Linkage Analysis of Candidate Genes in Autoimmune Thyroid Disease: 1. Selected Immunoregulatory Genes

Giuseppe Barbesino, Yaron Tomer, Erlinda Concepcion, Terry F. Davies, and David A. Greenberg the International Consortium for the Genetics of Autoimmune Thyroid Disease

Address all correspondence and requests for reprints to: Giuseppe Barbesino, M.D., Mount Sinai Medical Center, New York, New York 10128. E-mail: gb@doc.mssm.edu.

DOI: http://dx.doi.org/10.1210/jcem.83.5.4813

Received: December 11, 1997
Accepted: February 02, 1998
First Published Online: January 14, 2009

Abstract

Graves’ and Hashimoto’s diseases are autoimmune thyroid diseases in which the genetic contribution is complex. For this reason, identification of necessary susceptibility genes has been difficult. However, a number of immunoregulatory genes have been implicated by association studies, including: CTLA-4, a recently described protein involved in antigen presentation, located on chromosome 2q33; the T-cell receptor Vα and Vβ gene complexes, located on 14q11 and 7q35, respectively; and the Ig gene complex (IgH), located on 15q11. We used polymorphic microsatellite markers located within these genes, or gene complexes, to test for linkage (rather than association), to each of these candidates. Using markers within the loci allowed us to assume a fixed recombination fraction of 0.01 in the tested model.

Three hundred eight subjects from 48 multiplex families were studied, with 142 affected subjects. Using this set of families, we have previously shown evidence of linkage with a major susceptibility locus for Graves’ disease (GD-1) on 14q24.3–31, with a maximum lod (logarithm + odds) score of 2.1, at a penetrance of 80% and with a dominant mode of inheritance. In the present study, we obtained consistently negative lod scores for each of the candidate genes, assuming either dominant or recessive modes of inheritance. These data, therefore, showed evidence against linkage with all the candidate genes. Unlike association studies, linkage analyses detect major genetic influences on disease susceptibility exerted by the linked loci. The lack of linkage for the immunoregulatory genes that were studied indicated, therefore, that they were not major contributors to disease etiology.

THE AUTOIMMUNE thyroid diseases (AITDs), Graves’ disease (GD) and Hashimoto’s thyroiditis (HT), have very different phenotypes. However, they are thought to share common pathogenetic mechanisms encompassing autoimmune reactions to thyroid epithelial cells mediated by both cellular and soluble mechanisms (1). Epidemiological and clinical data are consistent with the view that genetic factors play an important role in disease susceptibility. Inheritance of these diseases, as with many autoimmune disorders, is complex and, most likely, due to the occurrence of multiple susceptibility genes and the existence of environmental factors modulating the effect of these genes (2, 3). Therefore, identifying susceptibility genes has been most difficult. Many studies have addressed the complex problem of gene identification by using association studies. The human histocompatibility
leucocyte antigen (HLA) complex locus has been extensively studied by this method and found to have a well-defined influence on susceptibility to AITD in some ethnic groups (3). However, using the classical linkage analysis approach, which detects genes of major influence, our laboratory and others were unable to find evidence of linkage with HLA in a series of AITD families (4, 5, 6, 7). These observations indicate that the HLA locus has only a small influence on the overall genetic predisposition to the AITDs. Additional studies have been performed with alleles of loci of non-HLA genes that have a central role in the immune response. These have included CTLA-4, the T-cell receptor Vα and Vβ chains, and the Ig heavy chain region (IGH). Some studies have found positive associations, whereas others have not (see Ref. 3 for an extensive review of the available data). Discrepancies between studies may be explained by genetic differences between populations and/or by the use of different polymorphisms for the same genes, making comparisons among studies difficult. However, even in the reported positive studies, the relative risk never exceeded 3.2, suggesting a limited influence on disease expression for the genes analyzed.

