A human type 1 diabetes susceptibility locus maps to chromosome 21q22.3

Patrick Concannon1,2, Suna Onengut-Gumuscu2,3, John A. Todd4, Deborah J. Smyth4, Flemming Pociot5, Regine Bergholdt5, Beena Akolkar6, Henry A. Erlich7, Joan E. Hilner8, Cécile Julier9, Grant Morahan10, Jørn Nerup5, Concepcion R. Nierras11, Wei-Min Chen2,12, Stephen S. Rich2 and the Type 1 Diabetes Genetics Consortium

1Department of Biochemistry and Molecular Genetics, 2Center for Public Health Genomics, and 3Department of Medicine, Division of Endocrinology and Metabolism, University of Virginia, Charlottesville, Virginia, USA
4Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK
5Steno Diabetes Center, Gentofte, Denmark
6Division of Diabetes, Endocrinology, and Metabolic Diseases, The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institutes of Health, Bethesda, MD
7Roche Molecular Systems, Alameda, CA, USA
8Division of Public Health Sciences, Wake Forest University Health Sciences, Winston-Salem, NC, USA
9Inserm U730 et CEA, Institut de Génomique Centre National de Génotypage, Evry, France
10Centre for Diabetes Research, The Western Australian Institute for Medical Research, and Centre for Medical Research, University of Western Australia, Australia
11Juvenile Diabetes Research Foundation, New York, NY, USA
12Department of Public Health Sciences, Division of Biostatistics and Epidemiology, University of Virginia, Charlottesville, VA, USA

Correspondence:
Patrick Concannon
Department of Biochemistry and Molecular Genetics
University of Virginia, Box 800733, Charlottesville, VA 22908 USA

Received 6 June 2008 and accepted 14 July 2008.
ABSTRACT

Objective. The Type 1 Diabetes Genetics Consortium (T1DGC) has assembled and genotyped a large collection of multiplex families for the purpose of mapping genomic regions linked to type 1 diabetes. In the current study, we tested for evidence of loci associated with type 1 diabetes utilizing genome-wide linkage scan data and family-based association methods.

Research Design and Methods: A total of 2,496 multiplex type 1 diabetes families were genotyped with a panel of 6,090 SNPs. Evidence of association to disease was evaluated by the pedigree disequilibrium test (PDT). Significant results were followed up by genotyping and analyses in two independent sets of samples: 2,214 parent-affected child trio families and a panel of 7,721 cases and 9,679 controls.

Results: Three of the SNPs most strongly associated with type 1 diabetes localized to previously identified type 1 diabetes risk loci: INS, IFIH1, and KIAA0350. A fourth strongly associated SNP, rs876498 (P=1.0×10^{-4}) occurred in the sixth intron of the UBASH3A locus at chromosome 21q22.3. Support for this disease association was obtained in two additional independent sample sets, type 1 diabetes families (OR=1.06, 95%CI=1.00-1.11, P=0.023) and cases and controls (OR=1.14, 95%CI=1.09-1.19, P=7.5x10^{-8}).

Conclusions: The T1DGC 6K SNP scan and follow-up studies reported here confirm previously reported type 1 diabetes associations at INS, IFIH1 and KIAA0350 and identify an additional disease association on chromosome 21q22.3 in the UBASH3A locus (OR=1.10, 95% CI=1.07-1.13, P=4.4x10^{-12}). This gene and its flanking regions are now validated targets for further re-sequencing, genotyping and functional studies in type 1 diabetes.
Genome-wide linkage and association studies focused on candidate genes or interval mapping in type 1 diabetes have previously identified a handful of risk loci that have been confirmed by multiple replications. Of these, the major risk factor(s) reside within the HLA region on chromosome 6p21. Other, non-HLA, type 1 diabetes loci have comparatively small effects on disease risk relative to HLA but comparable effect sizes to risk loci identified in other common disorders. These include the insulin (INS) locus, where variation is thought to impact the transcription and expression of insulin, modulating thymic selection of T cells specific for this autoantigen (1;2). Variants at the cytotoxic T lymphocyte antigen (CTLA4) locus are implicated in type 1 diabetes risk, potentially by altering the production of differentially spliced products from the locus which affects T cell activation (3;4). SNPs in the IL-2 receptor alpha (IL2RA) region, that encodes one chain of the heterodimeric IL-2 cytokine receptor, are associated with type 1 diabetes (5). Finally, a non-synonymous coding region SNP in the protein tyrosine phosphatase, non-receptor type 22 (PTPN22) gene has been associated with type 1 diabetes as well as a number of other autoimmune disorders (6-9). This SNP, which predicts an Arg to Trp substitution at position 620, increases the activity of the PTPN22 encoded phosphatase, Lyp, resulting in hypo responsiveness of T cells (10;11).

