Brief Genetics Report

A Comprehensive, Statistically Powered Analysis of GAD2 in Type 1 Diabetes

Gillian C.L. Johnson,1 Felicity Payne,1 Sarah Nutland,1 Helen Stevens,1 Eva Tuomilehto-Wolf,2 Jaakko Tuomilehto,2 and John A. Todd,1

GAD2 maps to chromosome 10p11.23 and encodes the 65-kDa isoform of GAD65, a major autoantigen in type 1 diabetes. The genetic variation that influences expression of preproinsulin mRNA, encoding another major autoantigen in type 1 diabetes, has already been shown to be genetically associated with disease. Previous reports that have assessed the association of GAD2 with type 1 diabetes have not used a dense map of markers surrounding the gene and have relied on very small clinical sample sizes. Consequently, no definite conclusions can be drawn from these studies. We have therefore systematically searched all exons and splice sites of the 3′ untranslated region (UTR), the 5′ UTR, and the 5′ upstream region of GAD2, for polymorphisms in 32 white European individuals. We have genotyped these polymorphisms in a maximum of 472 U.K. type 1 diabetic affected sib pair families exhibiting linkage to type 1 diabetes on chromosome 10p and have tested both single variants and haplotypes in the GAD2 region for association with disease. We subsequently followed up our results by genotyping a subset of these single-nucleotide polymorphisms in a maximum of 873 Finnish families with at least one affected child. Our results suggest that GAD2 does not play a major role in type 1 diabetes in these two European populations. Diabetes 51:2866–2870, 2002

GAD65 has been identified as a major autoantigen in type 1 diabetes, with autoantibodies against GAD65 present in the sera of 50–80% of newly diagnosed type 1 diabetic patients as compared with ~2% of the general population (1). It is still unclear whether immune responses against GAD are directly involved in the destruction of the islet β-cells or are merely a consequence of other pathogenic mechanisms that cause type 1 diabetes. However, in NOD mice, diabetes is characterized by early spontaneous T-cell reactivity to GAD, and prevention of this anti-GAD immunity results in the prevention of insulitis (β-cell autoimmunity) and diabetes (2–4). Although such compelling evidence for a primary role for GAD in human type 1 diabetes pathology has not been demonstrated, GAD autoantibodies can be detected several years before the onset of human disease (5), suggesting that autoimmunity to GAD could play a major role in the preclinical phase of type 1 diabetes (6).

Preproinsulin and GAD65, both strong candidates for primary type 1 diabetes autoantigens, are expressed in the thymus, where they can influence T-cell tolerance (7,8). Indeed, there is evidence to suggest that variation in thymic preproinsulin mRNA expression is associated with genetic predisposition to type 1 diabetes, encoded by a variable number tandem repeat locus 5′ of the insulin gene (IDDM2) (7,8), and it is possible that this could also be true for GAD65. GAD65 is encoded by GAD2, which maps to chromosome 10p11.23 (9,10). In addition to being a strong functional candidate, GAD2 is also a strong positional candidate for involvement in disease susceptibility because chromosome 10p11.2 shows evidence of linkage to type 1 diabetes in U.K. sib-pairs (11,12). Three previous studies have specifically attempted to address whether GAD2 is genetically associated with type 1 diabetes (12–14). However, none of these studies used a dense enough map of markers nor a large enough sample size to stand any reasonable chance of detecting a disease association, should one exist. Indeed, Reed et al. (12) recognized that only a high-resolution linkage disequilibrium (LD) map of single-nucleotide polymorphisms (SNPs) within the gene could comprehensively exclude a minor role for GAD2 in type 1 diabetes predisposition. Consequently, no definite conclusions can be drawn from these studies (12,13).

The entire contiguous genomic sequence of GAD2 can be found on a clone from the RPCI-13 BAC library, RP13-8106: AL389927 (available at www.sanger.ac.uk). The coding regions of GAD2 are organized into 16 exons spanning ~88 kb, and the sequence and genomic structure of rat, murine, and human GAD2 genes appears to be highly conserved (15–18). Indeed, comparing the noncoding 5′ flanking sequence of human GAD2 with upstream sequences in rat Gad2 (AF000195) reveals specific regions of high sequence similarity (>90%) in a region that has previously been suggested to contain a major TATA-less promoter that controls gene expression (17). In addition, sequence comparison between noncoding 5′ flanking regions of mouse and rat Gad2 also show high similarity.
diabetic (90%), again suggesting the presence of conserved regulatory elements controlling expression of the gene (16,18).

