Brief Genetics Report

Association Analysis of the Lymphocyte-Specific Protein Tyrosine Kinase (LCK) Gene in Type 1 Diabetes

John S. Hulme,[1] Bryan J. Barratt,[1] Rebecca C.J. Twells,[1] Jason D. Cooper,[1] Chris E. Lowe,[1] Joanna M.M. Howson,[1] Alex C. Lam,[1] Luc J. Smink,[1] David A. Savage,[2] Dag E. Undlien,[3] Cristian Guja,[4] Constantin Ionescu-Tîrgovişte,[4] Eva Tuomilehto-Wolf,[5] Jaakko Tuomilehto,[5,6] and John A. Todd[1]

Prior data associating the expression of lymphocyte-specific protein tyrosine kinase (LCK) with type 1 diabetes, its critical function in lymphocytes, and the linkage of the region to diabetes in the nonobese diabetic (NOD) mouse model make *LCK* a premier candidate for a susceptibility gene. Resequencing of *LCK* in 32 individuals detected seven single nucleotide polymorphisms (SNPs) with allele frequencies >3%, including four common SNPs previously reported. These and six other SNPs from dbSNP were genotyped in a two-stage strategy using 2,430 families and were all shown not to be significantly associated with type 1 diabetes. We conclude that a major role for the common LCK polymorphisms in type 1 diabetes is unlikely. However, we cannot rule out the possibility of there being a causal variant outside the exonic, intronic, and untranslated regions studied. *Diabetes* 53:2479–2482, 2004

Type 1 diabetes is believed to arise from the specific autoimmune destruction of the insulin-producing islet cells of the pancreas by autoreactive T-cells. The disease is mediated by the interaction of many genes and environmental factors (1). To date three genetic loci have been confirmed, the HLA region (chromosome 6p21), the insulin gene region (chromosome 11p15), and *CTLA4* (chromosome 2q33) (2,3).

In the NOD mouse model of type 1 diabetes, the susceptibility region *Idd9.1* on chromosome 4 includes the lymphocyte-specific protein tyrosine kinase gene (*LCK*). LCK has been associated with T-cell proliferative hyporesponsiveness in NOD mice. Reduced recruitment of CD4-associated LCK to the T-cell receptor complex results in deficient coupling of the T-cell receptor complex to downstream signaling events (4). It has also been reported by Nervi et al. (5) that hyporesponsiveness of T-cells in patients with type 1 diabetes correlates with reduced levels of LCK in resting T-cells. In addition, Nervi et al. (6) found no association between *LCK* polymorphisms and protein level in type 1 diabetic patients. However, as the authors acknowledged, this study was statistically underpowered to detect the weak or moderate genetic associations expected in a common multifactorial disease such as type 1 diabetes and the estimates of protein levels had wide CIs.

Human *LCK* is located on chromosome 1p35 (Ensembl 15.33.1: sequence AL121991.50.1.61515). It has 13 exons and two promoters with different 5′ untranslated regions (UTRs) active at different stages of T-cell development. The *LCK* proximal promoter is only active in thymocytes, whereas the distal promoter is active at all stages of T-cell development (7). The long form of *LCK* (ENST00000328410) includes the distal promoter and has an untranslated exon 1 with a first intronic interval of 22,953 bp (exon 1-2), whereas the short form includes the proximal promoter (Fig. 1). Translation of both forms start at the first ATG codon in exon 2. A putative untranslated exon 1′ in the short form of *LCK* (ENST0000033070) is supported by a single cDNA sequence (BC013200) derived from lymphoma tissue. However, further homology searches against dbEST (http://www.ncbi.nlm.nih.gov/dbEST) did not find any additional matches that would support exon 1′ in nonlymphoma tissues.

