The different approaches to the genetic analysis of autoimmune thyroid disease

A Allahabadia1 and S C L Gough1,2

1University of Birmingham, Queen Elizabeth Hospital, Edgbaston, Birmingham B15 2TH, UK
2Birmingham Heartlands Hospital, Bordesley Green East, Birmingham B9 5SS, UK

(Requests for offprints should be addressed to S C L Gough, Department of Medicine, University of Birmingham, Birmingham Heartlands Hospital, Birmingham B9 5SS, UK)

Abstract

Graves’ disease and Hashimoto’s thyroiditis are organ-specific autoimmune disorders of multifactorial aetiology with a polygenic mode of inheritance. Familial clustering and twin studies provide evidence for a genetic predisposition. Three main approaches have been used in the search for susceptibility loci: population-based case–control studies, classical linkage analysis, and intrafamilial linkage disequilibrium. Case–control studies are a sensitive method of gene detection and the collection of subjects is resource-efficient. However, they require prior knowledge of a candidate gene and are prone to inconsistent results due to false positives that may arise from population mismatch. Linkage analysis is a powerful tool for detecting ‘major’ genes that does not require a candidate gene and, therefore, a means of genome screening. This method, however, has limited power to detect genes of ‘modest’ effect, and the collection of sibpairs and multiple family members may be difficult. Intrafamilial linkage disequilibrium analysis is more sensitive than classical linkage analysis, requires only one affected offspring, and eliminates population mismatch. This approach has confirmed linkage disequilibrium of the HLA region with Graves’ disease, previously not detected by linkage analysis. Knowledge of a candidate locus is required, however, and this method cannot, therefore, at present be used for genome screening. It is likely that a combination of all three approaches will be required to identify susceptibility loci for autoimmune thyroid disease.

Introduction

Graves’ disease and Hashimoto’s thyroiditis are organ-specific autoimmune disorders of the thyroid gland, collectively known as autoimmune thyroid disease. They are multifactorial conditions that occur in genetically susceptible individuals, with environmental factors influencing presentation. The interaction of a permissive environmental factor with an immune system that fails to distinguish adequately between self and non-self may lead to the activation of pathogenic immune mechanisms. This breakdown in immune tolerance results either in stimulation of thyroid cells in Graves’ disease, producing biochemical hyperthyroidism, or in destruction of the cells in Hashimoto’s thyroiditis. Together, these conditions constitute the most common cause of altered thyroid function.

Evidence for genetic factors may be seen in the clustering of these diseases within families and in twin studies. The largest study assessing heredity in Graves’ disease indicated that more than 40% of patients had a family history of thyroid disease (Bartels 1941). Similarly, it is well recognised that Hashimoto’s thyroiditis also has a high degree of familial clustering. Both of these conditions are frequently found in members of the same family, supporting the genetic hypothesis and implying a shared genetic predisposition. Twin studies provide evidence for both a genetic and an environmental contribution to disease. Concordance rates for Graves’ disease in identical (monozygotic) twins have been reported to be as high as 30–40%, compared with 3–9% in non-identical (dizygotic) twins (Stenzky et al. 1985). As identical twins share all of the same genes and non-identical twins share, on average, half their genes, the difference in concordance rates can be explained by genetic factors. More recent work with Danish twins (Brix et al. 1998) confirmed the importance of genetic factors in the aetiology of Graves’ disease. The lower pairwise concordance rates, however, imply that the genetic effect might not be as strong as originally believed. In all epidemiological studies, concordance rates
Such genes are neither termed susceptibility loci, contributing to inheritance. Polygenic disorders with several genetic regions, termed susceptibility loci, contributing to inheritance. Such genes are neither sufficient alone nor necessary to cause disease (Heward & Gough 1997). The search for susceptibility loci is ongoing, with the aims of gaining a better understanding of disease pathophysiology and of being able to predict outcome of disease at onset and to identify other family members at risk of disease, all of which will improve disease management. Three main strategies have been applied in humans to determine the identity of susceptibility loci. These include population-based case–control studies, family-based linkage analysis, and family based linkage disequilibrium analysis (Table 1).

