

Absolute Risk of Childhood-Onset Type 1 Diabetes Defined by Human Leukocyte Antigen Class II Genotype: A Population-Based Study in the United Kingdom

A. PAUL LAMBERT, KATHLEEN M. GILLESPIE, GLENYS THOMSON, HEATHER J. CORDELL, JOHN A. TODD, EDWIN A. M. GALE, AND POLLY J. BINGLEY

Division of Medicine (A.P.L., K.M.G., E.A.M.G., P.J.B.), University of Bristol, Bristol BS10 5NB, United Kingdom; Department of Integrative Biology (G.T.), University of California, Berkeley, California 94720-3140; and Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory (H.J.C., J.A.T.), Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 2XY, United Kingdom

The autoimmune disease process leading to childhood-onset type 1 diabetes appears to start in infancy, and decisions on treatment to prevent initiation of autoimmunity will need to be based on genetic susceptibility alone. We set out to quantify the absolute risk associated with human leukocyte antigen (HLA) *DRB1-DQA1-DQB1* class II genotypes and to develop strategies for recruitment into primary prevention trials. HLA class II haplotype- and genotype-specific risks were derived from 753 United Kingdom families from the Bart's-Oxford population-based study of type 1 diabetes and combined with incidence data from the region to calculate the absolute risk of development of diabetes. A hierarchy of susceptibility was established for both HLA class II haplo-

types and genotypes, and the sensitivity and specificity of each genotype was established relative to age at disease onset. Highest risk was conferred by the genotype *DRB1*03-DQA1*0501-DQB1*0201/DRB1*0401-DQA1*0301-DQB1*0302* (5% absolute risk of diabetes by age 15 yr), although sensitivity was only 22.6%. Combining the six highest risk genotypes conferred similar risk but increased sensitivity to 36.6% and was most sensitive for diagnosis of diabetes before age 5 yr (48.4%), whereas inclusion of 11 genotypes achieved the same sensitivity for diagnosis for ages 10–14 yr. Analysis of genotype-specific risk should form the basis for design of future primary prevention trials in the general population. (*J Clin Endocrinol Metab* 89: 4037–4043, 2004)

THE AUTOIMMUNE DISEASE process that culminates in childhood-onset type 1 diabetes appears to start in infancy (1). Therefore, primary prevention should be attempted as soon after birth as possible, before immune responses directed against pancreatic β -cells are detectable. This implies that treatment decisions will be based on genetic susceptibility alone. Family history of type 1 diabetes is a useful starting point for identification of children at high genetic risk but identifies only 10–15% of future cases of childhood diabetes (2, 3). We need to develop robust means of assessing absolute genetic risk of diabetes including children with no family history of type 1 diabetes—means that will discriminate accurately between differing levels of risk and will permit appropriate recruitment into future intervention trials.

Type 1 diabetes is a multigenic disorder in which human leukocyte antigen (HLA) class II alleles make the greatest contribution to susceptibility. The effects of *HLA-DRB1* and *-DQB1* alleles are modified by alleles at other loci making up the *DRB1-DQA1-DQB1* haplotype (4–6), and the risk conferred by a class II genotype may differ from that predicted from the two haplotypes expressed, for example in the syn-

ergistic effect of *DRB1*03* and *04* in the highest risk heterozygous genotypes. Therefore, assessment of class II-associated risk is most appropriately based on genotype, and the most suitable measure for clinical studies is absolute risk rather than relative risk. We set out to quantify the absolute risk associated with three-locus (*DRB1-DQA1-DQB1*) HLA class II genotypes in the Oxford region of the United Kingdom and to develop strategies for recruitment into primary prevention trials.

The absolute risk associated with each genotype can be derived from knowledge of the relative risk related to the background risk of disease. We used a population-based study with high levels of case ascertainment, which allowed risk to be defined for genotypes conferring low and high susceptibility, permitted comparison of the sensitivity of these genotypes, and provided accurate data on diabetes incidence.

Patients and Methods

Patients

Families were identified from the Bart's-Oxford (BOX) study of childhood diabetes. This is a prospective, population-based family study that since 1985 has recruited more than 95% of the families of children who have developed type 1 diabetes before age 21 yr in the former Oxford Health Authority Region, United Kingdom (7, 8). By March 2002, 1746 families had been recruited into the study with 89% under regular follow-up. The study population is 95% white European, and the remainder originate mainly from the Indian subcontinent (data from Office of Population Censuses and Surveys for 1991). The classification of type 1 diabetes was based on assignment by the referring clinician and

Abbreviations: AFBAC, Affected family-based control; AUC, area under the curve; BOX, Bart's-Oxford; HLA, human leukocyte antigen; PC, ratio of haplotype and genotype frequencies in patients and controls; ROC, receiver-operator characteristic.

JCEM is published monthly by The Endocrine Society (<http://www.endo-society.org>), the foremost professional society serving the endocrine community.

was made on the basis of World Health Organization criteria (9) and a clinical requirement for insulin treatment from diagnosis. Patients with secondary diabetes, known genetic subtypes including Maturity Onset Diabetes of the Young, or clinical type 2 diabetes were not included in the study. The study was approved by the research ethics committees in all centers involved.

HLA class II typing

HLA class II typing for *DRB1*, *DQA1*, and *DQB1* was carried out on DNA from blood or mouth swab samples collected between 1998 and 2001 from 753 families remaining under follow-up in the study. Details of DNA extraction methods and HLA class II analysis have been published (10). Low-yield DNA samples from mouth swabs underwent whole-genome amplification by primer extension preamplification. HLA analysis was carried out by PCR using sequence-specific primers (11).

Statistical methods

For multiplex families in which type 1 diabetes was diagnosed in more than one child during the study period, only the first eligible child in the family was included in the analysis. In the absence of an independent control population, we used two analytical methods to define genotype frequency within the background population and compared results obtained from affected family-based controls (AFBACs) analyzed with homogeneity testing *vs.* pseudocontrols analyzed using a conditional logistic regression method.