To evaluate whether variants in GAD2 influence susceptibility to type 1 diabetes, we scanned all coding sequences, most of the 3' untranslated region (UTR), and 2.8 kb of the 5' flanking region for polymorphisms (including the 5' UTR and all regions showing similarity to 5' flanking sequences in rat and murine Gad2) in 32 individuals. We detected 22 polymorphisms in total, 3 of which were in coding regions, and each caused a nonsynonymous change in the amino acid sequence of the protein. In addition to the 3 cSNPs detected in GAD2, we also genotyped 10 additional noncoding polymorphisms in a maximum of 472 multiplex families from the U.K. already showing evidence of linkage to disease in this region (11,12). This dataset provided >99.9% power to detect an effect equivalent to that seen at IDDM2 (relative risk [RR] = 2, disease allele frequency = 0.7), with \( P = 0.05 \) and \( \sim 77\% \) power to detect an effect at \( P = 0.05 \) for a disease allele with 5% frequency and an RR of 1.5. We selected the noncoding SNPs on the basis of their patterns of heterozygosity in the original screening panel of 32 individuals. As specific groups of SNPs were exclusively heterozygous in the same individuals, it is likely they all marked the same ancestral haplotype and therefore would be redundant in a first-pass association study. Therefore, we only chose a selection from each group to genotype in the diabetic family panel in the first instance. We also genotyped all other SNPs that were found to be heterozygous in at least three individuals in our screening panel. Our SNP map at GAD2 therefore extended over 86.5 kb, from 2,625 bp upstream of the GAD2 initiation codon to 91 bp before the start of the last exon (Fig. 1).

We observed 12 haplotypes when reconstructing parental haplotypes from all SNP markers (Fig. 1). A feature of the genomic architecture observed in the GAD2 region is that SNP allele frequency is a good indicator of LD. Indeed, SNPs with very similar allele frequencies are in almost absolute LD and therefore mark certain haplotypes. For example, five SNPs whose minor allele frequency is \( \sim 8\% \) (−2,625 A→G, 12,669 A→G, 12,750 T→G, 29,208 C→T, and 53,842 G→C) are mainly present on just one haplotype, haplotype D, and those with a minor allele frequency of \( \sim 15\% \) (−1,937 C→T, −1,399 C→A, −243 A→G, and 84 A→G) are mainly present on another haplotype, haplotype C. Consequently, analyzing just two haplotype-tagging SNPs (19), 12,689 C→A→T, the SNP that segregates three alleles, and one of −1,937 C→T, −1,399 C→A, −243 A→G, or 84 A→G, can capture the majority of the variation that we observed in this 87-kb region (Fig. 1). Indeed, SNPs marking haplotype D and haplotype C represent the groups of polymorphisms that were heterozygous in the same individuals in our screening panel, confirming that a limited amount of additional information would have been gained if the remaining markers in these groups had been genotyped. We tested all individual SNPs for association with type 1 diabetes using the transmission-disequilibrium test (TDT) (Table 1). The group of five SNPs, mainly present on haplotype D, including both the 5' flanking SNP, −2,625 A→G, which maps within a region of high sequence similarity (>90%) to a 5' flanking region in rat Gad2, and also a SNP in exon 10 of GAD2, which causes a Gly-Glu amino acid change in the mature protein, showed a slight deviation from the expected Mendelian rate of transmission (\( P < 0.05 \)).

![FIG. 1. A: Polymorphisms detected and genotyped at GAD2. B: U.K. parental haplotype reconstruction using 13 SNPs spanning the GAD2 gene.](image)