In the present study, we present linkage analyses of a number of candidate immunoregulatory genes in a large series of multiplex AITD families. Genes studied included the T-cell receptor Vα and Vβ chains, the Ig heavy chain gene, and CTLA-4. Highly polymorphic microsatellite markers within or close to each selected locus were used in the analyses to be certain we were looking at effects associated with the genes in question. Negative lod (logarithm + odds) scores were observed with all the candidate genes studied, under all the inheritance models analyzed, with both GD and HT. Our results allowed us to exclude linkage with all the studied genes. These findings do not contradict previous reports of a positive association of GD and/or HT with some of the candidate genes we have analyzed, but the findings do indicate that these genes are not major determinants of genetic susceptibility to AITD.

Subjects and Methods

Patients

A total of 48 AITD multiplex families were analyzed. Forty-six families were of Caucasian descent, collected in New York, Italy, Britain, and Israel; 2 families were of Caribbean origin. Eleven families had at least 2 members with GD, whereas 20 families had at least 2 HT members. The remaining 17 families had at least two affected members, one of which had HT and the other had GD (mixed families). There were 29 families with 2 affected members, 14 families with 3, and 5 families with 4 or more, for a total of 142 patients. The diagnosis of GD was determined on the basis of documented clinical and laboratory evidence of present or past hyperthyroidism and the presence of at least one of the following: a diffuse goiter, positive TSH receptor antibody tests, or the presence of exophthalmos. The diagnosis of HT was based on past or present evidence of thyroid hormone-replaced primary hypothyroidism with any of the following: a diffuse, firm goiter, positive antithyroid peroxidase, or antithyroglobulin tests. Blood samples were collected from all available affected and nonaffected members of each family, after informed consent.

Allele determinations and typing

Whole blood was collected in tubes containing EDTA. Total genomic DNA was extracted using a commercial method (Puregene; Gentra Systems Inc., Minneapolis, MN). PCR reactions were conducted using fluorescent label primers, following the procedure of Weber (8). Briefly, each PCR reaction was performed in 10 μl of a mixture containing 50 ng total DNA, 0.12 U Taq polymerase (Perkin-Elmer, Foster City, CA), 5 pg fluorescent-labeled primers, deoxyribose dinucleotide triphosphate, and 1.5 mmol MgCl2 buffer. The primer sequences and marker characteristics are described in Table 1. Fluorescent labeled PCR products were denatured and separated on an ABI-310 (Applied Biosystems, Foster City, CA) automated sequencer. Allele typing was performed using Genotyper software in a semiautomated fashion.
Statistical analyses

Linkage analysis. Computer simulations were performed under different models, simulating a dataset of 50 families \(9\). The results showed maximum lod scores for a linked Mendelian locus ranging from 2.1 (at a penetrance of 20% and recessive mode of inheritance) to 19.25 (at 80% penetrance and dominant mode), suggesting that our dataset is sufficient to detect linkage in a wide range of models. Experimental family data were then analyzed using LIPED \(10\). The hypothesis of positive linkage was tested in different models, assuming a dominant or a recessive mode of inheritance. For each mode of inheritance, different levels of penetrance (20, 40, 50, 60, and 80%) were tested. Furthermore, to avoid the problem of age-dependent or reduced penetrance, an affected-only analysis also was performed on the dataset. Because all the markers studied were close to or within the respective candidate genes, a low (0.01) recombination fraction \(\theta\) was assumed in the model. Nonetheless, lod scores were obtained also for \(\theta\) up to 0.5 to examine the possibility of linkage to other genes in the vicinity of the markers studied. Also, a positive lod score at a distance from the analyzed marker could indicate heterogeneity (see below). The hypothesis of linkage was rejected when lod scores were < 2.0 or less. lod scores were first obtained considering all affected subjects, \(i.e.\) with either GD or HT (AITD affectedness). In a second step, to verify whether the studied genes could have a role in determining the type of AITD inherited, analyses were run with the same parameters as above, but classifying only persons with GD as affected or only with HT as affected. In each of these analyses, mixed families were included; and when the GD affectedness was examined, HT patients were considered as nonaffected and vice-versa. Because the total lod scores obtained with this method represented the algebraic summation of the lod scores obtained in each individual family, any results close to zero could have resulted from the coexistence of different families with very negative and very positive lod scores. This situation could have resulted from the phenomenon of genetic heterogeneity, in which the same phenotype was attributable to different genes in different families. Therefore, lod scores obtained for individual families were always examined in each of our analyses. In this situation, a test for heterogeneity was also applied in the linkage analysis.