AFBAC method. Data from family-based samples allowed unambiguous assignment of alleles to three-locus haplotypes (*DRB1-DQA1-DQB1*) in all but 34 families. In these 34 families, haplotypes were inferred using observed patterns of linkage disequilibrium. HLA haplotypes never transmitted to the affected individual were identified in all families (12). This AFBAC population provides an unbiased estimate of the overall population (control) HLA allele and haplotype frequencies, under the reasonable assumption of zero recombination between the marker (HLA) and disease, and random mating assumptions, *i.e.* no population stratification, admixture, or migration effects (12). Control genotypes were derived from the two nontransmitted AFBAC haplotypes in each family. Parental-transmitted *vs.* never-transmitted haplotypes were tested for differences using a χ^2 contingency table test for heterogeneity (12). The Bonferroni correction for multiple comparisons was deemed overly conservative for these data and was not applied in any of our tests. Tests for differences in predispositional and protective effects of HLA haplotypes and genotypes were performed using the odds ratio method for haplotypes and genotypes (13) and the relative predispositional effects method (14). For the relative predispositional effects method, the χ^2 contingency table test was performed at each round of analysis, rather than recalculation of expected values as in the original method, preventing bias toward positive associations. The ratio of haplotype and genotype frequencies in patients and controls (PC ratio) was calculated and used to provide a measure of susceptibility.

Conditional logistic regression method. Using genotype data on trios, the conditional logistic regression method generates between one and three pseudocontrols consisting of the alternative genotypes that could have been transmitted to the case (15). These are analyzed together with the case in a matched case-control design. Conditional logistic regression is used to fit models for the genotype relative risks. The 34 families in which unambiguous assignment of haplotypes was not possible were excluded from this analysis. This approach does not require assumptions of random mating, Hardy-Weinberg equilibrium, or absence of population stratification.

Sensitivity, specificity, and absolute risk of HLA class II genotypes

The sensitivity and specificity associated with each genotype were calculated according to age at onset of diabetes. Receiver-operator characteristic (ROC) curve analysis was used to evaluate the performance of genotype combinations in discriminating disease from nondisease (16). Successive genotypes were added in descending order of PC ratio. The area under the curve (AUC) with 95% confidence intervals was calculated assuming a nonparametric distribution. An AUC of 1.00 would

indicate that a strategy achieved 100% accuracy in identifying disease, whereas an AUC of 0.00 would indicate that all individuals were misclassified on the basis of the test, and an area of 0.5 would indicate that the test achieved random assignment of disease/nondisease status. The absolute risk of diabetes associated with each genetic marker was calculated for each age band based on the cumulative incidence of diabetes in the region over the period 1985–1996 (8) using the formula:

Absolute risk =

$$CI \times (\text{genotype frequency in cases} / \text{genotype frequency in controls})$$

where *CI* is the cumulative incidence of diabetes within the age band. The confidence interval around the estimated risk was calculated according to the methods of Miettinen (17).

Results

Patient characteristics

The analysis was based on 753 family trios (both parents and affected child) from the 1746 families in the BOX study cohort. Families were included in the analysis if full *HLA-DRB1-DQA1-DQB1* genotype data were available on all members of the trio. Of 753 probands, 426 (57%) were male, 97 (12.9%) had an affected first-degree relative at the time of diagnosis (the father was affected in 5.3% of families, the mother in 2.9%, and a sibling in 4.7%), and the median age at diagnosis was 10.5 yr (range 0.7–20.9 yr); 161 were diagnosed before age 5 yr, 195 at 5–9 yr, 285 at 10–14 yr, and 112 at 15–20 yr. These characteristics were similar to those of the whole BOX study cohort, which consists of 1816 individuals with type 1 diabetes from 1746 families, of which 1018 (56%) are male, 214 (12.3%) had an affected first degree relative at the time of diagnosis (the father was affected in 5.2% of families, the mother in 3.0%, and a sibling in 3.9%), and the median age at diagnosis was 10.3 yr (range 0.4–20.9 yr).

DRB1-DQA1-DQB1 haplotype associations with type 1 diabetes

The three-locus (*DRB1-DQA1-DQB1*) haplotypes from PCR using sequence-specific primer typing of 753 families with type 1 diabetes are listed in Table 1 and are ranked by PC ratio. These data show a strong predisposing effect for the *DRB1*03-DQA1*0501-DQB1*0201* and *DRB1*04-DQA1*0301-DQB1*0302* haplotypes. The relative susceptibility effect associated with each haplotype can be estimated by comparison of the PC ratios (Table 1). The conditional logistic regression model shows a similar hierarchy of susceptibility effects from the most predisposing to the most protective. The relative risks and *z* scores allow comparison of effects by haplotype.

Significant heterogeneity of haplotype effects was demonstrated among six *DRB1*04*-based haplotypes (0401, 0402, 0403, 0404, 0405, and 0408) (overall χ^2 42.8 on 5 df, $P = 4 \times 10^{-8}$) in a model incorporating additional haplotype effects for *DRB1*X-DQA1*0301-DQB1*0302* (*X* is not 04) and *X-Y* (*Y* is any haplotype at *DQA1-DQB1* other than 0301-0302). The *DRB1*0401-DQ8*, *DRB1*0405-DQ8*, and *DRB1*0402-DQ8* haplotypes were found to be predisposing, with odds ratios of 6.5 (5.0–8.6), 3.51 (1.8–6.9), and 1.8 (0.3–2.4), respectively, although the *DRB1*0405-DQ8* was only present at low frequency (0.025). *DRB1*0403* was protective (odds ratio of 0.2, 0.04–0.9).