In this study, we investigated whether there is statistical support for a type 1 diabetes locus associated with the common *LCK* polymorphisms, using a large family collection. *LCK* was resequenced in 32 U.K. Caucasian individuals with type 1 diabetes, including all the exons (1–13), the short introns (2–8 and 10–12), 4 kb of sequence 5′ of exon 1 (including the distal promoter), 2 kb of sequence 5′ of exon 2 (including the proximal promoter and the putative exon 1′), and 2 kb of sequence 3′ of exon 13, including the 3′ UTR (Fig. 1). Our resequencing included all of the single nucleotide polymorphism (SNP) locations

From the [1]Juvenile Diabetes Research Foundation (JDRF)/Wellcome Trust (WT) Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, U.K.; the [2]Department of Medical Genetics, Queen's University Belfast, Belfast City Hospital, Belfast, Northern Ireland; the [3]Institute of Medical Genetics, Ulleval University Hospital, University of Oslo, Oslo, Norway; the [4]Clinic of Diabetes, Institute of Diabetes, Nutrition, and Metabolic Diseases 'N. Paulescu', Bucharest, Romania; the [5]Diabetes and Genetic Epidemiology Unit, National Public Health Institute, University of Helsinki, Helsinki, Finland; and the [6]Department of Public Health, University of Helsinki, Helsinki, Finland.

Address correspondence and reprint requests to Prof. John A. Todd, JDRF/WT Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 2XY, U.K. E-mail: john.todd@cimr.cam.ac.uk.

Received for publication 8 October 2003 and accepted in revised form 26 May 2004.

Additional information for this article can be found in an online appendix at http://diabetes.diabetesjournals.org.

EMBL, European Molecular Biology Laboratory; LCK, lymphocyte-specific protein tyrosine kinase; SNP, single nucleotide polymorphism; UTR, untranslated region.

© 2004 by the American Diabetes Association.

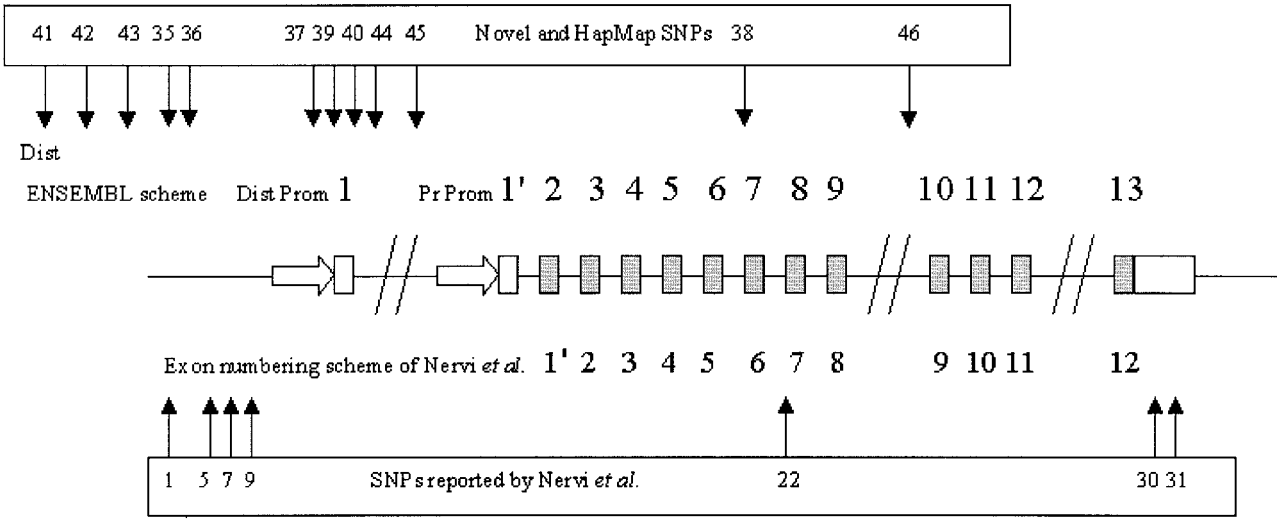

FIG. 1. Genomic organization of *LCK*. The two promoters (distal and proximal) are shown as white arrows. Exons are shown as boxes, the gray shading indicating the translated region, while the white boxes are UTRs. The // regions represent intronic gaps (of 19.9, 2.8, and 4.4 kb from left to right) in our resequencing. The SNPs are shown as black arrows, numbered according to Nervi et al. (6) and extended to include newly identified novel SNPs 35–40 and the HapMap SNPs 41–46. The upper annotation is according to Ensembl 15.33.1 and matches all available sequence information. The lower annotation is drawn from Nervi et al. (6) and is necessary to interpret previously published results, owing to inconsistencies between the exon numbering schemes.

reported by Nervi et al. (6). By using a sequencing panel of 32 affected individuals, we had an 88% probability of detecting SNPs with minor allele frequencies of 3.3 and 96% probability of detecting minor alleles of 5% frequency.

