Two Single Nucleotide Polymorphisms Identify the Highest-Risk Diabetes HLA Genotype

Potential for Rapid Screening

Jennifer M. Barker, Taylor M. Triolo, Theresa A. Aly, Erin E. Baschal, Sunanda R. Babu, Adam Kretowski, Marian J. Rewers, and George S. Eisenbarth

OBJECTIVE—People with the HLA genotype *DRB1**0301-*DQA1**0501-*DQB1**0201/*DRB1**04-*DQA1**0301-*DQB1**0302 (DR3/4-DQ8) are at the highest risk of developing type 1 diabetes. We sought to find an inexpensive, rapid test to identify DR3/4-DQ8 subjects using two single nucleotide polymorphisms (SNPs).

RESEARCH DESIGN AND METHODS—SNPs rs2040410 and rs7454108 were associated with DR3-*DQB1**0201 and DR4-*DQB1**0302. We correlated these SNPs with HLA genotypes and with publicly available data on 5,019 subjects from the Type 1 Diabetes Genetic Consortium (T1DGC). Additionally, we analyzed these SNPs in samples from 143 HLA-typed children who participated in the Diabetes Autoimmunity Study of the Young (DAISY) using Taqman probes (rs7454108) and restriction digest analysis (rs2040410).

RESULTS—With a simple combinatorial rule, the SNPs of interest identified the presence or absence of the DR3/4-DQ8 genotype. A wide variety of genotypes were tested for both SNPs. In T1DGC samples, the two SNPs were 98.5% (1,173 of 1,191) sensitive and 99.7% (3,815 of 3,828) specific for DR3/4-DQ8. In the DAISY population, the test was 100% (69 of 69) sensitive and 100% (74 of 74) specific. Overall, the sensitivity and specificity for the test were 98.57 and 99.67%, respectively.

CONCLUSIONS—A two-SNP screening test can identify the highest risk heterozygous genotype for type 1 diabetes in a time- and cost-effective manner. ***Diabetes*** **57:3152–3155, 2008**

We have the ability to identify subjects with a greater than 50% risk of developing anti-islet autoimmunity and type 1 diabetes on the basis of family history and HLA genotype (1,2). Siblings with the highest type 1 diabetes risk HLA genotype *DRB1**0301-*DQA1**0501-*DQB1**0201/*DRB1**04-*DQA1**0301-*DQB1**0302 (DR3/4-DQ8) who are identical by descent for the major histocompatibility complex region with a type 1 diabetic sibling have an 85% risk of developing diabetes-related autoimmunity by age 15 years and a 55% risk of developing type 1 diabetes by age 12 years (1). Children with multiple first-degree relatives with type 1 diabetes and high– or moderate–diabetes risk HLA genotypes are reported to have a 50% risk for the development of multiple diabetes-related autoantibodies and type 1 diabetes (2).

Prevention trials, including the Trial to Reduce Type 1 Diabetes in the Genetically at Risk (TRIGR) (3), the Nutritional Intervention to Prevent Diabetes (NIP-Diabetes) trial (4), and the Primary Oral and Intranasal Trial (Pre-POINT) are currently underway in genetically at-risk children. Many of these trials include first-degree relatives of people with diabetes as well as individuals with high-risk HLA genotypes, including DR3/4-DQ8. The identification of subjects for these trials requires large-scale HLA screening, with many children tested who do not have the highest type 1 diabetes risk. Current typing techniques for DR3/4-DQ8 often utilize coamplification of the *DQA1* and *DQB1* genes followed by multiple probe hybridization or direct sequencing. This technique uses sequence-specific oligonucleotides in a linear assay for hybridization with amplified product from DNA samples (5). Alleles are called with a customized typing program. Sequence-based typing techniques use PCR to amplify *DRB1* genes that are sequenced, with HLA type determined using special software (6). High-throughput screening systems that employ asymmetric PCR and hybridization of allele-specific probes as a first screening step to identify samples with specified HLA genotypes have been developed. Samples identified via these programs can be identified for further HLA genotyping (7). This method is cost and time consuming, as it may cost up to $31.44 to genotype 1 sample and takes up to 9 h to perform and analyze a set of 50 samples.