TABLE 1. Three-locus haplotypes of 753 families with type 1 diabetes

<i>DRB1-DQA1-DQB1</i> haplotype	No. transmitted (n = 1506)	No. nontransmitted (n = 1506)	PC ratio	χ^2	Odds ratio (95% CI)	Relative risk (95% CI)	z score	P value (z score)
0401-0301-0302	360	69	5.217	197.392	6.54 (5.00–8.56)	1.0 (Reference)		
0405-0301-0302	38	11	3.455	14.878	3.52 (1.79–6.91)	0.73 (0.29–1.89)	–0.64	0.522
03-0501-0201	525	205	2.561	140.274	3.40 (2.83–4.07)	0.42 (0.29–0.62)	–4.41	0.0001
0402-0301-0302	18	10	1.8	2.286	1.81 (0.83–3.93)	0.23 (0.07–0.82)	–2.26	0.024
0404-0301-0302	121	71	1.704	13.021	1.77 (1.31–2.39)	0.3 (0.18–0.5)	–4.69	0.0001
09-0301-0303	23	15	1.533	1.684	1.54 (0.80–2.97)	0.23 (0.09–0.55)	–3.26	0.001
08-NT-04	25	28	0.893	0.17	0.89 (0.52–1.54)	0.17 (0.08–0.36)	–4.64	0.0001
01-NT-0501	133	166	0.801	3.642	0.78 (0.61–0.99)	0.13 (0.08–0.21)	–8.60	0.0001
0401-0301-0301	82	130	0.631	10.868	0.61 (0.46–0.81)	0.09 (0.06–0.16)	–9.06	0.0001
02-NT-0502	8	18	0.444	3.846	0.44 (0.19–1.02)	0.04 (0.01–0.15)	–5.13	0.0001
07-0201-02	64	155	0.413	37.813	0.39 (0.29–0.52)	0.06 (0.04–0.1)	–10.96	0.0001
06-NT-0603	47	118	0.398	30.551	0.38 (0.27–0.54)	0.06 (0.04–0.11)	–9.44	0.0001
0403-0301-0302	2	10	0.2	5.333	0.20 (0.04–0.91)	0.07 (0.01–0.33)	–3.32	0.001
10-0501-0501	2	10	0.2	5.333	0.20 (0.04–0.91)	NA		
11-0501-0301	17	110	0.155	68.102	0.14 (0.09–0.24)	0.03 (0.01–0.06)	–9.99	0.0001
12-0501-0301	5	33	0.152	20.632	0.15 (0.06–0.38)	0.02 (0.006–0.07)	–6.30	0.0001
07-0201-0303	6	58	0.103	42.25	0.10 (0.04–0.23)	0.02 (0.006–0.05)	–7.15	0.0001
0407-0301-0301	1	11	0.091	8.333	0.09 (0.01–0.70)	0.01 (0.002–0.12)	–3.89	0.0001
02-NT-0602	3	204	0.015	195.174	0.01 (0–0.04)	0.005 (0.002–0.01)	–9.87	0.0001
Others	26	74						

CI, Confidence interval.

TABLE 2. Genotypes ranked by PC ratio

	Genotype: <i>DRB1-DQA1-DQB1</i>		Probands (n = 753)	PC ratio	χ^2	Odds ratio (95% CI)	Relative risk	z score	P value (z score)
	Haplotype 1	Haplotype 2							
1	03-0501-0201	0401-0301-0302	163	17.35	11.21	21.87 (11.23–42.6)	1.0		
2	0401-0301-0302	0401-0301-0302	27	17.08	5.97	17.68 (3.54–88.2)	0.77 (0.34–1.77)	–0.61	0.544
3	03-0501-0201	0405-0301-0302	24	16.03	8.73	16.52 (3.16–86.39)	0.52 (0.18–1.5)	–1.21	0.224
4	0401-0301-0302	08-NT-04	13	10.13	8.26	10.29 (1.67–63.3)	0.57 (0.19–1.70)	–1.01	0.312
5	0401-0301-0302	09-0301-0303	6	8.73	0.05	8.79 (0.72–106.9)	0.87 (0.18–4.14)	–0.17	0.865
6	03-0501-0201	0402-0301-0302	11	8.08	3.56	8.19 (1.38–48.72)	0.22 (0.05–0.99)	–1.98	0.048
7	03-0501-0201	0404-0301-0302	72	7.45	21.08	8.13 (4.12–16.04)	0.39 (0.2–0.76)	–2.74	0.006
8	03-0501-0201	09-0301-0303	13	6.37	3.10	6.46 (1.47–28.35)	0.19 (0.06–0.62)	–2.77	0.006
9	01-NT-0501	0401-0301-0302	47	6.18	7.27	6.52 (3.01–14.13)	0.48 (0.25–0.91)	–2.24	0.025
10	0401-0301-0302	0405-0301-0302	3	5.95	4.07	5.97 (0.3–118.23)	0.47 (0.04–5.13)	–0.62	0.537
11	03-0501-0201	03-0501-0201	71	5.09	4.60	5.51 (3.08–9.88)	0.22 (0.12–0.4)	–4.87	0.0001
12	0401-0301-0302	0404-0301-0302	12	3.69	9.90	3.73 (1.09–12.76)	0.13 (0.05–0.39)	–3.71	0.0001
13	0401-0301-0302	06-NT-0603	19	3.51	5.37	3.58 (1.37–9.36)	0.2 (0.08–0.49)	–3.54	0.0001
14	0401-0301-0302	07-0201-02	22	3.10	2.94	3.16 (1.35–7.41)	0.14 (0.07–0.3)	–5.27	0.0001
15	0404-0301-0302	0404-0301-0302	5	2.99	0	3.00 (0.52–17.32)	0.07 (0.02–0.29)	–3.78	0.0001
16	0405-0301-0302	07-0201-02	3	2.65	1.19	2.66 (0.31–23.13)	0.08 (0.01–0.45)	–2.86	0.004
17	01-NT-0501	0405-0301-0302	3	2.47	0.04	2.48 (0.3–20.49)	0.12 (0.02–0.92)	–2.05	0.041
18	07-0201-02	08-NT-04	6	2.08	22.95	2.09 (0.51–8.55)	0.08 (0.02–0.3)	–3.68	0.0001
19	0401-0301-0302	11-0501-0301	9	1.79	6.00	1.80 (0.6–5.37)	0.05 (0.02–0.14)	–5.56	0.0001
20	01-NT-0501	03-0501-0201	38	1.68	1.51	1.72 (1.01–2.92)	0.1 (0.05–0.2)	–6.62	0.0001
21	0401-0301-0301	0401-0301-0301	8	1.43	14.90	1.43 (0.48–4.23)	0.1 (0.03–0.37)	–3.48	0.001
22	0401-0301-0301	0401-0301-0302	8	1.34	6.87	1.35 (0.46–3.91)	0.07 (0.02–0.22)	–4.45	0.0001
23	01-NT-0501	08-NT-04	4	1.30	1.45	1.30 (0.29–5.75)	0.09 (0.02–0.47)	–2.88	0.0004
24	03-0501-0201	06-NT-0603	17	1.06	0.02	1.06 (0.53–2.11)	0.07 (0.03–0.16)	–6.32	0.0001
25	0401-0301-0301	07-0201-02	13	0.97	25.98	0.97 (0.45–2.10)	0.04 (0.02–0.11)	–6.86	0.0001
26	01-NT-0501	0401-0301-0301	13	0.91	4.58	0.91 (0.42–1.93)	0.05 (0.02–0.14)	–5.92	0.0001
27	01-NT-0501	0404-0301-0302	7	0.89	1.27	0.89 (0.32–2.49)	0.05 (0.01–0.15)	–5.02	0.0001
28	03-0501-0201	0401-0301-0301	12	0.68	9.62	0.67 (0.32–1.41)	0.03 (0.01–0.07)	–7.89	0.0001
29	03-0501-0201	12-0501-0301	3	0.67	0.91	0.67 (0.15–2.88)	0.03 (0.007–0.17)	–4.11	0.0001
30	0404-0301-0302	07-0201-02	4	0.55	0.25	0.54 (0.16–1.85)	0.05 (0.01–0.19)	–4.24	0.0001
31	03-0501-0201	07-0201-02	9	0.43	7.94	0.42 (0.19–0.92)	0.04 (0.02–0.09)	–7.20	0.0001
32	01-NT-0501	07-0201-02	3	0.18	1.24	0.17 (0.05–0.59)	0.01 (0.002–0.04)	–6.71	0.0001
	Others		85						