### Population-based case–control studies

Population-based case–control studies investigate association of a marker allele with disease by comparing the frequency of the allele in a diseased population with that in a disease-free population. If there is allelic association or linkage disequilibrium between the marker allele and the disease, then there will be a significant excess of the marker allele in the disease population. It is not possible, however, to state from such studies whether the associated allele is the primary disease-causing mutation, or whether the allele is acting as a marker and is in linkage disequilibrium with a nearby unknown aetiological mutation. Either way, such studies require prior knowledge of a ‘candidate’ gene with a known or suspected biological function being implicated in the pathogenesis of the disease.

Although widely used, this methodology has met with limited success in identifying susceptibility loci for autoimmune thyroid disease. Several case–control studies have demonstrated association of alleles of the class II MHC (HLA) region on chromosome 6p with Graves’ disease (Boehm et al. 1992, Yanagawa et al. 1993, Badenhoop et al. 1995, Barlow et al. 1996) and Hashimoto’s thyroiditis (Tandon et al. 1991). In the majority of these studies, results have been consistent, with allelic association reported between Graves’ disease and the alleles DR3 and DQA1*0501 (Boehm et al. 1992, Yanagawa et al. 1993, Badenhoop et al. 1995, Barlow et al. 1996). Because of the sensitivity of the case–control method and results already reported, most adequately sized, well matched data sets would be expected to identify the HLA region as a susceptibility locus.

The CTLA-4 gene on chromosome 2q33 is a good candidate gene for autoimmune thyroid disease, as it has an important role in T cell regulation (Bluestone 1997). It is a member of the same family of cell-surface molecules as CD28 and, along with CD28, it can bind to B7. The CTLA-4/B7 complex competes with the CD28/B7 complex and delivers negative signals to the T cell and affects T cell expansion, cytokine production and immune responses (Bluestone 1997). The CTLA-4 gene was first identified as a candidate gene for Graves’ disease (Yanagawa et al. 1995), and further case–control studies have consistently detected association of the CTLA-4 gene with Graves’ disease (Yanagawa et al. 1995, Nistico et al. 1996, Donner et al. 1997, Kotsa et al. 1997a), and Hashimoto’s thyroiditis (Kotsa et al. 1997a).

Several other loci, including the thyrotrophin receptor gene (Cuddihy et al. 1995, Kotsa et al. 1997b) and interleukin-1 receptor antagonist gene (Blakemore et al. 1995, Cuddihy & Bahn 1996), have yielded conflicting results in autoimmune thyroid disease. Such inconsistencies highlight the inherent problems with the population-based case–control method. False positive associations arising as a result of a chance event or random variation are highly likely in small data sets, and many case–control studies have been performed on data sets of inadequate numbers. Moreover, it is almost impossible to obtain perfectly matched diseased and normal control populations. False positives arising from population

<table>
<thead>
<tr>
<th>Locus</th>
<th>Population-based case–control analysis</th>
<th>Classical linkage analysis</th>
<th>Intrafamilial linkage disequilibrium analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA region</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>CTLA-4 gene</td>
<td>+</td>
<td>–</td>
<td>+*</td>
</tr>
<tr>
<td>TSH-R gene</td>
<td>+/−</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>β3-antagonist gene</td>
<td>+/−</td>
<td>–</td>
<td>−*</td>
</tr>
<tr>
<td>GD-1</td>
<td>+</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

+, positive result; −, negative result.

*S. C. L. Gough, unpublished observations.

---

*Table 1 Summary of genetic studies performed in autoimmune thyroid disease. See text for references*
stratification (Gough et al. 1995) are also more likely to occur in small data sets. As we shall see later in this review, the major problem of population stratification can be overcome by using family-based association studies. Despite their limitations, however, population-based case–control studies do have a number of advantages. First, the identification and collection of samples from subjects is far quicker and more resource-efficient than the collection of family samples. Secondly, it may be easier to recruit adequate numbers of subjects in late-onset diseases such as autoimmune hypothyroidism rather than recruit index cases with additional family members, including parents who are less likely to be alive. Finally, the population-based case–control study is more sensitive than some of the family-based studies and is more likely to detect genes of modest effect. However, the problem of population stratification can be significant (Gough et al. 1995) and it is difficult to ascertain if association is due to linkage disequilibrium or the result of population stratification. The trend, therefore, in the genetic analysis of autoimmune thyroid disease, is towards the use of family-based data sets.

