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-Tîrgovis¸te,4 Eva Tuomilehto-Wolf,5 Jaakko Tuomilehto,5,6 and John A. Todd1

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

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

<table>
<thead>
<tr>
<th>Nervi et al. SNP no.*</th>
<th>DIL SNP no.</th>
<th>Nucleotide variation</th>
<th>dbsNP no.</th>
<th>Position (Nervi et al.)</th>
<th>Location (Ensembl 15.33.1)</th>
<th>Minor allele frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Control subjects</td>
<td>Patients</td>
</tr>
<tr>
<td>1</td>
<td>4670</td>
<td>A to G</td>
<td>rs2291063</td>
<td>173</td>
<td>5′ exon 1</td>
<td>5.36</td>
</tr>
<tr>
<td>35</td>
<td>4671</td>
<td>G to T</td>
<td>ss16336875</td>
<td>215</td>
<td>5′ exon 1</td>
<td>0</td>
</tr>
<tr>
<td>36</td>
<td>4672</td>
<td>G to T</td>
<td>ss16336876</td>
<td>284</td>
<td>5′ exon 1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>4699</td>
<td>C to G</td>
<td>ss16336879</td>
<td>2945</td>
<td>5′ exon 1</td>
<td>0.6</td>
</tr>
<tr>
<td>7</td>
<td>4664</td>
<td>G to T</td>
<td>ss16336869</td>
<td>2615</td>
<td>5′ exon 1</td>
<td>20</td>
</tr>
<tr>
<td>9</td>
<td>4665</td>
<td>G to A</td>
<td>ss16336870</td>
<td>2795</td>
<td>5′ exon 1</td>
<td>1.81</td>
</tr>
<tr>
<td>37</td>
<td>4666</td>
<td>G to A</td>
<td>ss16336871</td>
<td>3802</td>
<td>Dist prom</td>
<td>0</td>
</tr>
<tr>
<td>39</td>
<td>7791</td>
<td>G to A</td>
<td>ss22970416</td>
<td>3858</td>
<td>Dist prom</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>7792</td>
<td>G to A</td>
<td>ss22970417</td>
<td>3960</td>
<td>Intron 1</td>
<td>0</td>
</tr>
<tr>
<td>38</td>
<td>4998</td>
<td>G to A</td>
<td>ss16336878</td>
<td>7131†</td>
<td>Exon 7</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>4668</td>
<td>G to T</td>
<td>ss16336872</td>
<td>7357†</td>
<td>Intron 7</td>
<td>1.06</td>
</tr>
<tr>
<td>30</td>
<td>4669</td>
<td>A to G</td>
<td>rs1042546</td>
<td>17010†</td>
<td>3′ UTR</td>
<td>10.63</td>
</tr>
<tr>
<td>31</td>
<td>4997</td>
<td>A to G</td>
<td>ss16336877</td>
<td>17027†</td>
<td>3′ UTR</td>
<td>0</td>
</tr>
</tbody>
</table>

*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.
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 1,782 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 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’), rs695161, 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 1,700 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 (16). 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 BC013290 and X13529, respectively, were used (EMBL release 33). The **LCK** gene 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/pubs/exp/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-genome.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 r² 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