Genotype effects

Table 2 shows genotypes ranked by PC ratio. The genotype with the greatest predisposing effect was the DRB1*03-DQA1*0501-DQB1*0201/DRB1*0401-DQA1*0301-DQB1*

0302 (odds ratio of 21.9, 11.2–42.6). The genotypes DRB1*0401-DQA1*0301-DQB1*0302/ DRB1*0401-DQA1*0301-DQB1*0302 and DRB1*03-DQA1*0501-DQB1*0201/DRB1*0405-DQA1*0301-DQB1*0302 were also strongly predisposing with odds

ratios of 17.7 (3.5–88.8) and 16.5 (3.2–86.4), respectively, and relative risks (relative to most predisposing) of 0.8 (0.3–1.8) and 0.5 (0.2–1.7). However, these three accounted for only 214 of 753 cases (28%).

Sensitivity and specificity of HLA class II genotypes

The sensitivity of individual genotypes is shown in Table 3, subdivided by age at diagnosis. The DRB1*03-DQA1*0501-DQB1*0201/DRB1*0401-DQA1*0301-DQB1*0302 genotype achieved an overall sensitivity of 22.6% (19.4–25.9) for those diagnosed up to age 14 yr. The specificity for the DRB1*03-DQA1*0501-DQB1*0201/DRB1*0401-DQA1*0301-DQB1*0302 genotype was, however, only 24.7% (21.2–28.1). The next ranked genotype (DRB1*03-DQA1*0501-DQB1*0201/DRB1*0405-DQA1*0301-DQB1*0302) achieved 3.4% sensitivity (2.0–4.8).

ROC curves for these HLA class II genotypes are shown in Fig. 1. The AUC for the children aged 0–4 yr was 0.93 (0.91–0.96), for the 5- to 9-yr curve 0.87 (0.83–0.9), for the 10- to 14-yr curve 0.86 (0.83–0.89), and for the 0- to 14-yr curve 0.88 (0.86–0.9), indicating that HLA class II genotype analysis achieved significantly better discrimination of disease from nondisease in the youngest age group. This difference

was attributable to the high sensitivity achieved by the highest risk genotypes in the 0- to 4-yr age group. If assignment of risk was based only on the DRB1*03-DQA1*0501-DQB1*0201/DRB1*0401-DQA1*0301-DQB1*0302 genotype, the sensitivity in the 0- to 4-yr age group was 33.5% (26.2–40.8) compared with 19.9% (14.3–25.5) and 18.3% (13.8–22.8) in the older age bands. If the six highest genotypes were taken into consideration, the sensitivity rose to 48.4% (40.7–56.1) in the 0- to 4-yr age group, 32.7% (26.1–39.2) in the 5- to 9-yr age group, and 28.96% (23.6–34.1) in those aged 10–14 yr. The increase in sensitivity associated with inclusion of the remaining 26 genotypes shown in Table 3 was similar in all of the three age bands.

The absolute risk of diabetes associated with each genotype calculated using incidence data for the study population for the period 1985–1996 (8) is shown in Fig. 2. The absolute risk of diabetes for the DRB1*03-DQA1*0501-DQB1*0201/DRB1*0401-DQA1*0301-DQB1*0302 genotype was 17 (7–40) cases per 1000 population by age 5 and 51 yr (31–82) cases per 1000 population by age 15 yr. The cumulative incidence of diabetes for the whole population, irrespective of genotype, was 0.6 cases per 1000 by age 5 yr and three cases per 1000 by age 15 yr.

