A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study

Hakon Hakonarson1,3*, Hui-Qi Qu5, Jonathan P. Bradfield1, Luc Marchand5, Cecilia E. Kim1, Joseph T. Glessner1, Rosemarie Grabs5, Tracy Casalunovo1, Shayne P. Taback6, Edward C. Frackelton1, Andrew W. Eckert1, Kiran Annaiah1, Margaret L. Lawson7, F George Otieno1, Erin Santa1, Julie L Shaner1, Ryan M Smith1, Chioma C Onyiah1, Robert Skraban1, Rosetta M. Chiavacci1, Luke J. Robinson1, Charles A. Stanley4, Susan E. Kirsch8, Marcella Devoto3,10, Dimitri S. Monos2,9, Struan F.A. Grant1,3, Constantin Polychronakos5*

1Center for Applied Genomics, and 2Department of Pathology and Laboratory Medicine, Abramson Research Center, The Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA.
3Department of Pediatrics and Division of Human Genetics, and 4Division of Endocrinology, The Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA.
2Departments of Pediatrics and Human Genetics, McGill University, Montreal H3H 1P3, Que’bec, Canada 6Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Manitoba, Canada. 7Division of Endocrinology, Children’s Hospital of Eastern Ontario, University of Ottawa, Ottawa, Ontario, Canada.
8Markham-Stouffville Hospital, Markham, Ontario, Canada.
9Department of Pediatrics University of Pennsylvania, School of Medicine Philadelphia, Pennsylvania 19104, USA.
10Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

Corresponding Authors:
Hakon Hakonarson, M.D., Ph.D.
Director, Center for Applied Genomics
1216E Abramson Research Center
3615 Civic Center Blvd.
Philadelphia, PA 19104 - 4318
hakonarson@chop.edu

Constantin Polychronakos M.D. F.R.C.P.C.
McGill University Health Center (Children's Hospital)
2300 Tupper, Montréal, Qc
Canada, H3H 1P3
constantin.polychronakos@mcgill.ca

Received for publication 13 September 2007 and accepted in revised form 10 January 2008.

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

Copyright American Diabetes Association, Inc., 2008
ABSTRACT

Objective: In Stage 1 of our genome-wide association (GWA) study for type 1 diabetes (T1D), one locus at 16p13 was detected (P=1.03x10^{-10}) and confirmed in two additional cohorts. Here we describe the results of testing, in these additional cohorts, 23 loci that were next in rank of statistical significance.

Research Design and Methods: Two independent cohorts were studied. The Type 1 Diabetes Genetics Consortium (T1DGC) replication cohort consisted of 549 families with at least one child diagnosed with diabetes (946 total affected) and DNA from both parents. The Canadian replication cohort consisted of 364 nuclear family trios with one T1D-affected offspring and two parents (1092 individuals).

Results: One locus at 12q13, with the highest statistical significance among the 23, was confirmed. It involves T1D association with the minor allele of rs1701704 (P=9.13×10^{-10}, OR=1.25, 95% CI=1.12, 1.40).

Conclusions: We have discovered a T1D locus at 12q13 which replicates in an independent cohort of T1D patients, and confers a comparable risk to T1D as the 16p13 locus we recently reported. These two loci are identical to two loci identified by the WGA study of the Wellcome Trust Case-Control Consortium, a parallel independent discovery that adds further support to the validity of the GWA approach.

ABBREVIATIONS. T1D, type 1 diabetes; GWA, genome-wide association.
Type 1 diabetes (T1D), a multifactorial disease with a strong genetic component, is due to the autoimmune destruction of pancreatic β-cells. The major T1D susceptibility locus, mapping to the HLA class II genes at 6p21(1) and encoding highly polymorphic antigen-presenting proteins, accounts for less than half of genetic T1D risk(2). The remainder is accounted for by many loci with more modest effects, of which only a few are known. These include: (1) the insulin (INS) VNTR(3), modulating thymic expression of and tolerance to insulin, a major T1D autoantigen(4; 5); (2) the Arg620Trp single-nucleotide polymorphism (SNP) at PTPN22, which affects the function of a negative regulator of TCR signaling(6); (3) Non-coding SNPs at IL2RA(7-9), which encodes the α chain of the IL2 receptor complex (CD25), an important modulator of immunity; (4) variants in the CTLA4 locus(10) whose protein product transmits inhibitory signals to attenuate T-cell activation. It is worth noting that all of these T1D-associated genes are expressed in cells with immune function and all except INS have been associated with other autoimmune disorders. Together, these loci explain little more than half of the genetic T1D risk, the remaining half being made up of loci whose number and effect size is still unknown. It is also worth remarking that these known loci represent the “low-hanging fruit,” involving candidate genes whose importance in autoimmunity was already known, thus limiting the value of genetic discovery to generate previously unsuspected physiological insights.