Transmission disequilibrium test (TDT). The TDT \(11, 12\) was applied to the case of the CTLA-4 gene, for which the 106-bp allele of the marker used has been previously associated with GD and HT \(13, 14, 15, 16\). Briefly, all available parents heterozygous for the 106-bp allele and any other of the alleles, with one or more affected offspring, were included. The deviation of the transmission of the 106-bp allele from the expected 50% ratio was then examined in the affected offspring by \(\chi^2\) -square analysis.

Results

The 48 families

At least 1 affected first-degree relative of the proband was present in each of the families, making our set of families a potentially highly informative one for analyzing linkage. Of the 48 families analyzed, 26 were nuclear families, whereas the others had more complex structure, with affected subjects in 3 generations and/or in parallel branches. A total of 308 subjects were analyzed, with 54 GD patients and 70 HT patients. The female/male ratio was 4.4 in GD patients and 10.1 in HT patients. To reduce the chance of mislabeling carrier subjects, because of age-dependent penetrance, children less than 18 years old were not included, unless affected.

AITD

Table 2 shows the maximum and minimum lod scores obtained for the four markers at \(\theta = 0.01\), when all AITD patients were considered affected. None of the markers showed a maximum lod score more than 1 in either the recessive or the dominant model and at any of the penetrances analyzed. Figure 1 shows total lod scores obtained in AITD families with all
four markers at $\theta = 0.01$, considering different levels of penetrance, in both the recessive and dominant mode of inheritance. Lod scores less than −2 were observed at all levels of penetrance, in both the dominant and recessive model, the only exception being CTLA-4, when assuming a penetrance of 20% and a recessive mode of inheritance. Even in this extreme model, the lod score was still negative (−1.6). Similar results were obtained in the affected-only analysis, where all the candidates showed lod scores less than −2 in both the recessive and dominant mode, with the only exception being CTLA-4 in the recessive mode (−1.4). In none of the models did any of the families show a lod score of +1.0 or higher, suggesting that heterogeneity could not explain our results. When examining lod scores obtained at higher $\theta$, no evidence for linkage at loci in the vicinity of any of the candidate genes was observed. In the case of CTLA-4, however, a lod score of 1.008 was observed, at a $\theta$ of 0.2, penetrance 80%, in the dominant model (Fig. 2). A low positive lod score may be found in cases of positive association and in the absence of linkage, at relatively high $\theta$ (17). This could be explained by the fact that, in the presence of an association, affected subjects are more likely to carry the associated allele (therefore, yielding higher lod scores). At the same time, the presence of many other affected subjects not carrying the allele is interpreted in the linkage calculations as a high frequency of recombinations, placing the maximum lod score at a higher than expected distance from the marker itself. To verify the hypothesis of positive association, we carried out a TDT on the available data, as described in Subjects and Methods. Another possible explanation for the results observed for CTLA-4 is the presence of heterogeneity. However, analysis of the data for CTLA-4, in a model assuming heterogeneity, showed a maximum lod score of 0.87 in only 25% of the families, making heterogeneity an unlikely explanation for our results.

TDT

Twenty-two of the parents of 1 or more AITD-affected members were heterozygous for the 106-bp allele of CTLA-4, with a total of 31 affected offsprings. Of these 31 affected family members, 16 inherited the 106-bp allele from 1 or 2 of the heterozygous parents, whereas 15 did not, a distribution which was very close to the 50% expected by chance alone ($P$, not significant). It must be noted, however, that our sample (not designed for the TDT specifically) is small and therefore able to detect only major associations.