TABLE 3. Sensitivity of individual genotypes

Genotype: DRB1-DQA1-DQB1		Age at Diagnosis						
Haplotype 1	Haplotype 2	0–4 yr			0–14 yr			
		Probands (n = 161)	Sensitivity (%; 95% CI)	Absolute risk (%; 95% CI)	Probands (n = 641)	Sensitivity (%; 95% CI)	Absolute risk (%; 95% CI)	
1	03-0501-0201	0401-0301-0302	54	33.5 (26.2–40.8)	1.7 (0.7–4)	145	22.6 (19.4–25.9)	5.1 (3.1–8.2)
2	0401-0301-0302	0401-0301-0302	6	3.7 (0.8–6.7)	1.1 (0.1–13.5)	23	3.6 (2.1–5)	4.8 (1.4–16.7)
3	03-0501-0201	0405-0301-0302	6	3.7 (0.8–6.7)	1.2 (0.1–14.8)	22	3.4 (2–4.8)	4.8 (1.3–17.45)
4	0401-0301-0302	08-NT-04	4	2.5 (0.1–4.9)	0.9 (0.05–16.8)	10	1.6 (0.6–2.5)	2.6 (0.5–12.8)
5	0401-0301-0302	09-NT-0303	0			5	0.8 (0.1–1.5)	2.4 (0.3–22.5)
6	03-0501-0201	0402-0301-0302	2	1.2 (0–3)	0.4 (0.01–12.2)	10	1.6 (0.6–2.5)	2.4 (0.5–11.8)
7	03-0501-0201	0404-0301-0302	16	9.9 (5.3–14.6)	0.5 (0.1–1.6)	62	9.7 (7.4–12)	2.1 (1.2–3.8)
8	03-0501-0201	09-NT-0303	7	4.3 (1.2–7.5)	1.0 (0.1–9.5)	13	2 (0.9–3.1)	2.1 (0.5–7.9)
9	01-NT-0501	0401-0301-0302	8	5 (1.6–8.3)	0.3 (0.07–1.4)	40	6.2 (4.4–8.1)	1.7 (0.8–3.5)
10	0401-0301-0302	0405-0301-0302	1	0.6 (0–1.8)	0.6 (0.003–100)	3	0.5 (0–1)	2.0 (0.1–30.3)
11	03-0501-0201	03-0501-0201	11	6.8 (2.9–10.7)	0.2 (0.07–0.8)	55	8.6 (6.4–10.7)	1.3 (0.7–2.2)
12	0401-0301-0302	0404-0301-0302	3	1.9 (0–4)	0.3 (0.03–2.9)	10	1.6 (0.6–2.5)	1.0 (0.3–3.5)
13	0401-0301-0302	06-NT-0603	5	3.1 (0.4–5.8)	0.3 (0.04–1.7)	15	2.3 (1.7–3.5)	0.9 (0.3–2.4)
14	0401-0301-0302	07-0201-02	5	3.1 (0.4–5.8)	0.2 (0.04–1.1)	18	2.8 (1.5–4.1)	0.8 (0.3–2.0)
15	0404-0301-0302	0404-0301-0302	0			5	0.8 (0.1–1.5)	1.0 (0.2–5.6)
16	0405-0301-0302	07-0201-02	1	0.6 (0–1.8)	0.3 (0.004–15.5)	3	0.5 (0–1)	0.9 (0.1–7.6)
17	01-NT-0501	0405-0301-0302	1	0.6 (0–1.8)	0.2 (0.004–13.4)	2	0.3 (0–0.7)	0.5 (0.05–5.6)
18	07-0201-02	08-NT-04	1	0.6 (0–1.8)	0.1 (0.004–2.4)	4	0.6 (0–1.2)	0.5 (0.09–2.2)
19	0401-0301-0302	11-0501-0301	2	1.2 (0–3)	0.1 (0.01–1.2)	7	1.1 (0.3–1.9)	0.5 (0.1–1.5)
20	01-NT-0501	03-0501-0201	7	4.3 (1.2–7.5)	0.09 (0.03–0.3)	35	5.5 (3.7–7.2)	0.5 (0.3–0.9)
21	0401-0301-0301	0401-0301-0301	0			7	1.1 (0.3–1.9)	0.4 (0.1–1.3)
22	0401-0301-0301	0401-0301-0302	1	0.6 (0–1.8)	0.05 (0.004–0.7)	7	1.1 (0.3–1.9)	0.4 (0.1–1.2)
23	01-NT-0501	08-NT-04	0			2	0.3 (0–0.7)	0.2 (0.03–1.3)
24	03-0501-0201	06-NT-0603	2	1.2 (0–3)	0.04 (0.007–0.2)	14	2.2 (1.1–3.3)	0.3 (0.1–0.6)
25	0401-0301-0301	07-0201-02	2	1.2 (0–3)	0.2 (0.02–4.1)	11	1.7 (0.7–2.7)	0.3 (0.1–0.6)
26	01-NT-0501	0401-0301-0301	3	1.9 (0–4)	0.06 (0.01–0.3)	12	1.9 (0.8–2.9)	0.3 (0.1–0.6)
27	01-NT-0501	0404-0301-0302	1	0.6 (0–1.8)	0.04 (0.003–0.4)	6	0.9 (0.2–1.7)	0.3 (0.08–0.8)
28	03-0501-0201	0401-0301-0301	2	1.2 (0–3)	0.03 (0.006–0.2)	10	1.6 (0.6–2.5)	0.2 (0.08–0.4)
29	03-0501-0201	12-0501-0301	0			3	0.5 (0–1)	0.2 (0.05–1.0)
30	0404-0301-0302	07-0201-02	0			4	0.6 (0–1.2)	0.2 (0.05–0.6)
31	03-0501-0201	07-0201-02	2	1.2 (0–3)	0.03 (0.006–0.14)	7	1.1 (0.3–1.9)	0.1 (0.05–0.3)
32	01-NT-0501	07-0201-02	1	0.6 (0–1.8)	0.02 (0.002–0.14)	2	0.3 (0–1.7)	0.04 (0.01–0.1)
	Others		7	4.3 (1.2–7.5)	0.004 (0.0002–0.006)	69	10.8 (8.4–13.2)	0.044 (0.04–0.05)

CI, Confidence interval.