Resequencing detected the four common SNPs identified in the previous report (6), SNPs 1, 7, 9, and 30 (Table 1). We also detected two SNPs (5 and 22), previously observed to be rare (~1%), at a frequency >3% in our panel and an additional novel SNP at a frequency >3% (SNP 35). Two further novel variants, SNPs 36 and 37, were found at frequencies of ~1.5%.

We did not, however, detect the remaining 28 previously reported rare polymorphisms (6) in our sequencing panel of 32 individuals. Therefore, we resequenced four regions in a panel of 96 additional individuals with type 1 diabetes to increase our power to identify rare SNPs (85% probability of detecting SNPs with minor allele frequencies of 1%) and to confirm the allele frequencies of the rare novel SNPs 36 and 37. The first region (~1.8 kb) encompassed exons 3, 4, 5, 6, 7, and 8 (including introns) and, hence, included the previously reported locations of the rare exonic SNPs observed in the study by Nervi et al. (6) and SNP 22, which had already been observed in our panel of 32 individuals. The second region (525 bp) included all five of the rare 3′ SNPs seen in the previous study and the common SNP 30 (3′ UTR), whereas the third region included SNP 37 and the fourth region included SNP 36.

In the first region (~1.8 kb), the sequencing of 96 additional individuals did not detect any of the rare exonic SNPs previously reported, but we were able to detect a new rare SNP in exon 7, SNP 38, which causes a Gly to Ser, nonsynonymous, nonconservative amino-acid coding

TABLE 1
SNP data, including position and frequency data from Nervi et al. (6)

						Minor allele frequency (%)			
						Nervi et al.		DIL	
Nervi et al. SNP no.*	DIL SNP no.	Nucleotide variation	dbSNP no.	Position (Nervi et al.)	Location (Ensembl 15.33.1)	Control subjects	Patients	32x panel	96x panel
1	4670	A to G	rs2291063	173	5′ exon 1	5.36	2.41	3.33	—
35	4671	C to T	ss16336875	215	5′ exon 1	0	0	3.33	—
36	4672	G to T	ss16336876	284	5′ exon 1	0	0	1.67	1.63
5	4999	C to T	ss16336879	2045	5′ exon 1	0.6	1.23	3.13	—
7	4664	G to T	ss16336869	2615	5′ exon 1	20	22.73	12.5	—
9	4665	G to A	ss16336870	2795	5′ exon 1	1.81	5.36	3.13	—
37	4666	G to A	ss16336871	3802	Dist prom	0	0	1.56	0.56
39	7791	G to A	ss22970416	3858	Dist prom	0	0	0	0.56
40	7792	G to A	ss22970417	3960	Intron 1	0	0	0	0.56
38	4998	G to A	ss16336878	7131†	Exon 7	0	0	0	0.56
22	4668	G to T	ss16336872	7357†	Intron 7	1.06	1.08	3.13	2.78
30	4669	A to G	rs1042546	17010†	3′ UTR	10.63	10.67	6.25	9.47
31	4997	A to G	ss16336877	17027†	3′ UTR	0	0.56	0	0.56

*SNP numbering system of Nervi et al. (6), retained and extended to include SNPs 35–40. †Previously published data (NCBI: BN000073) contained a gap of 21,578 bp between positions 4086 and 4187. (*LCK*: locus link no. 3932). DIL, Diabetes and Inflammation Laboratory, Cambridge, U.K.

TABLE 2
Genotyping data for stage 1 and stages 1 and 2 combined

Nervi et al. SNP no.*	DIL SNP no.	dbSNP no.	HapMap minor allele frequency (%)	Parental minor allele frequency (%)		TDT *P* values (corrected *P*)	
				Stage 1 (722 families)	Stages 1 and 2 (2,430 families)	Stage 1 (722 families)	Stages 1 and 2 (2,430 families)
41	7895	rs747020	6	5.11	—	0.47	—
42	7893	rs3795428	8	6.52	—	0.57	—
43	8551	rs669538	28	29.46	—	0.27	—
1	4670	rs2291063	—	4.27	—	0.27	—
35	4671	ss16336875	—	0.52	—	1.00	—
5	4999	ss16336879	—	1.52	—	0.91	—
7	4664	ss16336869	—	17.25	16.38	0.11	0.54 (0.15)
9	4665	ss16336870	—	3.54	—	1.00	—
44	8850	rs6422595	3	3.81	—	0.52	—
45	7892	rs1004420	11	15.31	—	0.77	—
22	4668	ss16336872	—	2.22	—	0.24	—
46	7894	rs695161	48	47.74	—	0.84	—
30	4669	rs1042546	—	7.95	6.28	0.02	0.24 (0.10)