Several studies have examined the possibility of predicting HLA alleles from existing single nucleotide polymorphism (SNP) data (8–10); however, only one article has provided data on predicting specific HLA alleles from individual SNPs. de Bakker et al. (8) reported an association between HLA types and SNPs. SNP rs2040410 A allele was associated with *DRB1**0301, and rs7454108 C allele with *DQB1**0302. We have developed and tested the ability of these two SNPs to identify individuals with the DR3/4-DQ8 genotype in subjects within the Type 1 Diabetes Genetics Consortium (T1DGC) and the Diabetes Autoimmunity Study in the Young (DAISY). This novel method adds to existing knowledge by utilizing SNP technology to quickly identify individuals with the DR3/4-DQ8 genotype and may be beneficial to prevention trials because it provides high-throughput screening in a time- and cost-effective manner.

From the Barbara Davis Center for Childhood Diabetes, University of Colorado at Denver, Denver, Colorado.
Corresponding author: Jennifer M. Barker, jennifer.barker@uchsc.edu.
Received 5 May 2008 and accepted 5 August 2008.
Published ahead of print at http://diabetes.diabetesjournals.org on 11 August 2008. DOI: 10.2337/db08-0605.

TABLE 1
Characteristics by cohort

	n	Probands	Parents	DR3/4-DQ8	Not DR3/4-DQ8	Sensitivity DR3/4-DQ8 (*n*)	Specificity not DR3/4-DQ8 (*n*)
BDA	1,598	772	737	429	1,169	97.67% (419 of 429)	99.23% (1,160 of 1,169)
Denmark	675	309	228	150	525	98.67% (148 of 150)	99.81% (524 of 525)
HBDI	2,037	894	828	433	1,604	98.61% (427 of 433)	99.88% (1,602 of 1,604)
Poland	344	126	218	62	282	100.00% (62 of 62)	99.65% (281 of 282)
U.K.	365	182	151	117	248	100.00% (117 of 117)	100.00% (248 of 248)
DAISY	143	NA	NA	69	74	100.00% (69 of 69)	100.00% (74 of 74)
N	5,162	2,283	2,162	1,260	3,902	98.57% (1,242 of 1,260)	99.67% (3,889 of 3,902)

Data are *n* unless otherwise indicated. BDA, British Diabetic Association; HBDI, Human Biological Data Interchange.

RESEARCH DESIGN AND METHODS

The T1DGC enrolled sibling pairs with type 1 diabetes and their parents. HLA typing and two standard dense Illumina SNP panels of the major histocompatibility complex region were completed. We first evaluated 344 Polish subjects from the T1DGC, as this was the original dataset available to our group. We expanded the analysis to an additional five T1DGC cohorts—British Diabetic Association (n = 1,911), Denmark (n = 776), Human Biological Data Interchange (n = 2,605), Joslin Diabetes Center (n = 486), and U.K. (n = 519)—totaling 6,297 subjects from 1,240 families (Table 1). Subjects were enrolled with informed consent, and the study was approved by the local institutional review board or ethics committee.

We typed 143 subjects with a variety of *HLA-DR/DQ* genotypes from DAISY. DAISY has HLA-typed more than 30,000 newborns in Denver, Colorado, at the *HLA-DR/DQ* loci (11,12) and has followed a subset of these children with high, intermediate, and low risk with measurement of autoantibody levels (insulin, GAD65, and [insulinoma-associated protein 2] IA-2) (12). DAISY also follows a cohort of siblings and offspring of patients with type 1 diabetes. Subjects were enrolled with informed parental consent, and the study was approved by the institutional review board at the University of Colorado Denver.