### TABLE 1

<table>
<thead>
<tr>
<th>Marker</th>
<th>Allele</th>
<th>Transmitted</th>
<th>Nontransmitted</th>
<th>( P^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>−2,625 A→G</td>
<td>A</td>
<td>127</td>
<td>91</td>
<td>0.02</td>
</tr>
<tr>
<td>−1,937 C→T</td>
<td>C</td>
<td>242</td>
<td>230</td>
<td>0.6</td>
</tr>
<tr>
<td>−1,399 C→A</td>
<td>A</td>
<td>186</td>
<td>191</td>
<td>0.82</td>
</tr>
<tr>
<td>−243 A→G</td>
<td>A</td>
<td>175</td>
<td>167</td>
<td>0.69</td>
</tr>
<tr>
<td>84 A→G</td>
<td>A</td>
<td>145</td>
<td>136</td>
<td>0.64</td>
</tr>
<tr>
<td>2,414 C→A</td>
<td>A</td>
<td>12</td>
<td>17</td>
<td>0.43</td>
</tr>
<tr>
<td>7,882 G→A</td>
<td>A</td>
<td>11</td>
<td>13</td>
<td>0.67</td>
</tr>
<tr>
<td>12,669 A→G</td>
<td>A</td>
<td>114</td>
<td>76</td>
<td>0.01</td>
</tr>
<tr>
<td>12,689 C→A→T</td>
<td>A</td>
<td>80</td>
<td>108</td>
<td>0.04</td>
</tr>
<tr>
<td>12,750 T→G</td>
<td>G</td>
<td>77</td>
<td>117</td>
<td>0.01</td>
</tr>
<tr>
<td>29,208 C→T</td>
<td>C</td>
<td>100</td>
<td>78</td>
<td>0.11</td>
</tr>
<tr>
<td>53,842 G→C</td>
<td>C</td>
<td>134</td>
<td>169</td>
<td>0.05</td>
</tr>
<tr>
<td>83,897 T→A</td>
<td>A</td>
<td>172</td>
<td>200</td>
<td>0.2</td>
</tr>
</tbody>
</table>

\(^*\)Robust variance estimators were used in the calculation of \( P \) values to ensure a valid test of association.
common haplotypes captured by the two SNPs, 12,689 C→A→T and 84 A→G, revealed weak evidence for a protective effect of haplotype D (41.4% transmission, P = 0.045). No other haplotypes analyzed showed any significant deviations from Mendelian transmission to affected offspring. The two SNPs defining the most common SNP haplotypes were also analyzed in conjunction with two microsatellite markers, D10S197 and GAD (13), in the region (Fig. 1). None of the four SNP/microsatellite haplotypes with a parental frequency >5% showed any evidence of association to disease (data not shown).

Although no additional U.K. families were available to extend the weak evidence for the association of SNPs on haplotype D with type 1 diabetes, 873 simplex families were available from Finland, the country with the highest incidence of type 1 diabetes (20). Initially, we genotyped seven of the GAD2 polymorphisms (−2,625 A→G, −1,937 C→T, 84 A→G, 12,669 A→G, 12,689 C→A→T, 12,750 T→G, and 83,897 T→A) in 224 families to evaluate LD/haplotype patterns of these markers in the Finnish population. The four most common haplotypes comprising these six markers were identical in both populations (Table 2). Therefore, it seemed reasonable to assume that if a variant on haplotype D (with a frequency of ~8% in the U.K.) was causing susceptibility to type 1 diabetes in the U.K., it might also have the same effect in Finland. As the three SNPs that mark haplotype D in the U.K. were in absolute LD in the Finnish families studied, the remaining families were genotyped with only the −2,625 A→G polymorphism to evaluate association of this haplotype/SNP in Finnish diabetic individuals. In a total of 814 fully genotyped families, this SNP showed no evidence of association to type 1 diabetes at P = 0.05 (assuming a multiplicative model [21]). Because our replication study was therefore reasonably powered, the initial weak result obtained in our U.K. families is likely to represent a false positive.

When a disease is known to be clinically heterogeneous, it may be beneficial to stratify the dataset on the basis of relevant predefined criteria (11,13). The HLA complex on chromosome 6p21.31 (IDDM1) contains genes encoding cell surface molecules that display peptides for immune recognition, with HLA-DRB1*03 (HLA-DR3) and HLA-DRB1*04 (HLA-DR4) haplotypes showing the greatest association to type 1 diabetes in European populations (22). GAD is a major autoantigen in type 1 diabetes, and peptide binding to class II molecules could be allotype specific. Indeed, it has been suggested that the early appearance of islet autoantibodies is common for DR3/4 heterozygous and DR4/4 homozygous offspring of type 1 diabetic parents (23) and that antibodies to GAD are specifically associated with DR3 alleles in siblings of type 1 diabetic subjects (24).

Offspring were therefore selected on the basis of having at least one copy of the high-risk HLA-DR3 or HLA-DR4 haplotypes to test whether the weak association seen with SNP 2,625 A→G in the original U.K. population was influenced by specific interactions with IDDM1. There was, however, no evidence for HLA-specific heterogeneity in the distribution of the 2,625A→G*G allele in unrelated patients from either the U.K. or Finland (Table 3).