*SNP numbering system of Nervi et al. (6), retained and extended to include HapMap SNPs 41, 42, 43, 44, 45, and 46. Only SNPs with minor allele frequencies >3% in the SNP discovery phase or the HapMap database were genotyped in stage 1. Only SNPs 7 and 30, which had a transmission/disequilibrium test (TDT) *P* value <0.2 (in stage 1), were genotyped in stage 2. The data for stages 1 and 2 were combined and the TDT test performed on the entire dataset.

change. However, the low minor allele frequency of 0.56% for SNP 38 was considered too low to obtain a reliable result for disease association in our current family collection. The intronic SNP 22 was detected in the panel of 96 individuals (2.78%) at a frequency similar to that seen in the panel of 32 individuals (3.13%).

In the second region (525 bp), we were able to detect the rare SNP 31 at the same minor allele frequency (0.56%) as previously reported (6). We were also able to detect the 3′ UTR SNP 30 at a frequency similar to that previously reported (~10%). However, we did not detect the remaining three rare SNPs in this region. The failure to detect 11 of the 12 rare SNPs in these two regions, 8 of which had been reported in patients with type 1 diabetes, may result from differing allele frequencies between our U.K. panel and the French panel used by Nervi et al. (6). In the third region, which included SNP 37, we detected two further novel SNPs, 39 and 40, both at minor allele frequencies of 0.56% in our panel of 96 individuals. The fourth region confirmed the allele frequency of SNP 36, but did not detect any further rare SNPs (Table 1).

A two-stage genotyping strategy was used for this study, incorporating the concept of stopping for futility after the first stage. Therefore, SNPs were only genotyped in stage 2 if results from stage 1 offered the possibility of a significant overall result (8). Stage 1 comprises 722 multiplex families (454 U.K. and 268 U.S.; providing 1,340 parent-child trios) who were genotyped and tested for association using the transmission-disequilibrium test (9). Stage 2 comprises 1,708 mostly simplex families (926 Finnish, 330 U.K., 233 Romanian, 159 Norwegian, and 60 U.S.; providing 1,733 parent-child trios) who were only genotyped if the *P* value for stage 1 association was ≤0.20. The two-stage design results in only a small loss of power compared with genotyping stages 1 and 2 together (online appendix 1 [available from http://diabetes.diabetesjournals.org]).

Seven SNPs with minor allele frequencies >3% in the SNP discovery phase (Table 1) were genotyped in stage 1. Only SNPs 7 and 30 had *P* values ≤0.20 and, consequently, were genotyped in stage 2 families (Table 2). Association was then tested for these two SNPs in the entire family collection (stages 1 and 2 combined), using a significance target of $P = 0.001$. As we used a two-stage genotyping strategy, the *P* values from the association tests using the combined data of stages 1 and 2 should be corrected for the possibility of stopping for futility after the first stage. Since a positive result must survive two statistical tests, the *P* values are reduced by the correction (8). The results for SNP 7 (uncorrected $P = 0.54$, corrected $P = 0.15$) and SNP 30 (uncorrected $P = 0.24$, corrected $P = 0.10$), together with the stage 1 results, show no evidence of association for the seven SNPs genotyped (Table 2).