Recently, the approach of genome-wide association scanning has been applied to type 1 diabetes (12-14). The combination of high-density SNP coverage, and large sample sizes for both initial screening and follow-up genotyping in these studies, has both confirmed known type 1 diabetes loci and identified a number of novel candidate genes/regions. New associations with type 1 diabetes have been identified and replicated for SNPs in two regions on chromosome 12, 12q24 near C12orf30 and 12q13 near erythroblastic leukemia viral oncogene homolog 3 (ERBB3), a region on 16p13 near C-type lectin domain family 16, member A (CLEC16A/KIAA0350), a region on 18p11 near protein tyrosine phosphatase, non-receptor type 2 (PTPN2), and a region on 2q24 near interferon induced with helicase C domain 1 (IFIH1) (15;16). Detailed information about each of these type 1 diabetes associated loci can be found in T1DBase (www.t1dbase.org).

Most of the non-HLA type 1 diabetes risk loci are characterized by odds ratios in the 1.15 to 1.3 range and, despite the recent increase in the number of such loci identified and confirmed, it is unlikely that these known loci can account for all of the genetic risk for type 1 diabetes (15). Therefore, in the current study we took advantage of a recently completed genome-wide scan for linkage in type 1 diabetes multiplex families to explore whether alleles at any of the 6,090 SNPs genotyped displayed evidence of association with type 1 diabetes.

RESEARCH DESIGN AND METHODS

Subjects. Families utilized for initial screening were ascertained from 9 different sources, the four T1DGC networks, Asia-Pacific (N=228), Europe (N=585), North America (N=370) and UK (N=124), the Diabetes UK Warren Repository (N=429), the Human Biological Data Interchange repository (N=424), the Joslin Diabetes Center (N=112), the Steno Diabetes Center (N=146), and Sardinia (N=78). Details on recruitment sites and the numbers of families contributed for each of the T1DGC networks are accessible from the public side of the T1DGC website (www.t1dgc.org). Minimum entry criterion for families included the presence of at least one affected sibling pair; availability of one or both biological parents was preferred. Eligibility requirements for...
affected individuals included documented type 1 diabetes with onset at less than 35 years of age, insulin use within 6 months of diagnosis, and the absence of any concomitant disease or disorder associated with diabetes. In total, 2,496 families were genotyped. Both parents were available for 72.9% of families; 18.8% had only a single parent. Most families (91.7%) had exactly two affected full siblings; 4.1% had three and 0.3% had four or more. The remaining families had pairs that were either incomplete or were half siblings.

The family replication set of samples consisted of 2,214 parent-affected child trio families derived from two sources. The GoKinD study of diabetic nephropathy contributed 578 families. Affected offspring in these families had type 1 diabetes diagnosed before age 31, had initiated insulin therapy within one year of diagnosis, and had continued insulin treatment uninterrupted since initiation. The remaining 1,636 families were ascertained through the Danish Society of Childhood Diabetes and were all of Danish Caucasoid origin. Probands in these families were diagnosed with type 1 diabetes before 15 years of age and had continued insulin treatment since diagnosis. The case-control replication set consisted of 17,400 subjects, 7,721 type 1 diabetes cases recruited as part of the JDRF/Welcome Trust British case collection and 9,679 controls from the British 1958 Birth Cohort and the Wellcome Trust Case Control Blood Service (13;17).