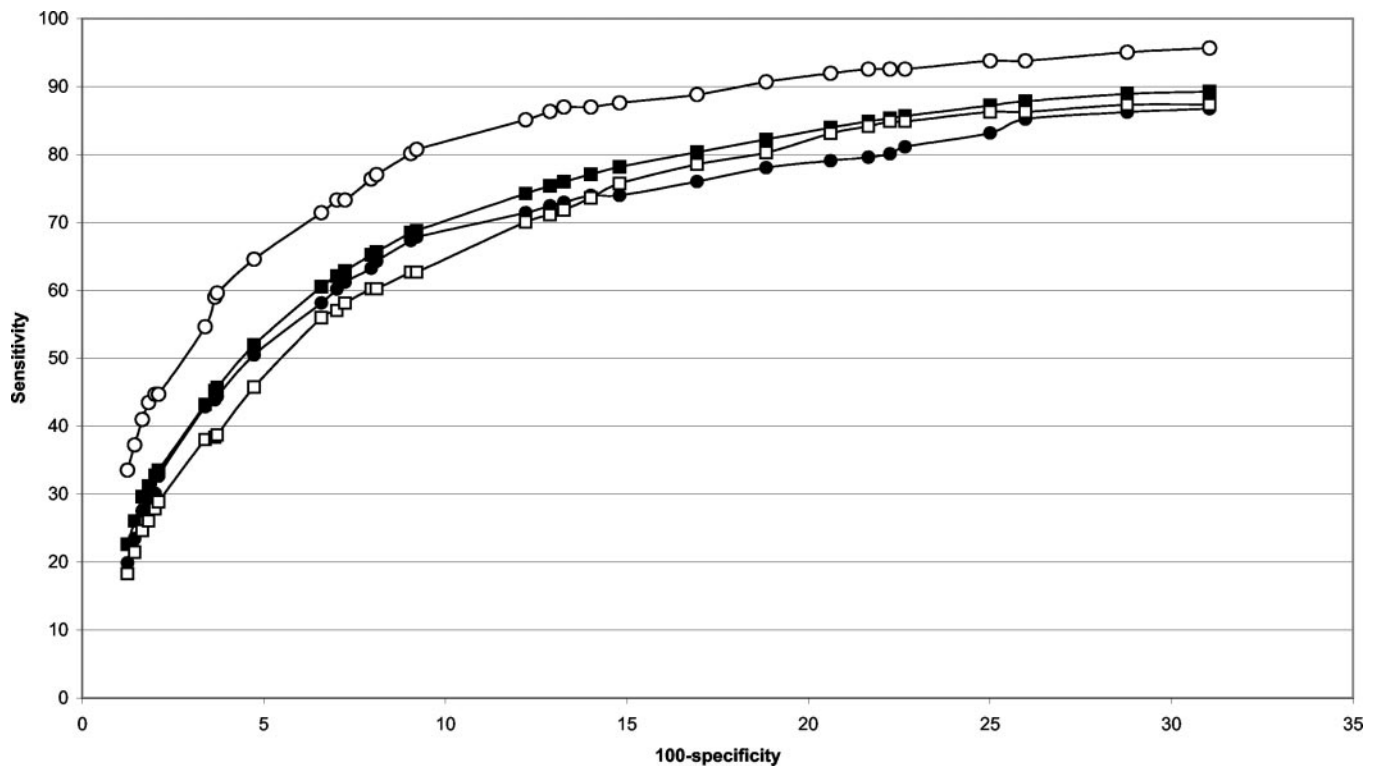


FIG. 1. ROC curves showing sensitivity against 100-specificity for HLA class II genotypes in type 1 diabetes added sequentially in descending order of PC ratio as given in Table 3. Patients are subdivided by age at diagnosis of diabetes. *White circle*, Diagnosed age 0–4 yr; *black circle*, 5–9 yr; *white square*, 10–14 yr; *black square*, 0–14 yr.

Discussion

Allelic variations of the HLA class II genes *DRB1* and *DQB1* on chromosome 6 make the major contribution to type 1 diabetes genetic susceptibility in Caucasian populations, and correlations have been reported between their allele frequencies and the incidence of type 1 diabetes in different countries (18, 19). However, this genetic locus, termed *IDDM1*, is complex, with epistasis between *DRB1* and *DQB1* manifested by particular *DRB1-DQB1* haplotype disease associations, trans or genotype effects probably involving *DRB1*, *DQB1*, and *DQA1*, and other as yet unidentified genes near the class II region modifying class II risk. We observed a hierarchy of susceptibility effects for HLA class II haplotypes, ranging across a 200-fold risk gradient (Table 1). Analysis of the risk associated with *DRB1*04* subtypes in the presence of the *DQB1*0302* allele showed that, even within this high-risk group, there is 20-fold difference between the strongly predisposing *0401* allele and the protective *0403* allele. Genotype-specific risk cannot be inferred from haplotype-associated risk, and we were able to assign positions within a hierarchy of risk to genotypes occurring at lower frequency, albeit with wider confidence intervals, as well as to the common high-risk genotypes (Table 2).

Recruitment to our family study was based on ascertainment of sporadic cases from the general population and had no bias toward familial diabetes. Although we did not obtain genetic samples from all families, the demographic characteristics of the large subset we studied are closely matched to those of the whole study population and are therefore likely to be representative. Because our study is population-based with validated incidence data, we were able to determine the absolute

risk of diabetes associated with individual genotypes, with the caveat that some high-ranking genotypes were found in only a small number of cases so that the confidence intervals around the absolute risk were wide. The overall cumulative incidence of type 1 diabetes by age 15 yr within the Oxford region was three cases per 1000, but the risk rose to 44 cases per 1000 for individuals carrying any one of the six highest risk genotypes. This approximates to the level of risk conferred by a first degree family history of diabetes but applies to a much greater proportion of future cases (32%) because only 12.9% of the children with diabetes in the study had an affected parent or sibling (8). Practical application of a predictive marker requires consideration of sensitivity as well as absolute risk and also of any potential variation in the genetic associations of different disease phenotypes, for example according to age of onset (20). The highest risk *DRB1*03-DQA1*0501-DQB1*0201/DRB1*0401-DQA1*0301-DQB1*0302* genotype was strongly associated with diabetes onset before the age of 5 yr, but the proportion of cases carrying this genotype declined sharply above this age. Equally, the sensitivity of all *DRB1*03/DRB1*04* genotypes in children diagnosed below the age of 5 yr was almost 50% but fell to 23% in those diagnosed aged 10–14 yr. The ROC curve shown in Fig. 1 portrays the balance of sensitivity and specificity achieved by sequential addition of genotypes over the 0- to 14-yr age range and within each 5-yr band.

Accurate definition of genotype frequency within the background population is fundamental to the approach we have used. In the absence of a large control population from the same region, we elected to use family-based controls. Two independent approaches are available, AFBAC methods and condi-