We cannot exclude the possible effects of long-range gene expression regulatory elements, which might lie outside the areas resequenced in our study. To address this issue, we used the National Center for Biotechnology Information (NCBI) database (Build 117) to identify 53 dbSNPs in a 60-kb region containing *LCK*, between dbSNPs rs747020 (5′) and rs7543692 (3′), on chromosome 1 (http://www.ncbi.nlm.nih.gov). Ten of these dbSNPs are in the resequenced regions, of which only SNPs 1 (rs2291063) and 30 (rs1042546) were detected in our study. This low concordance (20%), combined with the number of dbSNPs (24 in the large introns and the 19 in the extreme 5′ and 3′ regions), would make genotyping all of these SNPs inefficient. However, the region does contain seven SNPs published as part of the International HapMap Project, and six of these were polymorphic, with minor allele frequencies ≥3% in a panel of 30 trios (http://www.hapmap.org). These six SNPs, rs747020 (5′), rs3795428 (5′), rs669538 (5′), rs6425795 (intron 1), rs1004420 (intron 1), and rs695161 (intron 9), were genotyped in stage 1 families and tested for association. All six SNPs have stage 1 transmission-disequilibrium test *P* values >0.20 and were, therefore, not genotyped in stage 2 families (Table 2).

None of the 13 SNPs genotyped in this study show association, suggesting that, for the samples and gene regions studied here, it is unlikely that common variants of

LCK have a significant role in susceptibility to type 1 diabetes. When a more comprehensive SNP map becomes available, it will, however, be possible to test the alternative hypothesis, which suggests that a distant regulatory variant of *LCK* exists and modifies type 1 diabetes susceptibility.

RESEARCH DESIGN AND METHODS

All families were Caucasian and of European descent, with two parents and at least one affected child. Stage 1 families consisted of 722 multiplex families from the U.K. British Diabetic Association Warren 1 repository (10) and the U.S. Human Biological Data Interchange (11). Stage 2 families consisted of 1,708 simplex and multiplex families from the U.K., Belfast (12), Norway (13), Finland (14), Romania (15), and the further U.S. Human Biological Data Interchange (11). The Cambridge Local Research Ethics Committee gave full ethical approval, and informed consent was obtained for the collection and use of these DNA samples from all subjects.

Computational analysis. The gene structures of the short and long forms of *LCK* were determined using Blat (16). Full-length human mRNAs with the European Molecular Biology Laboratory (EMBL)/GenBank accession of BC013200 and X13529, respectively, were used (EMBL release 33). The output was converted into an ACeDB format (http://www.acedb.org), and the gene structures were checked in a lightweight ACeDB interface. Confirmed gene structures were exported into a Gbrowse viewer (http://www.gmod.org) (17) via General Feature Format (GFF) (http://www.sanger.ac.uk). Identified polymorphisms were uploaded into an in-house database directly from Gap4 files. Mapping against the human genome assembly was performed using an in-house program, which utilized BLAST (basic local alignment search tool; http://www.ncbi.nlm.nih.gov), Ensembl Perl API package (http://www.ensembl.org), and BioPerl (http://bioperl.org). Mapping data were exported into Gbrowse for visualization via GFF. For further details of genome informatics methods, see Burren et al. (18).

Polymorphism identification. DNA aliquots (20 ng) from 32 or 96 individuals with type 1 diabetes were PCR amplified as ~1,500-bp overlapping amplicons, and 2 μl of the products were sequenced with internal primer pairs to generate three ~500-bp overlapping sequences within these amplicons. Sequencing reactions were performed using the ABI Prism Big Dye terminator kit according to manufacturer's instructions (Applied Biosystems, Foster City, CA) and products electrophoresed on an ABI Prism 3700 DNA Analyzer. SNPs were identified using the Staden package (http://www.mrc-lmb.cam.ac.uk/pubseq/staden_home.html).

Genotyping. SNPs were genotyped by either TaqMan (Applied Biosystems) or Biplex Invader (Third Wave Technologies, Madison, WI) assays, according to the manufacturer's instructions, with an overall 97% success rate of scorable genotypes.

Statistical analysis. Statistical analyses were performed within the Stata package (http://www.stata.com), making specific use of the Genassoc routines (http://www-gene.cimr.cam.ac.uk/clayton/software/stata). Correction for two-stage analysis (8) was performed in the *R* package (http://cran.r-project.org). Pairwise values of D' and r^2 were calculated for all 13 SNPs genotyped in stage 1 and are shown in online appendix 2.

ACKNOWLEDGMENTS

We thank the Juvenile Diabetes Research Foundation, the Wellcome Trust, Novo Nordisk, the Novo Nordisk Foundation, the Academy of Finland, and the Sigrid Juselius Foundation for funding support.