Genotyping. Families included in the original genome-wide linkage scan were all genotyped at the Center for Inherited Disease Research (CIDR) using the Illumina Human Linkage-12 Genotyping Beadchip consisting of 6,090 SNPs. Genotyping of SNP rs876498 in replication samples was performed using an Eclipse genotyping assay from Nanogen. As a quality control for accuracy of this assay, a blinded re-sampling and genotyping of 366 type 1 diabetes families from the initial genome-wide scan was performed.

Statistical analysis. Prior to statistical genetic analyses, genotype data were evaluated for Mendelian errors using PedCheck (18). PREST (19) was used to estimate the likelihood of each specified relationship in pedigrees given the genotyping information. Based on these analyses, 43 families were removed from further analysis due to the presence of a duplicate family, a duplicate sample within the family, non-resolvable family errors or missing genotype information. Allelic association with type 1 diabetes within multiplex families was assessed using the pedigree disequilibrium test (PDT) (20), which provides a valid test of association in the presence of linkage in families with multiple affected individuals. The PDT examines the discrepancy of alleles transmitted from heterozygote parents to affected offspring, as well as the allele difference between affected and unaffected siblings. Both PDT-SUM and PDT-AVE tests (21) were carried out. Since these tests are asymptotically equivalent, the PDT-AVE test in which counts of trios and pairs for each pedigree are weighted is primarily reported here. PLINK (v.1.02) (22) was used to assess allelic association to type 1 diabetes in trio families via the transmission/disequilibrium test (23) and in case-control samples. A logistic regression model was used to estimate effect sizes, with familial correlations being taken into account by the GEE method (24).

To combine results of the original genome scan and follow-up studies a weighted z score–based fixed effects meta-analysis method was used. In brief, a z statistic summarizing the magnitude of the P value for association and direction of effect was generated for each study. An overall z statistic was then computed as a weighted average of the individual statistics, and a corresponding P value for that statistic was computed. The weights were proportional to the square root of the number of individuals in each study and scaled such that the squared weights
summed to 1. For the meta-analysis of the effect size, the inverse variance was used as weights for each study.

RESULTS

After data-cleaning, PDT was performed on the data from 5,943 markers genotyped in the family collection. Table 1 lists all SNPs with \( P < 0.001 \) in this analysis. Only a single marker, rs1004446, displayed evidence of association with type 1 diabetes at a genome-wide significance level (\( P = 2.5 \times 10^{-9} \)). This SNP is located in the \textit{INS-IGF2} region on chromosome 11p15 and has previously been reported to be associated with type 1 diabetes (14). Indeed, among the 16 SNPs listed in Table 1, five corresponded to previously confirmed T1D loci identified through genome-wide scanning: rs1004446 (\textit{INS-IGF2}), a non-synonymous coding SNP rs1990760 (\textit{IFIH1}), rs887864 (intron 18 of \textit{CLEC16A/KIAA0350}), and two SNPs, rs169679 and rs1011094, which flank the \textit{HLA} region.

Given that the top three most associated SNPs ranked by statistical significance in Table 1 correspond to confirmed type 1 diabetes risk loci, we examined the fourth-ranked SNP, rs876498, for which there was also substantial evidence of association with type 1 diabetes (\( P=1.0 \times 10^{-4} \) for PDT-AVE test and \( P=1.0 \times 10^{-5} \) for PDT-SUM test). Further impetus for follow-up study of this SNP came from its location within the sixth intron of the gene \textit{UBASH3A}, which is expressed predominantly in T cells and is involved in the regulation of signaling from the antigen receptor (25). The putative disease-associated allele was relatively common with a minor allele frequency (MAF) of 0.45 in unaffected founders. Blinded validation genotyping was carried out in 366 of the original 2,496 T1DGC families by a second methodology to confirm the initial finding. A concordance rate of 100% was observed between the original and validation study. In order to confirm and extend the initial finding, rs876498 was genotyped in additional independent populations, both family-based and case-control (Table 2). Overall, there was highly significant evidence of association between the minor allele at rs876498 and type 1 diabetes (OR = 1.10, 95% CI 1.07-1.13, \( P = 4.4 \times 10^{-12} \)).