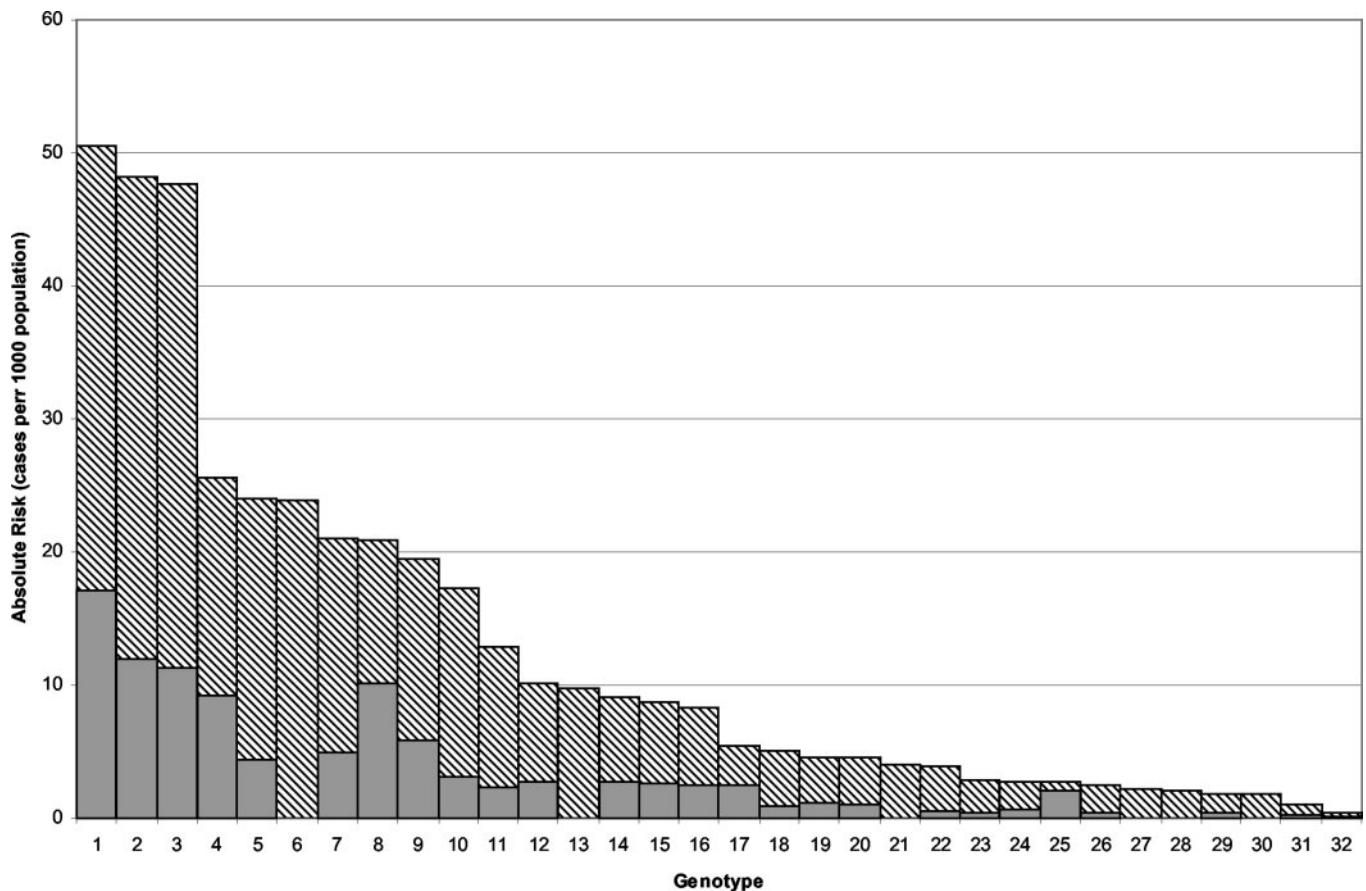


FIG. 2. Absolute risks associated with HLA class II genotypes for diagnosis of type 1 diabetes. Absolute risk by age 5 yr is shown by *black bars*, and absolute risk by age 15 yr is shown by *hatched bars*. The background cumulative incidence of type 1 diabetes was 0.6 per 1000 by age 5 yr and three per 1000 by age 15 yr (9). Genotypes are numbered as in Table 3.

tional logistic regression. Because AFBAC methods for analyzing such data rely on a number of assumptions that are not required by the conditional logistic regression method, we opted to use both approaches and to compare the results. Specifically, the AFBAC method assumes random mating in both the parental and grandparental population to satisfy the condition of Hardy-Weinberg equilibrium in the overall population from which the parental pool of this sample belongs and also assumes unambiguous determination of transmitted and nontransmitted haplotypes. Discarding families in which these haplotypes cannot be determined can potentially produce a bias in AFBAC-estimated haplotype frequencies. In the event, it was reassuring to find that both methods produce highly concordant results for both haplotype and genotype susceptibility effects (Tables 1 and 2).

On the basis of our findings, we are able to calculate the sample size potentially required for future controlled trials of primary interventions in our region. We might, for example, be willing to use a relatively safe intervention in children who had a 10 in 1000 risk of diabetes within the study period. If we consider a 50% reduction in progression to disease to be clinically relevant and want 80% power to detect this effect with 5% significance, 1468 children with this risk would need to be included in each treatment group. One option would be to include only neonates with the three highest risk genotypes (numbers 1–3) and study them for 5 yr. This would require

around 300,000 children to be screened and would identify some 41% of future cases. Another option would be to include neonates with any of the first 13 highest risk genotypes and study them for 15 yr. This would require 60,000 children to be screened and would identify some 63% of future cases. Using a surrogate measure, such as the appearance of strongly predictive islet antibody combinations, as the trial endpoint could reduce the length of follow-up, but the logistics of primary prevention trials in type 1 diabetes remain formidable.

Although the role of the HLA class II region in determining genetic susceptibility to type 1 diabetes is well established, surprisingly few large studies have attempted to optimize the use of genetic information for screening in the general population. In 1992, the Childhood Diabetes in Finland study analyzed 757 families and assigned absolute levels of risk on the basis of extended HLA class I and II haplotype, showing that three haplotypes, including a novel Finnish susceptibility haplotype, accounted for 26% of the cases in their population (21). The absolute risk associated with *HLA-DQA1* and *-DQB1* alleles has been reported for the Belgian population based on 1866 islet autoantibody-positive type 1 diabetes patients and 750 controls (22). Four genotypes were identified as conferring a highly significant disease risk ($P < 10^{-6}$). These were carried by 9% of controls and 60% of patients diagnosed before 40 yr (70% of those diagnosed under 5 yr) and, as a group, these four genotypes conferred

an absolute risk of developing diabetes before age 40 yr of 2.6%. The prospective Finnish Type 1 Diabetes Prediction and Prevention project proposed a two-step strategy for identification of neonates at high risk and has also undertaken some prospective evaluation. This strategy is based on initial screening for any of the five *DQB1* alleles 02, 0301, 0302, 0602, and 0603, followed by testing for low risk *DQA1* alleles on a *DQB1**02 background and *DRB1**04 subtypes in individuals with *DQB1**0302 (23). This would be expected to identify 50% of the future cases of type 1 diabetes in the Finnish population using a combination of the two-step screening process with follow-up for immune markers (24). Screening of 31,526 children born between 1994 and 1999 identified 14% in the high genetic risk category, and 17 (77%) of the 22 children from the cohort who have developed type 1 diabetes carried the high-risk genetic susceptibility alleles. The feasibility of newborn population screening by testing cord blood for selected *DRB1* and *DQB1* alleles has also been demonstrated in 5045 babies from the Denver general population (25).

We have optimized genetic risk assessment by genotype analysis within a large well-characterized population, using family-based controls. This has allowed us to assign absolute risk of progression to diabetes for a given genotype. Consideration of the extended haplotype incorporating the HLA class I and III regions could extend the power of this approach, although it remains to be seen whether this will modify the risk estimates given here to any substantial extent. As recently illustrated for the insulin gene *IDDM2* locus in a German population, susceptibility loci elsewhere in the genome may also aid risk assessment (26). Empirical data such as these can be applied to natural history studies, allowing environmental risk factors to be evaluated in subgroups at differing levels of genetic susceptibility. They also provide an improved basis for design of future primary prevention studies in the general population.

Acknowledgments

We thank Kyla Chandler, Sue Coe, Rhona Langley, Katie Lawrence, Natasha Rook, Jo Saunby, and Sarah Valovin for technical assistance. We are grateful to the study fieldworkers, the pediatricians, physicians, and diabetes nurse specialists in the Oxford region, and all the families participating in the BOX study for their continuing help and support.

Received December 4, 2003. Accepted May 10, 2004.