Typing of SNPs. T1DGC samples were typed with Illumina multiplex technology with an extension/ligation reaction (details available at the Illumina Technology and Applications Web site [www.illumina.com]). This analysis utilized the dense standard Illumina mapping and exon-centric panels (1,536 SNPs in each panel with 115 overlapping SNPs; 2,837 of 3,072 SNPs successfully typed with a 92% SNP locus success rate). Both SNPs (rs7454108 and rs2040410) we used to identify DR3/4-DQ8 individuals were included in the Illumina panel. There were no data for rs2040410 in the Joslin T1DGC Illumina panels; therefore, individuals from that cohort were excluded.

We utilized two typing techniques to analyze the individual SNPs rs7454108 and rs2040410 in DAISY samples. We genotyped 143 samples for rs7454108 using Taqman probes (hybridization/extension reaction with a fluorescence detection system developed by Applied Biosystems). Given nearby polymorphisms, we were unable to use Taqman technology to analyze rs2040410 and instead utilized restriction enzyme digestion.

We designed primers to sequence the region surrounding rs2040410 after identifying known SNPs in the region (forward: 5′ TGT GCT GAG AGT TCC AGC CT 3′; reverse: 5′ CAC AAG GAC TCA TGG CTT GG 3′). We genotyped 40 samples at the rs2040410 SNP using the 3130 Capillary Sequencing Machine (BigDye Terminator v1.1, Applied Biosystems). A restriction enzyme site that distinguished the two alleles for this SNP was identified (BsrGI, digested for at least 16 h at 37°C; New England BioLabs). The restriction digest showed 100% correlation with the sequencing and was used to complete typing for rs2040410 in 119 samples.

HLA genotyping. We performed *DRB1* and *DQB1* genotyping on DAISY samples using linear arrays of immobilized sequence-specific oligonucleotides similar to previously described methodology (5).

Statistical analysis. We performed two-by-two contingency analysis (Fisher's exact test) and determined the sensitivity and specificity of the two marker SNPs for identifying the DR3/4-DQ8 genotype in previously typed individuals.

RESULTS

Figure 1 illustrates the location of rs2040410 (R^2 for *DRB1**0301 = 0.872) (8) and rs7454108 (R^2 for *DQB1**0302 = 1.0) (8) in relation to classical HLA genes. The rs2040410 A allele is associated with *DRB1**0301, and the rs7454108 C allele is associated with *DQB1**0302. We hypothesized that subjects with HLA genotype DR3/4-DQ8 would have the SNP genotype AG/CT. Of 6,641 T1DGC subjects, 5,019 (75.6%) had HLA genotyping and data for the SNPs of interest. Of the 1,191 subjects with DR3/4-DQ8, 1,121 (94.1%) had AG/CT and 52 (4.6%) had AA/CT. Of the 3,828 subjects that were not DR3/4-DQ8, only 12 (0.3%) had the genotype AG/CT and 1 (0.03%) had the genotype AA/CT. Using the two genotypes AG/CT and AA/CT as markers of the DR3/4-DQ8 genotype, we identified subjects with DR3/4-DQ8 with 98.5% sensitivity and 99.7% specificity (Table 2). Of the 13 false-positives, 4 (30.8%) were *DRB1**0301 and *DQB1**0302 but not DR4-