Classical linkage analysis

Linkage is defined as the tendency of genes to be inherited together as a consequence of their physical proximity on a single chromosome. Linkage analysis requires no prior knowledge of the likely chromosomal position of a disease gene. If a chromosomal marker of known location is inherited together with a disease ( cosegregation ) and the cosegregation is statistically significant across a large number of families, then the chromosomal marker must be located close to the disease gene on the same chromosome and the two are said to be linked. Linkage analysis can be performed as part of genome-wide screening or simply on a single chromosomal region of interest, using a series of microsatellite markers. These markers are generally scattered throughout the genome, and comprise of sequences of 2, 3 or 4 nucleotides repeated a variable number of times. They can be highly polymorphic, with several alleles at a single locus. Microsatellites are usually analysed in data sets of large numbers of families in which DNA is available from two affected siblings, with or without parental DNA. Under a model of Mendelian inheritance, siblings would be expected to share zero alleles 25% of the time, one allele 50% of the time and two alleles 25% of the time (Fig. 1). If there is a significant excess of allele sharing between sibpairs with disease, genetic linkage is present. Once preliminary evidence of linkage is obtained within a data set, it is crucial that confirmation is sought in further families. This form of linkage analysis is termed ‘non-parametric’, and has the advantage that it requires no knowledge of a model of inheritance that is unknown in complex diseases.

The approach of linkage analysis has been used successfully in type 1 diabetes, with a genome-wide search identifying at least 14 susceptibility loci (Davies et al. 1994). Genome screens have been performed in other complex diseases with inconclusive results, most notably in multiple sclerosis (Ebers et al. 1996, Sawcer 1996, The Multiple Sclerosis Genetics Group et al. 1996). In autoimmune thyroid disease, no genome screens have been performed to date, but linkage analysis has been performed at various regions of interest, including the HLA region on chromosome 6p (Roman et al. 1992, Ratanachayivong et al. 1994, Shields et al. 1994), the thyrotrophin receptor gene ( TSH-R ) on chromosome 14q (De Roux et al. 1996) and GD-1 on 14q31 (Tomer et al. 1997). Studies of the HLA region have been performed in both Graves’ disease and Hashimoto’s thyroiditis (Roman et al. 1992), but have...
failed to detect linkage. Similar results have been obtained for the CTLA-4 gene (Barbesino et al. 1998) and the TSH-R gene (De Roux et al. 1996). Although authors claim to have data sets of sufficient size to detect linkage, positive results for regions such as HLA have not been found and they suggest, therefore, that this region can be exerting only a modest effect on disease. Loci including the HLA region may well be exerting a modest effect on disease but, as the total genetic contribution to autoimmune thyroid disease may also be modest, such regions are likely to be major contributors to overall genetic susceptibility. These regions should not, therefore, be disregarded. Graves’ disease and type 1 diabetes are often found to co-exist in the same individuals (Payami et al. 1989). As the HLA region is the major genetic determinant in type 1 diabetes (Todd 1995), it could be hypothesised that HLA is playing a similar role in Graves’ disease.

Tomer et al. (1997) have suggested linkage of Graves’ disease to a region on chromosome 14q31 at a distance of 25cM from the TSH-R gene, a region that overlaps the type 1 diabetes susceptibility locus, IDDM11. These results are important, although at present they await replication in additional data sets.

Classical linkage analysis is more likely to identify susceptibility loci in diseases with a greater genetic contribution (Todd 1995). The most common measure of genetic contribution to an individual disease is \( \lambda_s \) – the lifetime risk to siblings, divided by the general population frequency of the disease. Values of \( \lambda_s \) are highest in monogenic disorders (e.g. cystic fibrosis, \( \sim 500 \)) and considerably lower in complex diseases (e.g. type 2 diabetes, \( \sim 3-5 \)), hence linkage analysis has been more successful in single gene disorders. In Graves’ disease, we have estimated a \( \lambda_s \) of around 10 on the basis of epidemiological data provided in studies by Bartels (1941) and Howell-Evans et al. (1967). However, follow-up data from the Whickam survey have shown that Graves’ disease is more common than previously believed, with a prevalence of approximately 2% (Vanderpump et al. 1995). If this figure is correct the genetic contribution would be even smaller, with a \( \lambda_s \) of 2.