GD and HT

Results similar to the combined AITD grouping were obtained when the affectedness status of GD or HT were separately considered (Table 3). In general, the data set was less informative, because of the smaller number of families available for each analysis. Mainly negative results were observed; although, in all the models studied, the lod score did not reach a sufficiently negative level to be able to reject linkage with total certainty.

Discussion

We have applied the principles of linkage analysis to the study of a number of common immunoregulatory genes in the inherited susceptibility to GD and HT. Although environmental factors, such as infection, may play an important role (2), genetic factors have a major role in predisposition to AITD (3). GD, HT, and thyroid autoimmune phenomena run in families (18, 19, 20). A significant concordance rate for these diseases is found in monozygotic twins (36% in a recent report), higher than in dizygotic twins (0%), suggesting an important role for both genetic and environmental factors (21). GD and HT can occur in different members of the same family (22), as was observed in our data set, probably on the basis of of a common genetic background for the two diseases. The genetic transmission of AITD, however, seems to be complex and does not represent a simple Mendelian model (3). Indeed, the familial pattern is more characteristic of multigenic diseases, in which multiple genes may contribute to the clinical phenotype (23). The available studies have mainly addressed the identification of susceptibility genes through the approach of association studies. These have compared the
frequency of a polymorphism, at a candidate gene locus in a genetically homogeneous affected population, with the one observed in a genetically similar control population. This approach is considered very sensitive but has limitations. In particular, the matching of the control population for genetic background is crucial, to be able to accept statistically significant results as a real effect of disease presence. Moreover, in any association study, the marker loci studied have to play a pathogenic role in the phenotype generation or must be in linkage disequilibrium with a disease allele at the disease locus. Finally, association can identify genes that are not necessary for disease expression but merely increase the risk for the disease (11). It has been suggested that the problem of the control group selection and/or of the population stratification that is inherent in any population association study may be overcome by the TDT, which examines the distribution of one parental allele in the affected offspring (12). In the case of a positive association, the affected offspring of parents carrying one diseased and one normal allele would show a prevalence of the former allele higher than the 50% expected by simple chance.

Many association studies have looked at immunoregulatory genes. The HLA locus has been most intensely studied, and several haplotypes have been shown to increase the risk for GD and/or for HT, with a maximum relative risk of 3.8 for Caucasian GD patients with the HLA-DQA10501 haplotype (13, 24). Among the non-HLA genes, the T-cell receptor α and β genes have been deemed particularly interesting because of the central role of their products in the normal and aberrant immune response. A polymorphism in the Vα chain gene was first noted to be associated with HT in a small series of patients (25), but subsequent studies have failed to confirm this finding (26). A similar situation was found in GD, in which some studies have shown association of the disease with the T-cell receptor β chain (27), whereas others have not (28, 29). The Ig heavy chain gene also has yielded positive results in association studies (30) but not in linkage analyses (6). Recently, much attention has been given to CTLA-4, a protein involved in antigen presentation. CTLA-4 has emerged as a central factor in the homeostasis and down regulation of the immune response (31) and, therefore, represents an important candidate gene. The CTLA-4 locus has been shown to be linked to IDDM and associated with GD (14) and, subsequently, also with HT, with similar risk ratios (32, 33, 15, 16). Thus, CTLA-4 represents a significant susceptibility gene in addition to HLA.