DISCUSSION

In the current study, we took advantage of available SNP genotyping data from a linkage study of 2,496 multiplex families with type 1 diabetes affected sib pairs. The ability to use family-based testing for association allowed the combined analysis of subjects ascertained from multiple different geographic areas while reducing concerns regarding the potential effects of population stratification. As an additional check for potential heterogeneity from family collection in the 9 T1DGC regions two approaches were used. First, a meta-analysis combining the 9 individual PDT results provided a \( P = 2.8 \times 10^{-5} \), the same magnitude of significance as our combined data analysis. Second, a Q statistic based on the association effect and standard error (SE) was computed. The association heterogeneity test (\( Q=5.15, \text{df}=7 \) [one group has missing estimates], \( P=0.64 \)) indicates no statistically significant heterogeneity across the regions of the T1DGC data. The validity of our test for association can also be shown from the genomic control number 1.02 when the whole genome scan results were examined (26).

Although SNP coverage was extremely sparse for genome-wide association testing, an advantage was that the SNP selection in the linkage panel of markers differs from the higher density panels used in previous genome-wide association studies. Three of the top four markers ranked by \( P \) value corresponded to previously established type 1 diabetes risk loci. The fourth-ranked SNP mapped to chromosome 21q22.3; no locus
associated at genome-wide significance levels with type 1 diabetes, or indeed, any autoimmune disorder, has previously been mapped to this chromosome. The region had, however, been previously linked to type 1 diabetes in Scandinavian families (27), and a region approximately 10 Mb centromeric to rs876498 has been fine-mapped in Danish families (28). Furthermore, a type 1 diabetes risk locus on chromosome 21 might have been anticipated given the increased frequency of Down's syndrome, caused by trisomy for chromosome 21, among type 1 diabetes patients (29). The findings reported here for rs876498 were validated by a blinded 15% resample using an alternative genotyping approach and then replicated in independent sets of parent-affected child trio families and case-controls. These studies provide compelling evidence for association of the minor allele at this SNP, rs876498, with type 1 diabetes.

The SNP rs876498 is located within the sixth intron of a gene variously referred to as Ubiquitin associated and SH3 domain containing protein 3A (UBASH3A) (30), T-cell ubiquitin ligand (TULA) (31) and Suppressor of T-cell signaling 2 (Sts-2) (32). Based on the available data in HapMap (release 23a) rs876498 is centrally located in a haplotype block of approximately 14 kb contained entirely within UBASH3A. Flanking genes in the region are not obvious candidates on functional grounds for involvement in type 1 diabetes etiology and include the trefoil domain containing family members TFF1 and TFF2, the transmembrane serine protease TMRSS3, RSPH1 which encodes a male meiotic metaphase chromosome associated protein, and the glycerol 3-phosphate transporter, SLC37A1.

UBASH3A is expressed predominantly in T cells where it interacts with c-CBL through its SH3 domain and binds to ubiquitin and ubiquitylated proteins via its UBA domain (for review see (25)). In T-cells it acts, in part, by inhibiting c-CBL mediated down-regulation of protein tyrosine kinases such as ZAP70 that are activated upon T-cell receptor (TCR) stimulation (32). In this regard, UBASH3A is reminiscent of another identified type 1 diabetes risk locus, PTPN22, whose product, LYP, also interacts with c-CBL but directly downregulates some of the same protein tyrosine kinases by dephosphorylation (6;33). UBASH3A has other functions in T-cells that include participation in pro-apoptotic pathways (34). The functions of UBASH3A make it an appealing candidate to contribute to type 1 diabetes risk. While it may be premature to focus exclusively on this single gene based upon the present data, the significant and replicated findings of association reported here strongly suggest that a gene or genes in this region contributes to risk for type 1 diabetes.