A number of linkage studies have been performed in type 1 diabetes, identifying several loci, each with an individual contribution to the overall genetic susceptibility, known as the locus \( \lambda_c \). Outside the HLA region, most loci such as IDDM2 have a locus \( \lambda_c \) of between 1·1 and 1·3 (Todd 1995). It has been estimated that thousands of families would be required to detect a locus with such a magnitude (Risch & Merikangas 1996). If there was a major gene in autoimmune thyroid disease with a \( \lambda_c \) of between 1·4 and 2, approximately 400 families would be required to detect it. In complex disease, it is also likely that loci act in a multiplicative model in which the loci interact, so that the overall risk when all are present is greater than the sum of the independent risk of each individual locus. For example, if a disease had a \( \lambda_c \) of 8 with three genes, then each with locus would have a \( \lambda_c \) of 2 in a multiplicative model. It is, therefore, not difficult to see why linkage studies in autoimmune thyroid disease that have used small data sets of between 14 (De Roux et al. 1996) and 48 families (Barbesino et al. 1998) have failed to detect linkage. One of the methods adopted in order to overcome this problem has been to use parametric linkage in multigenerational pedigrees (Barbesino et al. 1998), which increases the size of the data set by using all available relatives. This method, however, requires a knowledge of the mode of inheritance and likely penetrance, both of which have had to be estimated. Such estimates are very approximate, may strongly influence the results of the studies and should, therefore, be viewed with caution. Furthermore, multiethnic data sets (Tomer et al. 1997, Barbesino et al. 1998) have been used, and it is recognised from population association studies that allele frequencies vary in different ethnic groups (Allahabadia et al. 1998).

In summary, the approach of linkage analysis in groups of affected sibpairs or multiplex families (in which parental DNA is also available), is a powerful tool for detecting ‘major genes’ in complex diseases, particularly as no candidate is required. However, this method has limited power to detect genes of modest effect. The lack of replication and inconclusive results seen in multiple sclerosis (Ebers et al. 1996, Sawcer et al. 1996, The Multiple Sclerosis Genetics Group 1996) highlights the importance of the size of the data sets to be used. In contrast, replication of linkage data seen in type 1 diabetes (Todd 1995) is almost certainly the result of large numbers of families within data sets and also the presence of parental DNA, allowing analysis identical by descent as opposed to identical by state (Fig. 1). Unfortunately, because of the late onset of the disease, autoimmune thyroid disease will be similar to multiple sclerosis, in that it is difficult to collect samples from families that include both parents.

Intrafamilial linkage disequilibrium

The method of linkage disequilibrium analysis (Fig. 2) within families is gaining recognition as a powerful alternative to classical linkage analysis in the search for susceptibility genes in complex disease. The Transmission Disequilibrium Test (TDT) is a test for linkage in the presence of linkage disequilibrium (Spielman et al. 1993). This procedure evaluates the transmission of alleles from a heterozygous parent to one or more offspring. Under simple Mendelian inheritance, we would expect all alleles to have a 50% chance of being transmitted to offspring. If, however, one of the alleles is associated with disease, then the expected transmission to diseased offspring would be greater than 50%. The observed frequency of transmission of an allele is compared with the expected frequency of 0·5
Linkage disequilibrium is the non-random association of specific alleles at two or more neighbouring chromosomal loci. In this figure, marker ‘a’, lying in close chromosomal proximity to the disease mutation, is more likely to be inherited through successive generations on the founder chromosome than is marker ‘b’. In accordance with this diagram, it would be expected that, if allele ‘a’ was found to be associated with disease, it would be likely to be in linkage disequilibrium with a disease-causing mutation.