The recent development of high throughput genotyping array technologies has enabled us(11) and others(12; 13) to perform genome-wide association (GWA) studies in search of the remaining T1D loci. The first successful use in T1D involved screening of 12,000 nonsynonymous SNPs (nsSNP), which found T1D association with rs1990760, involving an Ala946Thr substitution on the IFIHI gene (Interferon-Induced with Helicase C domain 1)(14). More recently, the Wellcome Trust Case-Control Consortium (WTCCC) tested 2,000 T1D cases and 3,000 controls for 500 k SNPs (Affymetrix GeneChip) (12), and four novel associations (12q24, 12q13, 16p13 and 18p11) were solidly replicated in 4,000 cases and 5,000 controls, plus an additional 2,997 T1D family trios, by the Juvenile Diabetes Research Foundation (JDRF)/Wellcome Trust Diabetes and Inflammation Laboratory(WTDIL)(13).

Our Stage 1 results were recently published(11), reporting one novel T1D locus at Chr16p that reached genome-wide significance in Stage 1 and coincided with one of the novel WTCCC findings. We used the Illumina Hap550 array to type 550 k SNPs, selected to tag most of the haplotype structure of the human genome, per Phase 2 of the International HapMap project (http://www.hapmap.org). A set of 483 subjects with T1D and both their parents, plus 563 T1D probands and 1,146 controls, all of European ancestry, was examined. For the case-control data, population stratification was corrected by principal component analysis in the EIGENSTRAT implementation(15).

Both studies independently mapped this association to a 300kb linkage disequilibrium (LD) block on 16p, encompassing two genes of unknown function, KIAA0350 (now renamed C-type lectin 16A, symbol CLEC16A) and LOC642451.

Pending completion of a full Stage 2, in which the 1,536 markers with the highest statistical significance are being tested in additional cohorts, we decided to fast-track to replication 24 SNPs (23 distinct loci) that, in rank of statistical significance, came next to the 16p locus.
This number was arbitrarily determined by the multiplexing capacity of our genotyping platform and corresponds to a cutoff of approximately $10^4$ (Supplementary Table 1). The power of our GWA to detect a range of effect sizes at that level is given in Supplementary Fig. 1.

RESEARCH DESIGN AND METHODS

Subjects. The sample sets were used for WGA scanning have been previously described in detail(11). (1) The Type 1 Diabetes Genetics Consortium (T1DGC) replication cohort consisted of 549 families with at least one child diagnosed with diabetes (946 total affected) and DNA from both parents available as of the July 2005 data freeze (https://www.t1dgc.org). The samples were collected in Europe, North America and Australia and most subjects were of European ancestry. Criteria were age at diagnosis below 35 years and uninterrupted treatment with insulin within six months of diagnosis. For siblings of probands diagnosed under the age of 35, the age-at-diagnosis limit was extended to 45 if they were lean and had positive antibodies and/or low C-peptide levels at diagnosis. (2) The Canadian cohort consisted of 364 nuclear family trios with one T1D-affected offspring and two parents (1092 individuals). The samples were collected in pediatric diabetes clinics in Montreal, Toronto, Ottawa and Winnipeg. All patients were diagnosed under the age of 18, were treated with insulin since diagnosis and none has stopped treatment for any reason since. Ethnic backgrounds were of mixed European descent. The Research Ethics Board of the Montreal Children’s Hospital and other participating centres approved the study, and written informed consent was obtained from all subjects.

Genotyping. Genotypes for this study were obtained using the Sequenom iPLEX assay (Sequenom, Cambridge, MA). The 90 CEU (European-descent individuals genotyped HapMap) were included as accuracy controls. Call rate of each of the three 12q13 SNPs was $>98.2\%$, and no Mendelian error was found. The genotypes of each SNP were in Hardy-Weinberg equilibrium in the parents.