The Norwegian Study Group for Childhood Diabetes was responsible for the collection of the Norwegian families analyzed in this study and the Human Biological Data Interchange and Diabetes U.K. for the U.S. and U.K. collections, respectively. We thank Sarah Nutland and Helen Rance for DNA preparation.

REFERENCES

1. Todd JA: From genome to aetiology in a multifactorial disease, type 1 diabetes. *Bioessays* 21:164–174, 1999
2. Reijonen H, Concannon P: Genetics of type 1 diabetes. In *Joslin's Diabetes Mellitus*. 14th ed. Kahn RC, Weir GC, Eds. Philadelphia, Lippincott, Williams and Wilkins, 2003
3. 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, Allahabadia A, Nithiyananthan R, Tuomilehto-Wolf E, Tuomilehto J, Bingley P, Gillespie KM, Undlien DE, Ronningen KS, Guja C, Ionescu-Tirgovişte 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
4. Zhang J, Salojin K, Delovitch TL: Sequestration of CD4-associated Lck from the TCR complex may elicit T cell hyporesponsiveness in nonobese diabetic mice. *J Immunol* 160:1148–1157, 1998
5. Nervi S, Atlan-Gepner C, Kahn-Perles B, Lecine P, Vialettes B, Imbert J, Naquet P: Specific deficiency of p56lck expression in T lymphocytes from type 1 diabetic patients. *J Immunol* 165:5874–5883, 2000
6. Nervi S, Nicodeme S, Gartioux C, Atlan C, Lathrop M, Reviron D, Naquet P, Matsuda F, Imbert J, Vialettes B: No association between *lck* gene polymorphisms and protein level in type 1 diabetes. *Diabetes* 51:3326–3330, 2002
7. Voronova AF, Adler HT, Sefton BM: Two lck transcripts containing different 5′ untranslated regions are present in T cells. *Mol Cell Biol* 7:4407–4413, 1987
8. Lowe C, Cooper JD, Chapman JM, Barratt BJ, Twells RCJ, Green EA, Savage DA, Guja C, Ionescu-Tirgovi°te C, Tuomilehto-Wolf E, Tuomilehto J, Todd JA, Clayton DG: Cost-effective analysis of candidate genes using htSNPs: a staged approach. *Genes Immun* 5:301–305, 2004
9. Martin ER, Kaplin NL, Weir BS: Tests for linkage and association in nuclear families. *Am J Hum Genet* 61:439–448, 1997
10. Bain SC, Todd JA, Barnett AH: The British Diabetic Association–Warren Repository. *Autoimmunity* 7:83–85, 1990
11. Lernmark A, Ducat L, Eisenbarth G, Ott J, Permutt MA, Rubenstein P, Spielman R: Family cell lines available for research. *Am J Hum Genet* 47:1028–1030, 1990
12. Patterson CC, Carson DJ, Hadden DR: Epidemiology of childhood IDDM in Northern Ireland 1989–1994: low incidence in areas with highest population density and most household crowding: Northern Ireland Diabetes Study Group. *Diabetologia* 39:1063–1069, 1996
13. Undlien DE, Akselsen HE, Joner G, Dahl-Jorgensen K, Aagenaes 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
14. Tuomilehto J, Lounamaa R, Tuomilehto-Wolf E, Reunanen A, Virtala E, Kaprio EA, Akerblom HK: Epidemiology of childhood diabetes mellitus in Finland: background of a nationwide study of type 1 (insulin-dependent) diabetes mellitus: the Childhood Diabetes in Finland (DiMe) Study Group. *Diabetologia* 35:70–76, 1992
15. Ionescu-Tirgovi°te 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
16. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: The human genome browser at UCSC. *Genome Res* 12:996–1006, 2002
17. Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, Lewis S: The generic genome browser: a building block for a model organism system database. *Genome Res* 12:1599–1610, 2002
18. Burren OS, Healy BC, Lam AC, Schuilenburg H, Dolman GE, Everett VH, Laneri D, Nutland S, Rance HE, Payne F, Smyth D, Lowe C, Barratt BJ, Twells RCJ, Rainbow DB, Wicker LS, Todd JA, Walker NM, Smink LJ: Development of an integrated genome informatics, data management and workflow infrastructure: a toolbox for the study of complex disease genetics. *Hum Genomics* 1:98–109, 2004