Everett V, Smink LJ, Lam AC, Cordell HJ, Walker NM, Bordin C, Hulme J, Motzo C, Cucca F, Hess JF, Metzker ML, Rogers J, Gregory S, Allahabadi A, Nithiyananthan R, Tuomilehto-Wolf E, Tuomilehto J, Bingley P, Gillespie KM, Undlien DE, Ronningen KS, Guja C, Ionescu-Tirgoviste C, Savage DA, Maxwell AP, Carson DJ, Patterson CC, Franklyn JA, Clayton DG, Peterson LB, Wicker LS, Todd JA, Gough SC: Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 423:506–511, 2003
 22. Undlien DE, Akselsen HE, Joner G, Dahl-Jorgensen K, Aagaenaes O, Sovik O, Thorsby E, Ronningen KS: No difference in the parental origin of susceptibility HLA class II haplotypes among Norwegian patients with insulin-dependent diabetes mellitus. *Am J Hum Genet* 57:1511–1514, 1995
 23. Ionescu-Tirgoviste C, Guja C, Herr M, Cucca E, Welsh K, Bunce M, Marshall S, Todd JA: Low frequency of HLA DRB1*03-DQB1*02 and DQB1*0302 haplotypes in Romania is consistent with the country's low incidence of type I diabetes. *Diabetologia* 44 (Suppl. 3):B60–B66, 2001
 24. Gardner SG, Bingley PJ, Sawtell PA, Weeks S, Gale EA: Rising incidence of insulin dependent diabetes in children aged under 5 years in the Oxford region: time trend analysis: the Bart's-Oxford Study Group. *BMJ* 315:713–717, 1997
 25. Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA: Haplotype tagging for the identification of common disease genes. *Nat Genet* 29:233–237, 2001
 26. Barratt BJ, Payne F, Rance HE, Nutland S, Todd JA, Clayton DG: Identification of the sources of error in allele frequency estimations from pooled DNA indicates an optimal experimental design. *Ann Intern Med* 66:393–405, 2002
 27. Mein CA, Barratt BJ, Dunn MG, Siegmund T, Smith AN, Esposito L, Nutland S, Stevens HE, Wilson AJ, Phillips MS, Jarvis N, Law S, de Arruda M, Todd JA: Evaluation of single nucleotide polymorphism typing with invader on PCR amplicons and its automation. *Genome Res* 10:330–343, 2000