Our polymorphism search at GAD2 included all coding sequences and a 5′ flanking region that likely plays a role in controlling expression of the gene (17). Moreover, as the mean level of LD between pairs of loci within the gene was considerable (Dmean = 0.86), it is possible that we would have also been able to detect, indirectly, the effects of any unidentified functional intronic variants. Combined with the large number of affected offspring (n = 914) studied in our first-pass scan, this should have allowed us to detect an effect equivalent to that of the insulin gene/IDDM2 locus (RR = 2, disease allele frequency = 0.7) or larger at P < 0.05 with >99.9% power. However, we only found very modest evidence for a disease association with a single haplotype (P = 0.045), a finding that was not replicated in a second large dataset comprising 814 families with at least one affected offspring. Therefore, the data from this first comprehensive evaluation of the role of GAD2 in type 1 diabetes suggest that genetic variation within this gene is unlikely to contribute significantly to the development of the disease. It is possible that a functional variant of a GAD2 transcription control element, which is out of LD with the SNPs we have typed, exists some distance from the GAD2 structural gene.
TABLE 3
Proportion of chromosomes from unrelated patients with a given HLA phenotype carrying the −26,25A>G*G allele.

<table>
<thead>
<tr>
<th></th>
<th>DR3+</th>
<th>DR3−</th>
<th>DR4+</th>
<th>DR4−</th>
<th>DR3/4</th>
<th>Non-DR3/4</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.K.</td>
<td>29/498 (5.8)</td>
<td>14/284 (4.9)</td>
<td>33/610 (5.4)</td>
<td>10/172 (5.8)</td>
<td>22/352 (6.2)</td>
<td>21/430 (4.9)</td>
</tr>
<tr>
<td>Finland</td>
<td>24/308 (7.8)</td>
<td>29/432 (6.7)</td>
<td>36/530 (6.8)</td>
<td>17/210 (6.1)</td>
<td>12/188 (6.4)</td>
<td>41/552 (7.4)</td>
</tr>
</tbody>
</table>

Data are n (%)..

However, our results also suggest that present resources would best be directed at other candidate genes or regions within the putative IDDM10 locus.

RESEARCH DESIGN AND METHODS

Subjects. We genotyped polymorphisms at GAD2 in a maximum of 472 multiplex families originating from the U.K. in which both sibs were diagnosed with type 1 diabetes (the Diabetes U.K. Warren 1 repository). We also genotyped markers in a maximum of 573 Finnish families in which at least one child had been diagnosed with type 1 diabetes.

Polymorphism detection. Polymorphisms in GAD2 were detected by genotyping 15 parents of type 1 diabetic sibs, 4 type 1 diabetes subjects, and 10 white control subjects not ascertained for disease status using DHPLC (WAVE). All individuals screened were white and born in the U.K. All regions screened were within 5 kb of the 5′UTR and 3 kb of the 3′UTR of the gene.

Genotyping. All polymorphisms were genotyped using Invader, as described on the Invader website (http://www-gene.cimr.cam.ac.uk/clayton/software/).

Statistical analysis. Parental haplotypes were reconstructed from pedigreed data using software available from Frank Dudbridge (http://www.lgmp.wisc. ac.uk/~fdudbrid/software/).

The TDT was performed with STATA using the genassoc package available from David Clayton’s website (http://www-gene.cimr.cam.ac.uk/clayton/software/). To ensure a valid test of association in multiplex families, robust variance estimators were used in the calculation of allele frequencies. TDT on multilocus haplotypes was performed using the software of Frank Dudbridge.

Contingency table analysis was used to compare allele frequencies of −2,625 A−G between groups of diabetic offspring stratified by HLA genotype. Only one diabetic subjects was chosen from each family. The P values shown are not corrected for multiple tests.

ACKNOWLEDGMENTS

This work was funded by the Wellcome Trust, the Juvenile Diabetes Research Foundation International, Diabetes U.K., and grants from the Finnish Academy (38387, 46558, 52114, and 51225), the Novo Nordisk Foundation, and the Diabetes Research Foundation International, Diabetes in Childhood.

REFERENCES

15. Bu DF, Tobin AJ: The exon-intron organization of the genes (GADI and GADI2) encoding two human glutamate decarboxylases (GADI6) suggests that they derive from a common ancestral GAD. Genomics 21:222–228, 1994