We addressed the role of some of these genes (IGH, CTLA-4, TCR Vα, and TCR Vβ) by the method of linkage analysis. Linkage analysis has been used successfully in mapping a large number of genes that cause monogenic disease. However, it can also be used in locating susceptibility genes for complex multigenic diseases, as shown by the example of IDDM (34). Linkage has several advantages over association: it allows the identification of genes that may be necessary for disease development and does not involve possible biases caused by mismatched control populations. Moreover, the markers employed may be many million base pairs away from the disease locus and still show linkage. The marker loci need only be in the genetic neighborhood of the disease gene. Furthermore, linkage analysis allows whole genome scanning without any previous knowledge of the locus sought (35). However, linkage yields no information on disease mechanisms and becomes less sensitive as the number of disease loci increases. Also, linkage requires a large amount of data and is less sensitive, in the sense that it fails to detect loci with small overall effects on the total susceptibility to a multigenic disease.

In the present study, we used the candidate gene approach and chose highly polymorphic microsatellite markers from the noncoding sequence of each candidate gene or gene complex. This allowed us to assume low recombination fractions in the linkage model, thus adding a great deal of power to the analysis. By these means, we were able to fully exploit the intrinsic informativeness of the family data set, and we explored a wide range of inheritance models. In a subset of these families, we have previously shown evidence suggestive of linkage to a locus on chromosome 14q31, which we have termed GD-1, thus demonstrating that our set of families was adequate to screen the genome for candidate loci (36, 37).
The data presented here allowed the rejection of linkage with AITD for a number of important immunoregulatory genes that have yielded conflicting results in the earlier association studies, such as the TCR Vα and Vβ chains and the Ig heavy chain. The same results applied to CTLA-4, which has been consistently associated with AITD in a number of independent laboratories. We, therefore, tested the hypothesis of association for CTLA-4 by the TDT in our set of families. Because our series was not designed for such an approach, the data available (mainly dependent on the number of available heterozygous parents) was limited and was likely to detect only a large effect of CTLA-4. However, the 106-bp allele was transmitted to almost 50% of the 31 affected offspring, providing no evidence for an association. It should be remembered that in the earlier published association studies, affected cases of GD and HT were randomly selected, independently of their family history. For our linkage approach, probands were selected based on the high incidence of thyroid autoimmune diseases in their families. It is, therefore, possible that different genes have different roles in familial AITD, as compared with sporadic AITD, explaining the discrepancy observed in our data with CTLA-4. Indeed, an association may only be present in a nonfamilial phenotype.

In conclusion, the linkage approach reported here initiates the process of excluding minor contributing genes on the way to finding the major genes that cause the human AITDs.

1 This work was supported, in part, by NIH Grants DK-52464, DK-35764, and DK-45011 (to T.F.D.) and DK-02498 (to Y.T.), the David Owen Segal Endowment, the Consorzio Pisa Ricerche-University of Pisa (to G.B.), and NIH Grants NS-27941, MH-48858, and DK-31775 (to D.A.G.). Additional funding was received from the Endocrine Fellows Foundation (to G.B.).

2 Additional members of the International Consortium for the Genetics of Autoimmune Thyroid Disease included: Drs. Rhoda Cobin (New York, NY), Luca Chiovato and Aldo Pinchera (Pisa, Italy), Sandra McLachlan (San Francisco, CA), Bernard Rees Smith (Cardiff, Wales), Fred Clark and Eric Young (Newcastle upon Tyne, UK), and Meir Berezin (Tel Hashomer, Israel). These contributors were not responsible for the content of the present manuscript.

Table 1. Characteristics of the microsatellite markers used in the study

<table>
<thead>
<tr>
<th>Gene locus</th>
<th>Chromosomal location</th>
<th>Marker name</th>
<th>Primers (5′–3′)</th>
<th>Allele size range (bp)</th>
<th>PIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgH</td>
<td>14q32</td>
<td>IgH@</td>
<td>tgttgaagaagggagtctgtttgcactcatgtttg</td>
<td>190–200</td>
<td>0.67</td>
</tr>
<tr>
<td>CTLA-4</td>
<td>2q33</td>
<td>CTLA4 (+)</td>
<td>gccagtgcatggaaagtggccgactcatgg</td>
<td>80–132</td>
<td>0.91</td>
</tr>
<tr>
<td>TCR A</td>
<td>14q11–12</td>
<td>TCRA</td>
<td>aacatccgctgctatagcatgccc</td>
<td>186–200</td>
<td>0.84</td>
</tr>
<tr>
<td>TCR B</td>
<td>7q35</td>
<td>R–G</td>
<td>atgtgactgtaagaaacagt</td>
<td>126–156</td>
<td>0.83</td>
</tr>
</tbody>
</table>

PIC, the polymorphic informative content of each marker.