Statistics. T1D association was tested by the Family Based Association Test (FBAT) software (http://www.biostat.harvard.edu/~fbat/fbat.html)(16), based on the transmission disequilibrium test (TDT) method. Considering that most of the T1DGC families have multiple siblings, the option of the empirical variance was used in the FBAT statistics to permit a robust but unbiased test of genetic association. The age-of-onset effect of the SNPs in T1D patients was tested by the Kruskal-Wallis Test.

RESULTS AND DISCUSSION

As shown in Table 1, the T1D association at the12q13 locus was replicated in both sets, with similar odds ratios. In a combined analysis of the two WGA cohorts and a pool of the two replication cohorts by Fisher’s method, all three SNPs attained $P$-values ranging from $2.48 \times 10^{-8}$ to $9.13 \times 10^{-10}$ (Table 2). These data provide convincing evidence of a novel T1D locus. The other loci fast-tracked to Stage 2 were shown in Supplementary Table 1. While this report was in preparation, the WTCCC results were published, reporting four novel T1D associations, two of which coincide with our own two novel findings, at 16p (13) and the one reported here, 12q13. We consider this simultaneous independent discovery of non-candidate loci a potent validation of the GWA approach.

The three SNPs are located in a 250 kb block of tight LD on Chr12q and are highly correlated, with $r^2$ values (0.608 to 0.855). T1D risk was conferred by the minor allele in all three. The odds ratio (OR) and 95% CI of the haplotype of the 3 SNPs’ minor alleles was 1.25 (1.12, 1.40). The genotypic association of the SNPs best fits an additive
model of susceptibility on a log scale (Supplementary Table 2). There was a trend, with borderline statistical significance, of an effect on age-of-onset of T1D (Supplementary Table 3): homozygotes of the risk allele may have the disease onset 1 year earlier than the homozygotes of the protective allele. This is clearly the same locus as the one reported by Todd et al.(13), as the T1D-associated rs2292239 is in tight LD with our markers (Supplementary Fig.2). In addition, imputed genotype counts for our three significant markers in the WTCCC data showed highly significant T1D association, as shown in a meta-analysis of the two studies for the 12q13 locus (Supplementary Table 4). Significance was increased by several orders of magnitude. None of the other loci tested in this report (Supplementary Table 1) achieved statistical significance as a result this joint analysis.

The associated 250 kb LD block, encompasses several genes that now become candidates for the T1D association (Fig.1, and Supplementary Fig. 3). RAB5B (MIM:179514) encodes a member of the RAS oncogene family, which may be involved in vesicular trafficking at the plasma membrane(17). SUOX (MIM:606887) encodes a liver-specific sulfite oxidase, involved in the degradation of sulfur-containing amino acids. Neither gene has known function in immunity or pancreatic development. IKZF4 (MIM:606239), encodes a better functional candidate, a zinc finger protein specifically expressed in lymphocytes and implicated in the control of lymphoid development. ERBB3 (MIM:190151) encodes a member of the epidermal growth factor receptor (EGFR) family of receptor tyrosine kinases, involved in the regulation of cell proliferation or differentiation. Interestingly, it interacts with proteins with important immune functions, such as LYN, ITK, and FER(18). CDK2 (MIM: 116953) encodes a member of the Ser/Thr protein kinase family whose activity is regulated by its protein phosphorylation; members of this gene family have been recently implicated in GWA studies of type 2 diabetes(19-22). Fine mapping and functional studies will be required to identify the causative variant and generate functional insights from this genetic finding.

A more immediate conclusion from this finding is the support it provides to the robustness of the GWA approach, through identification of the same loci in two independent studies. Further, it is worth noting that our power to include in this fast-track replication a locus of magnitude comparable to the one reported here was only 38%. This means that our full second stage, plus future stages and combined meta-analysis are likely to reveal additional such loci.

ACKNOWLEDGMENTS

We thank all the patients, their parents and the healthy control subjects for their participation in the study. We gratefully acknowledge the use of DNA samples from the Type 1 Diabetes Genetics Consortium (T1DGC), funded by NIH grant U01-DK62418. This work was funded by the Juvenile Diabetes Research Foundation International and Genome Canada through the Ontario Genomics Institute. H.Q.Q. is supported by a fellowship from the Canadian Institutes of Health Research.