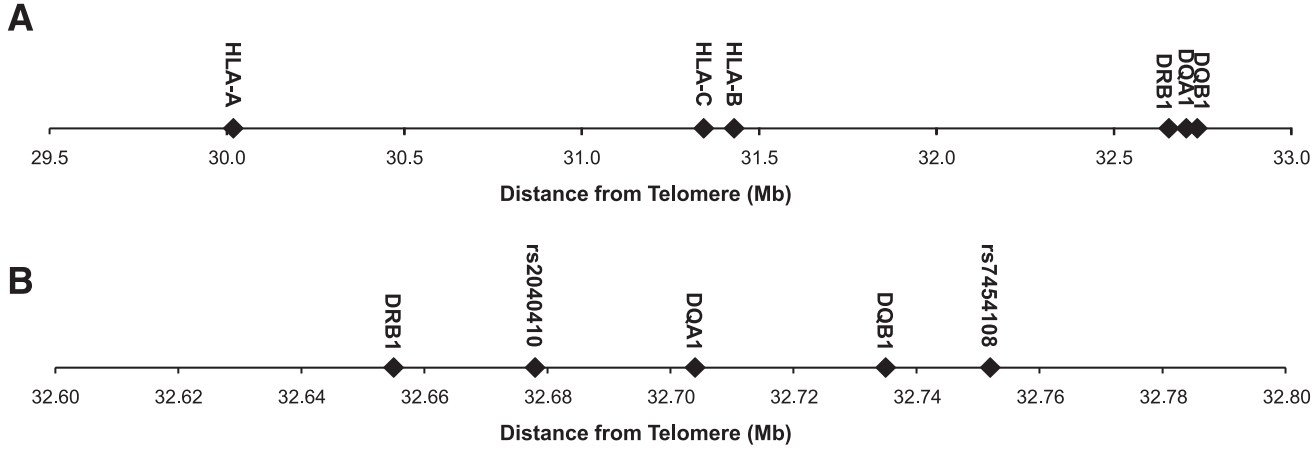

FIG. 1. Location of SNPs (♦) in relation to classical HLA loci (*A*) and *HLA-DRB1* and -*DQB1* loci (*B*). Numbers are distance in megabase (Mb) pairs from the telomere.

TABLE 2
Evaluation of rs2040410 and rs7454108 in subjects participating in the British Diabetic Association, Denmark, Human Biological Data Interchange, Poland, and U.K. T1DGC cohorts and in subjects in the DAISY population analyzed by restriction digest and Taqman

	DR3/4-DQ8	Not DR3/4-DQ8
n	1,191	3,828
T1DGC subjects		
rs2040410 A/G and rs7454108 C/T	1,121 (94.1)	12 (0.3)
rs2040410 A/A and rs7454108 C/T	52 (4.4)	1 (0.03)
rs2040410 G/G and rs7454108 C/T	7 (0.6)	1,384 (36.2)
rs2040410 A/G and rs7454108 C/C or T/T	9 (0.8)	1,086 (28.4)
rs2040410 A/A or G/G and rs7454108 C/C or T/T	2 (0.2)	1,345 (35.1)
DAISY subjects		
rs2040410 A/G and rs7454108 C/T	69 (100)	0 (0)
rs2040410 A/A and rs7454108 C/T	0 (0)	0 (0)
rs2040410 G/G and rs7454108 C/T	0 (0)	27 (36.5)
rs2040410 A/G and rs7454108 C/C or T/T	0 (0)	16 (21.6)
rs2040410 A/A or G/G and rs7454108 C/C or T/T	0 (0)	31 (41.9)

Data are *n* (%) unless otherwise indicated.

DQ8 (Online Appendix Table 1 [available at http://dx.doi.org/10.2337/db08-0605]).

In analysis of the British Diabetic Association cohort with DR3/4-DQ8, 87.4% (376 of 429) of subjects had the genotype AG/CT and 10.0% (43 of 429) had the genotype AA/CT, compared with 97.8% (745 of 762) and 1.18% (9 of 762), respectively, in the other cohorts ($P < 0.0001$). The unique frequency of the AA genotype of rs2040410 with the DR3/4-DQ8 genotype in this population suggests population differences within the U.K., as the U.K. cohort did not have increased frequency of the AA rs2040410 genotype.