Address all correspondence and requests for reprints to: Prof. Polly Bingley, Diabetes and Metabolism, Medical School Unit, Southmead Hospital, Bristol BS10 5NB, United Kingdom. E-mail: Polly.Bingley@bristol.ac.uk.

This work was supported by the Wellcome Trust, by Diabetes UK, and by the Juvenile Diabetes Research Foundation. G.T. is funded by the National Institutes of Health (Grant NIH GM35326).

References

1. Atkinson MA, Eisenbarth GS 2001 Type 1 diabetes: new perspectives on disease pathogenesis and treatment. *Lancet* 358:221–229
2. Bloom A, Hayes TM, Gamble DR 1975 Register of newly diagnosed diabetic children. *Br Med J* 3:580–583
3. Dahlquist G, Blom L, Holmgren G, Hagglof B, Larsson Y, Sterky G, Wall S

- 1985 The epidemiology of diabetes in Swedish children 0–14 years: a six-year prospective study. *Diabetologia* 28:802–808
4. Cucca F, Lampis R, Congia M, Angius E, Nutland S, Bain SC, Barnett AH, Todd JA 2001 A correlation between the relative predisposition of MHC class II alleles to type 1 diabetes and the structure of their proteins. *Hum Mol Genet* 10:2025–2037
5. Noble JA, Valdes AM, Cook M, Klitz W, Thomson G, Erlich HA 1996 The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian multiplex families. *Am J Hum Genet* 59:1134–1148
6. Thomson G, Robinson WP, Kuhner MK, Joe S, MacDonald MJ, Gottschall JL, Barbosa J, Rich SS, Bertrams J, Baur MP, Partanen J, Tait B, Schober E, Mayr WB, Ludvigsson J, Lindblom B, Farid NR, Thompson C, Deschamps I 1988 Genetic heterogeneity, modes of inheritance, and risk estimates for a joint study of Caucasians with insulin-dependent diabetes mellitus. *Am J Hum Genet* 43:799–816
7. Bingley PJ, Gale EAM 1989 The incidence of insulin-dependent diabetes in England: a study in the Oxford region 1985–6. *Br Med J* 298:558–560
8. Gardner SG, Bingley PJ, Sawtell PA, Weeks S, Gale EAM, The BRT's-Oxford Study Group 1997 Rising incidence of insulin dependent diabetes in children aged under 5 years in the Oxford Region: time trend analysis. *BMJ* 315:713–717
9. World Health Organization 1985 Diabetes mellitus: report of a World Health Organization Study Group. Geneva: World Health Organization, 1985
10. Gillespie KM, Valovin SJ, Saunby J, Hunter KM, Savage DA, Middleton D, Todd JA, Bingley PJ, Gale EAM 2000 HLA class II typing of whole genome amplified mouth swab DNA. *Tissue Antigens* 56:530–538
11. Bunce M, O'Neill CM, Barnado M, Krausa P, Morris PJ, Welsh KI 1995 Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 and DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). *Tissue Antigens* 46:355–367
12. Thomson G 1995 HLA disease associations: models for the study of complex human genetic disorders. *Crit Rev Clin Lab Sci* 32:183–219
13. Woolf B 1955 On estimating the relation between blood group and diseases. *Ann Eugen* 19:251–253
14. Payami H, Joe S, Farid NR, Stenszky V, Chan SH, Yeo PP, Cheah JS, Thomson G 1989 Relative predispositional effects (RPEs) of marker alleles with disease: HLA-DR alleles and Graves disease. *Am J Hum Genet* 45:541–546
15. Cordell HJ, Clayton D 2002 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* 70:141
16. Zweig MH, Campbell G 1995 Receiver-operating characteristic (ROC) plots: a fundamental tool in clinical medicine. *Clin Chem* 39:561–567
17. Miettinen O 1976 Estimability and estimation in case-referent studies. *Am J Epidemiol* 103:226–235
18. Dorman JS, LaPorte RE, Stone RA, Trucco M 1990 Worldwide differences in the incidence of type 1 diabetes are associated with amino acid variation at position 57 of the HLA-DQ β chain. *Proc Natl Acad Sci USA* 87:7370–7374
19. Ronningen KS, Keiding N, Green A on behalf of the Genomic Marker contributors and the EURODIAB ACE Study Group 2001 Correlations between the incidence of childhood-onset type 1 diabetes in Europe and HLA genotypes. *Diabetologia* 44(Suppl 3):B51–B59
20. Gillespie KM, Gale EAM, Bingley PJ 2002 High familial risk and genetic susceptibility in early onset childhood diabetes. *Diabetes* 51:210–214
21. Tienari PJ, Tuomilehto-Wolf E, Tuomilehto J, Peltonen L 1992 HLA haplotypes in type 1 (insulin-dependent) diabetes mellitus: molecular analysis of the HLA-DQ locus. The DIME Study Group. *Diabetologia* 35:254–260
22. Van der Auwera BJ, Schuit FC, Weets I, Ivens A, Van Autreve JE, Goris FK; Belgian Diabetes Registry 2002 Relative and absolute HLA-DQA1-DQB1 linked risk for developing type 1 diabetes before 40 years of age in the Belgian population: implications for future prevention studies. *Hum Immunol* 63:40–50
23. Nejentsev S, Sjoroos M, Soukka T, Knip M, Simell O, Lovgren T, Ilonen J 1999 Population-based genetic screening for the estimation of Type 1 diabetes mellitus risk in Finland: selective genotyping of markers in the HLA-DQB1, HLA-DQA1 and HLA-DRB1 loci. *Diabet Med* 16:985–992
24. Kupila A, Muona P, Simell T, Arvilommi P, Savolainen H, Hamalainen AM, Korhonen S, Kimpimaki T, Sjoroos M, Ilonen J, Knip M, Simell O 2001 Feasibility of genetic and immunological prediction of type I diabetes in a population-based birth cohort. *Diabetologia* 44:290–297
25. Rewers M, Bugawan TL, Norris JM, Blair A, Beaty B, Hoffman M, McDuffie RS Jr, Hamman RF, Klingensmith G, Eisenbarth GS, Erlich HA 1996 Newborn screening for HLA markers associated with IDDM: diabetes autoimmunity study in the young (DAISY). *Diabetologia* 39:807–812
26. Walter M, Albert E, Conrad M, Keller E, Hummel M, Ferber K, Barratt BJ, Todd JA, Ziegler A-G, Bonifacio E 2003 IDDM2/insulin VNTR modifies risk conferred by IDDM1/HLA for the development of type 1 diabetes and associated autoimmunity. *Diabetologia* 46:712–720