Conflict of Interest statement: None declared.
REFERENCES

12. 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


7
8


TABLE 1. The T1D-associated 12q13 SNPs in two independent cohorts

<table>
<thead>
<tr>
<th>SNP</th>
<th>Minor allele</th>
<th>Frequency</th>
<th>Informative families *</th>
<th>S †</th>
<th>E(S) †</th>
<th>Var(S) ‡</th>
<th>Z value</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1DGC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs773107</td>
<td>G</td>
<td>0.322</td>
<td>217</td>
<td>329</td>
<td>290</td>
<td>208</td>
<td>2.73</td>
<td>0.006400</td>
</tr>
<tr>
<td>rs10876864</td>
<td>G</td>
<td>0.417</td>
<td>226</td>
<td>395</td>
<td>350</td>
<td>229</td>
<td>2.98</td>
<td>0.002885</td>
</tr>
<tr>
<td>rs1701704</td>
<td>C</td>
<td>0.346</td>
<td>216</td>
<td>341</td>
<td>296</td>
<td>211</td>
<td>3.09</td>
<td>0.002019</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canadian set</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs773107</td>
<td>G</td>
<td>0.339</td>
<td>200</td>
<td>180</td>
<td>160</td>
<td>75</td>
<td>2.37</td>
<td>0.017736</td>
</tr>
<tr>
<td>rs10876864</td>
<td>G</td>
<td>0.452</td>
<td>212</td>
<td>210</td>
<td>191</td>
<td>87</td>
<td>2.09</td>
<td>0.036293</td>
</tr>
<tr>
<td>rs1701704</td>
<td>C</td>
<td>0.357</td>
<td>216</td>
<td>189</td>
<td>169</td>
<td>81</td>
<td>2.22</td>
<td>0.026268</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs773107</td>
<td>G</td>
<td>0.329</td>
<td>417</td>
<td>509</td>
<td>449</td>
<td>283</td>
<td>3.56</td>
<td>0.000374</td>
</tr>
<tr>
<td>rs10876864</td>
<td>G</td>
<td>0.432</td>
<td>438</td>
<td>605</td>
<td>540</td>
<td>316</td>
<td>3.64</td>
<td>0.000278</td>
</tr>
<tr>
<td>rs1701704</td>
<td>C</td>
<td>0.350</td>
<td>432</td>
<td>530</td>
<td>465</td>
<td>292</td>
<td>3.80</td>
<td>0.000148</td>
</tr>
</tbody>
</table>

* Number of nuclear families informative for (with a non-zero contribution to) FBAT analysis; † Observed allele number in the affected offspring; ‡ Expected allele number in the affected offspring; § Variance of allele distribution among the affected offspring.
**TABLE 2.** The combined $P$-values of T1D association of the three 12q13 SNPs

<table>
<thead>
<tr>
<th>SNP</th>
<th>CHR</th>
<th>Position</th>
<th>WGA case-control *</th>
<th>WGA Families</th>
<th>Stage 2</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs773107</td>
<td>12</td>
<td>54,655,773</td>
<td>2.89×10^{-5}</td>
<td>7.81×10^{-4}</td>
<td>3.74×10^{-4}</td>
<td>2.48×10^{-8}</td>
</tr>
<tr>
<td>rs10876864</td>
<td>12</td>
<td>54,687,352</td>
<td>8.39×10^{-5}</td>
<td>2.97×10^{-4}</td>
<td>2.78×10^{-4}</td>
<td>2.47×10^{-9}</td>
</tr>
<tr>
<td>rs1701704</td>
<td>12</td>
<td>54,698,754</td>
<td>9.89×10^{-6}</td>
<td>1.62×10^{-3}</td>
<td>1.48×10^{-4}</td>
<td>9.13×10^{-10}</td>
</tr>
</tbody>
</table>

*Corrected by the EIGENSTRAT method(15).
Figure 1. T1D association in the 12q13 region. The combined results of T1D association tests of the case-control cohort and the family cohort is shown. The red bars, corresponds to $-\log_{10}(P)$ values, represent the SNPs of significant T1D association with $P<0.05$. The LD map is based on our genotyping data of the family cohort. D’ values (%) are shown in the boxes, and the empty boxes have D’=1. The grey scale represents the $r^2$ values. The T1D associated SNPs map to a LD region in the middle of the figure, while the strongest association is from the region around the RAB5B gene. The $r^2$ values drastically drop beyond the two ends of the block containing the T1D associated SNPs. Relatively high D’ values continue for an additional ~250 kb in the telomeric direction, but beyond the block shown there is no T1D associated SNP and no HapMap SNPs with substantial $r^2$ values to those in the block shown (extended LD diagram in Supplementary Fig. 3).
FIGURE 1