Of the 2,162 parents in the T1DGC population, 200 were DR3/4-DQ8 and 97.0% (194 of 200) were AA/CT or AG/CT, while 99.7% (1,956 of 1,962) of subjects that were not DR3/4-DQ8 were not AA/CT or AG/CT. There were 2,283 probands in this population; 99.0% (901 of 910) of DR3/4-DQ8 probands were AA/CT or AG/CT, and 99.5% (1,366 of 1,373) of non–DR3/4-DQ8 probands were not AA/CT or AG/CT.

We evaluated 143 subjects in the DAISY study, 69 with DR3/4-DQ8. Subjects were chosen to represent a wide range of HLA genotypes, with a focus on DR3/4-DQ8 individuals (Online Appendix Table 2). All 69 subjects with DR3/4-DQ8 were AG/CT, and none of the subjects without DR3/4-DQ8 were AG/CT or AA/CT. Using the two genotypes AG/CT and AA/CT as markers of the DR3/4-DQ8 genotype, we identified subjects with DR3/4-DQ8 with 100% sensitivity and specificity (Table 1).

Overall, the sensitivity of the test was 98.57% and the specificity was 99.67% (Table 1). The rs2040410 A allele was present in 2,921 of the 3,018 (96.8%) individuals with *DRB1**0301. The rs7454108 C allele was present in 3,315 of the 3,353 (98.9%) individuals with *DQB1**0302.

DISCUSSION

In the U.S., approximately 1 in 300 individuals develop type 1 diabetes by age 18 years (13,14). The incidence of type 1 diabetes is increasing (15), especially at the youngest ages (16,17). Type 1 diabetes is characteristically preceded by the presence of anti-islet autoantibodies (18). Several groups of individuals have been identified as having a risk of type 1 diabetes greater than 50% by genetic factors alone (1,2). These groups have been proposed as ideal cohorts for intervention before the onset of autoimmunity, which may be key because it has proven difficult to halt the autoimmune process once it begins. Identification of these high-risk subjects for clinical trials will require screening many subjects for HLA genotypes. DR3/4-DQ8 individuals comprise between 30 and 50% of all children developing diabetes and thus represent an important high-risk group but only a subset of all children developing diabetes. It is likely that initial trials of immunomodulation (e.g., Pre-POINT) in genetically susceptible but autoantibody-negative children will primarily enroll only the highest-risk individuals, such as subsets of those with DR3/4-DQ8.

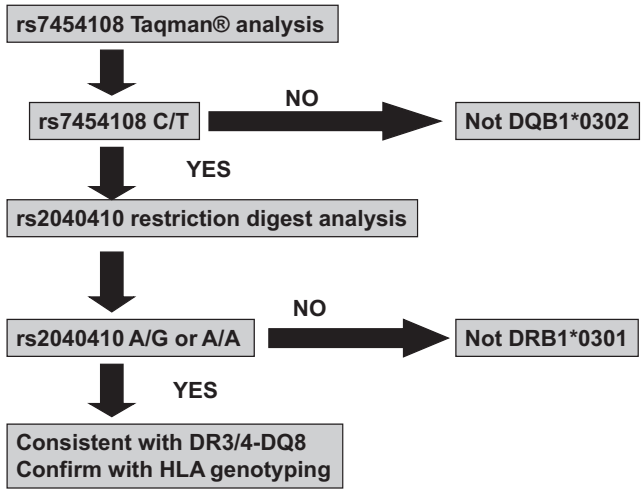

FIG. 2. Proposed algorithm for identifying subjects with DR3/4-DQ8 with SNP testing.

We confirmed the observations made by de Bakker et al. (8) of the association of *DRB1**0301 with the A allele of rs2040410 and *DQB1**0302 with the C allele of rs7454108. We attempted to exploit this association to identify subjects with the highest-risk HLA genotype for type 1 diabetes and were able to discern a variety of *DR/DQ* genotypes with our two-SNP test. The Taqman assay was specific enough to differentiate *DQB1**0301, -0302, -0303, and -0304 in the DAISY population, identifying DR4-DQ8 alleles exclusively and resulting in excellent specificity when both SNPs were evaluated.