- See more at: http://press.endocrine.org/doi/10.1210/jcem.83.5.4813#sthash.NdcV1jpe.dpuf
Table 2. Maximum and minimum lod scores obtained for each of the markers, at $\theta = 0.01$, when all patients with AITD were considered as affected

<table>
<thead>
<tr>
<th>Gene</th>
<th>Marker</th>
<th>Maximum lod score</th>
<th>Minimum lod score</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGH</td>
<td>IGH@</td>
<td>−2.911</td>
<td>−8.368</td>
</tr>
<tr>
<td>CTLA-4</td>
<td>CTLA4(+)</td>
<td>−1.622</td>
<td>−5.896</td>
</tr>
<tr>
<td>TCR-A</td>
<td>TCRA</td>
<td>−6.626</td>
<td>−16.939</td>
</tr>
<tr>
<td>TCRB</td>
<td>R–A</td>
<td>−9.058</td>
<td>−22.412</td>
</tr>
</tbody>
</table>

- See more at: http://press.endocrine.org/doi/10.1210/jcem.83.5.4813#sthash.NdcV1jpe.dpuf

Table 3. Maximum and minimum lod scores obtained at $\theta = 0.01$, when patients with GD alone or with HT alone were considered affected

<table>
<thead>
<tr>
<th>GD</th>
<th>Max lod</th>
<th>Min lod</th>
<th>Max lod</th>
<th>Min lod</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGH@</td>
<td>0.097</td>
<td>−5.455</td>
<td>0.282</td>
<td>−16.803</td>
</tr>
<tr>
<td>CTLA4(+)</td>
<td>−1.202</td>
<td>−6.875</td>
<td>0.522</td>
<td>−8.792</td>
</tr>
<tr>
<td>TCRA</td>
<td>−0.344</td>
<td>−9.873</td>
<td>−3.779</td>
<td>−15.536</td>
</tr>
<tr>
<td>R–A</td>
<td>0.671</td>
<td>−17.666</td>
<td>−3.012</td>
<td>−17.879</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HT</th>
<th>Max lod</th>
<th>Min lod</th>
<th>Max lod</th>
<th>Min lod</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- See more at: http://press.endocrine.org/doi/10.1210/jcem.83.5.4813#sthash.NdcV1jpe.dpuf
Figure 1. Lod score obtained with the four markers in the AITD families, assuming different penetrances in the linkage model. Results obtained in both the recessive (open circles) and dominant (filled squares) models are shown.
Figure 2. Lod scores obtained with the CTLA-4 microsatellite marker, assuming different recombination fractions (θ, on the x-axis) and different penetrances, assuming a dominant mode of inheritance.

We are indebted to Drs. P. F. Watson and A. P. Weetman (Sheffield, UK) for providing samples of DNA genotyped for CTLA-4, allowing correct sizing of alleles in our system.


15. **Kotsa K, Watson PF, Weetman AP.** 1997 A CTLA-4 gene polymorphism is associated with both Graves disease and autoimmune hypothyroidism. *Clin Endocrinol (Oxf).* 46:551–554. [CrossRef], [Medline]


23. **McLachlan SM.** 1993 The genetic basis of autoimmune thyroid disease: time to focus on chromosomal loci other than the major histocompatibility complex (HLA in man) [editorial; comment]. *J Clin Endocrinol Metab.* 77:605A–605C. [Abstract], [Medline]