by the $\chi^2$ statistic; if this is significantly greater than 0.5, linkage in the presence of linkage disequilibrium exists. This approach requires DNA for analysis from only one affected offspring and both parents, although at least one of the parents must be heterozygous at the candidate locus in order that a family be informative. Frequently, one or more unaffected offspring are included in the analyses, as they can be useful in excluding ‘meiotic’ segregation distortion as an alternative explanation to linkage disequilibrium. The TDT may also be used in families with two or more affected offspring. This technique was first used to detect linkage of the insulin gene region in type 1 diabetes (Spiegelman et al. 1993). Early case–control studies had shown population association of the class I variable number of tandem repeats of the insulin gene region with type 1 diabetes (Bell et al. 1984), but subsequent family studies using affected sibpairs failed to confirm linkage (Hitman et al. 1985). More recently, the TDT has confirmed linkage in the presence of linkage disequilibrium for the MHC class II region with Graves’ disease (Heward et al. 1998), with alleles previously shown to be associated with Graves’ disease in various case–control studies. The confirmation of linkages previously not detected by classical linkage analysis in type 1 diabetes and Graves’ disease clearly illustrates the greater power of linkage disequilibrium analysis as a tool in the search for susceptibility loci. This approach is more likely to identify genes of low to modest effect and those with a lower population frequency, because far fewer families are required to reach statistical significance (Spiegelman et al. 1993, Risch & Merikangas 1996). Copeman et al. (1995) showed that, using microsatellite markers in type 1 diabetes, allelic association could be detected with 100 families for a gene with a $\lambda_s$ of 1·1–1·3. In Graves’ disease, 54 informative families allowed us to detect the HLA region with a $P$ value of 0·04 on a locus with a relative risk of 2·7 (Heward et al. 1998). These studies in type 1 diabetes and Graves’ disease suggest that a minimum data set of 50–100 informative families would be required to detect linkage disequilibrium. Practically, collection of family samples for TDT analysis is easier than collecting affected sibpairs for linkage analysis, as only one affected offspring is needed. This technique, however, may be used with even greater power if two or more affected offspring are available. Hence, the TDT can be used in multiplex families that have already been collected for genome-wide searches and linkage analysis. The main limitation of linkage disequilibrium analysis is that it requires knowledge of candidate genes before the test can be performed, and a polymorphism within the gene, or one in strong disequilibrium with it, must be available. Until a large number of genes (up to 100 000) and their polymorphisms have been identified, genome-wide screens will not be possible using linkage disequilibrium analysis. The human genome project, with the probable mapping of all human genes together with the availability of single nucleotide polymorphisms with automated chip-based technology within the next few years, will revolutionise this type of study.

Conclusions

Each of the methods of genetic analysis described in this review has its merits and limitations. Although case–control studies may represent a potential short cut, there is a high risk that false positive results will be obtained as a result of population stratification. This technique is, therefore, inappropriate on its own as a means of identifying susceptibility genes. Case–control studies have an important role in confirming associations detected in family-based studies. Linkage analysis is useful for screening the genome for ‘major’ genes, but has limited power to detect genes of modest effect, particularly when only affected sibpairs without parents are available for analysis. Linkage disequilibrium analysis is gaining recognition as a powerful mechanism for gene identification, but the required knowledge of candidate genes and polymorphisms limits its use, at present, to the investigation of individual loci. As we await mapping of the entire genome and the development of technology capable of rapid analysis of many polymorphisms, a combination of the present available approaches will need to be used to identify all susceptibility loci involved in the development of autoimmune thyroid disease.

Acknowledgements

A Allahabadi is a Smith and Nephew Medical Research Fellow. We acknowledge the support of the Wellcome
Trust (Grant number M/95/3717), Eh Lilly UK and the Trustees of the former United Birmingham Hospitals.

References


Barths ED 1941 Heredity in Graves’ Disease. *Bartels ED 1941 Heredity in Graves’ Disease.*


Cuddihy RM, Duthie CM & Bahn RS 1995 A polymorphism in the extracellular domain of the thyrotropin receptor is highly associated with autoimmune thyroid disease in females. *Thyroid* **5** 89–95.


Received 28 July 1998
Revised manuscript received 2 November 1998
Accepted 9 November 1998