We have found that the use of two SNPs can identify subjects with DR3/4-DQ8 across different populations with greater than 99% accuracy. The assays can be performed in most laboratories, results are available within a relatively short period of time, and the cost of performing these tests is much less than that of traditional HLA genotyping. Therefore, we propose an algorithm for the use of the two-SNP test to identify individuals with a high probability of the DR3/4-DQ8 genotype (Fig. 2) for further standard HLA genotyping. Employing this algorithm makes performing standard testing on all subjects unnecessary and greatly reduces the cost and time required to test a large group of samples. This would allow rapid and relatively inexpensive screening of samples for more detailed HLA genotyping. It is estimated to cost $4 per sample to run the TaqMan assay and $2.05 per sample to perform the restric-

tion digest. Approximately $6 per sample is one-fifth the cost of traditional HLA genotyping at $31.44 per sample.

To examine the applicability of the two-SNP test in the general population, we applied Bayes theorem using the prevalence of DR3/4-DQ8 in the Denver population (2.4%) (11). In a hypothetical population of 1,000 people, 24 would have DR3/4-DQ8. Using our 98.57% sensitivity and 99.67% specificity, 23 of these individuals would be identified as DR3/4-DQ8 and 3 individuals without the genotype would be incorrectly classified as DR3/4-DQ8, giving a positive predictive value of 87.6% (23 of 26). One individual with DR3/4-DQ8 would be incorrectly classified as not and 973 individuals without DR3/4-DQ8 would be correctly classified, giving a negative predictive value of 99.9% (973 of 974). The prevalence of the DR3 allele in Denver is 22.7%, yielding a positive predictive value of 93.7% and a negative predictive value of 99.0% for the rs2040410 A allele. The prevalence of the *DQB1**0302 allele in Denver is 21.9%. The rs7454108 C allele has a positive predictive value of 97.5% and a negative predictive value of 99.7%. Although we recommend further traditional typing to confirm the presence of high-risk genotypes, the positive predictive value of this two-SNP test is quite high in a relatively low frequency population and demonstrates the utility of this two-SNP test.

ACKNOWLEDGMENTS

This research utilizes resources provided by the T1DGC, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the National Institute of Allergy and Infectious Diseases, the National Human Genome Research Institute, the National Institute of Child Health and Human Development, the Juvenile Diabetes Research Foundation International (JDRF), and NIDDK Grant U01 DK062418. J.M.B. is supported by JDRF Early Career Development award 11/2005/15. DAISY was funded by NIDDK, and oral glucose tolerance testing was performed in the pediatric General Clinical Research Center, which was supported by grant RR00069 from the General Clinical Research Centers Program, National Center for Research Resources at the National Institutes of Health. DAISY research was supported by NIDDK Grant DK32493, the Diabetes Endocrine Research Center, Clinical Investigations and Bioinformatics Core Grant P30 DK 57516, and the Children's Diabetes Foundation.