ACKNOWLEDGEMENTS

This research utilizes resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01 DK062418. Further support was provided by a grant from the NIDDK (DK46635) to PC and a joint JDRF and Wellcome Trust grant to the Diabetes and Inflammation Laboratory at Cambridge which also received support from the National Institute for Health Research Cambridge Biomedical Centre.
Table 1: Pedigree disequilibrium testing results from 6K SNP scan at $P < 0.001$

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>SNP</th>
<th>Position</th>
<th>P value</th>
<th>Locus</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>rs1004446</td>
<td>2126719</td>
<td>$2.5 \times 10^{-9}$</td>
<td>INS</td>
</tr>
<tr>
<td>2</td>
<td>rs1990760</td>
<td>162949558</td>
<td>$4.1 \times 10^{-5}$</td>
<td>IFIH1</td>
</tr>
<tr>
<td>16</td>
<td>rs887864</td>
<td>11066386</td>
<td>$7.1 \times 10^{-5}$</td>
<td>KIAA0350</td>
</tr>
<tr>
<td>21</td>
<td>rs876498</td>
<td>42714896</td>
<td>$1.0 \times 10^{-4}$</td>
<td>UBASH3A</td>
</tr>
<tr>
<td>11</td>
<td>rs2076837</td>
<td>14931161</td>
<td>$2.5 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs877783</td>
<td>72985946</td>
<td>$3.6 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>rs6767</td>
<td>5021488</td>
<td>$4.4 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>rs942434</td>
<td>7277013</td>
<td>$5.8 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs169679</td>
<td>28964551</td>
<td>$6.1 \times 10^{-4}$</td>
<td>HLA</td>
</tr>
<tr>
<td>22</td>
<td>rs885978</td>
<td>17572780</td>
<td>$6.4 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>rs1504</td>
<td>36977669</td>
<td>$6.8 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>rs860732</td>
<td>179360006</td>
<td>$7.0 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>rs1011094</td>
<td>28883961</td>
<td>$7.9 \times 10^{-4}$</td>
<td>HLA</td>
</tr>
<tr>
<td>7</td>
<td>rs1543851</td>
<td>64250422</td>
<td>$8.4 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>rs1872283</td>
<td>107880067</td>
<td>$9.0 \times 10^{-4}$</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>rs353087</td>
<td>15165591</td>
<td>$9.9 \times 10^{-4}$</td>
<td></td>
</tr>
</tbody>
</table>

$^a$Nucleotide position on indicated chromosome.

$^b$Most proximal locus to indicated SNP.
Table 2: Follow-up genotyping for rs876498 in type 1 diabetes families, cases and controls.

<table>
<thead>
<tr>
<th>Study population</th>
<th>Effective sample size</th>
<th>P</th>
<th>OR (95%CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original genome scan (2,496 ASP families)</td>
<td>6,172 trios + DSPs(^\text{a})</td>
<td>1.0x10(^{-4})</td>
<td>1.09 (1.04-1.13)</td>
</tr>
<tr>
<td>Follow-up: (2,214 parent-affected child trios)</td>
<td>890 trios</td>
<td>0.023</td>
<td>1.06 (1.00-1.11)</td>
</tr>
<tr>
<td>Follow-up: (7,721 affected individuals, 9,679 controls)</td>
<td>17,400</td>
<td>7.5x10(^{-8})</td>
<td>1.14 (1.09-1.19)</td>
</tr>
<tr>
<td>Overall</td>
<td>22,635</td>
<td>4.4x10(^{-12})</td>
<td>1.10 (1.07-1.13)</td>
</tr>
</tbody>
</table>

\(^\text{a}\)Discordant sib pairs
REFERENCES

collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. *Am J Hum Genet* 76:561-571, 2005


13. Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661-678, 2007


22. Purcell,S, Neale,B, Todd-Brown,K, Thomas,L, Ferreira,MA, Bender,D, Maller,J, Sklar,P, de Bakker,P, Daly,MA, Sham,PC: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559-575, 2007