REFERENCES

1. Aly TA, Ide A, Jahromi MM, Barker JM, Fernando MS, Babu SR, Yu L, Miao D, Erlich HA, Fain PR, Barriga KJ, Norris JM, Rewers MJ, Eisenbarth GS: Extreme Genetic Risk for Type 1A Diabetes. *Proc Natl Acad Sci U S A* 103:14074–14079, 2006
2. Bonifacio E, Hummel M, Walter M, Schmid S, Ziegler AG: IDDM1 and multiple family history of type 1 diabetes combine to identify neonates at high risk for type 1 diabetes. *Diabetes Care* 27:2695–2700, 2004
3. Study design of the Trial to Reduce IDDM in the Genetically at Risk (TRIGR). *Pediatr Diabetes* 8:117–137, 2007
4. Kishiyama CM, Chase HP, Barker JM: Prevention strategies for type 1 diabetes. *Rev Endocr Metab Disord* 7:215–224, 2006
5. Bugawan TL, Erlich HA: Rapid typing of HLA-DQB1 DNA polymorphism using nonradioactive oligonucleotide probes and amplified DNA. *Immunogenetics* 33:163–170, 1991
6. Wang JP, Zhou ZG, Lin J, Huang G, Zhang C, Yang L, Yuan Y, Zhou HF, Zhou M, Hou C, Zhou WD, Peng H, Hagopian WA: Islet autoantibodies are associated with HLA-DQ genotypes in Han Chinese patients with type 1 diabetes and their relatives. *Tissue Antigens* 70:369–375, 2007
7. Kiviniemi M, Hermann R, Nurmi J, Ziegler AG, Knip M, Simell O, Veijola R, Lovgren T, Ilonen J: A high-throughput population screening system for the estimation of genetic risk for type 1 diabetes: an application for the TEDDY (the Environmental Determinants of Diabetes in the Young) study. *Diabetes Technol Ther* 9:460–472, 2007
8. de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, Ke X, Monsuur AJ, Whittaker P, Delgado M, Morrison J, Richardson A, Walsh EC, Gao X, Galver L, Hart J, Hafler DA, Pericak-Vance M, Todd JA, Daly MJ, Trowsdale J, Wijmenga C, Vyse TJ, Beck S, Murray SS, Carrington M, Gregory S, Deloukas P, Rioux JD: A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet* 38:1166–1172, 2006
9. Walsh EC, Mather KA, Schaffner SF, Farwell L, Daly MJ, Patterson N, Cullen M, Carrington M, Bugawan TL, Erlich H, Campbell J, Barrett J, Miller K, Thomson G, Lander ES, Rioux JD: An integrated haplotype map of the human major histocompatibility complex. *Am J Hum Genet* 73:580–590, 2003
10. Leslie S, Donnelly P, McVean G: A statistical method for predicting classical HLA alleles from SNP data. *Am J Hum Genet* 82:48–56, 2008
11. Rewers M, Bugawan TL, Norris JM, Blair A, Beaty B, Hoffman M, McDuffie RSJr, Hamman RF, Klingensmith G, Eisenbarth GS, Erlich HA: Newborn screening for HLA markers associated with IDDM: diabetes autoimmunity study in the young (DAISY). *Diabetologia* 39:807–812, 1996
12. Barker JM, Barriga K, Yu L, Miao D, Erlich H, Norris JN, Eisenbarth GS, Rewers M: Prediction of autoantibody positivity and progression to type 1 diabetes: Diabetes Autoimmunity Study in the Young (DAISY). *J Clin Endocrinol Metab* 89:3896–3902, 2004
13. Rewers M, LaPorte RE, King H, Tuomilehto J: Trends in the prevalence and incidence of diabetes: insulin-dependent diabetes mellitus in childhood. *World Health Stat Q* 41:179–189, 1985
14. Patrick SL, Moy CS, LaPorte RE: The world of insulin-dependent diabetes mellitus: what international epidemiologic studies reveal about the etiology and natural history of IDDM. *Diabetes Metab Rev* 5:571–578, 1989
15. Onkamo P, Vaananen S, Karvonen M, Tuomilehto J: Worldwide increase in incidence of Type I diabetes–the analysis of the data on published incidence trends. *Diabetologia* 42:1395–1403, 1999
16. 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
17. Karvonen M, Viik-Kajander M, Moltchanova E, Libman I, LaPorte R, Tuomilehto J: Incidence of childhood type 1 diabetes worldwide. Diabetes Mondiale (DiaMond) Project Group. *Diabetes Care* 23:1516–1526, 2000
18. Eisenbarth GS: Type I diabetes mellitus: a chronic autoimmune disease. *N Engl J Med* 314:1360–1368, 1986