Ann. Hum. Genet. (2000), **64**, 207–213
Printed in Great Britain

Adaptation of the extended transmission/disequilibrium test to distinguish disease associations of multiple loci: the Conditional Extended Transmission/Disequilibrium Test

B. P. C. KOELEMAN*, F. DUDBRIDGE, H. J. CORDELL AND J. A. TODD
Wellcome Trust Centre for Molecular Mechanisms in Disease, University of Cambridge, UK

(Received 20.9.99. Accepted 27.3.00)

SUMMARY

Linkage and association studies in complex diseases are used to identify and fine map disease loci. The process of identifying the aetiological polymorphism, the molecular variant responsible for the linkage and association of the chromosome region with disease, is complicated by the low penetrance of the disease variant, the linkage disequilibrium between physically-linked polymorphic markers flanking the disease variant, and the possibility that more than one polymorphism in the most associated region is aetiological. It is important to be able to detect additional disease determinants in a region containing a cluster of genes, such as the major histocompatibility complex (MHC) region on chromosome 6p21. Some methods have been developed for detection of additional variants, such as the Haplotype Method, Marker Association Segregation Chi-squares (MASC) Method, and the Homozygous Parent Test. Here, the Extended Transmission/Disequilibrium Test is adapted to test for association conditional on a previously associated locus. This test is referred to as the Conditional Extended TDT (CETDT). We discuss the advantages of the CETDT compared to existing methods and, using simulated data, investigate the effect of polymorphism, inheritance, and linkage disequilibrium on the CETDT.

INTRODUCTION

One of the main methodological issues in mapping loci in a common disease is to distinguish between a candidate aetiological allele and an allele at a neighbouring locus that is in linkage disequilibrium (LD) with the aetiological allele. The LD between the loci may confound the analysis. The analysis can be even more complicated if genes of related function are clustered or adjacent on a chromosome and more than one of them have disease-associated alleles. For example, over fifty diseases map to the major histocompatibility complex (MHC), a region containing many genes involved in immunology (Beck & Trowsdale, 1999). Fine mapping of the underlying disease genes is complicated by the strong linkage disequilibrium in the MHC region, which extends over several Mb (Carrington, 1999). Furthermore, some MHC genes have been shown to have a joint effect on disease susceptibility. For example, from a combination of biological and genetic studies it is now established that the HLA class II genes *DRB1* and *DQB1* are the primary determinants in type 1 diabetes (Cucca & Todd, 1996).

Methods have been developed to detect further genetic factors interacting and/or in linkage

Correspondence: John A. Todd, Wellcome Trust Centre for Molecular Mechanisms in Disease, Wellcome Trust/MRC Building, Addenbrookes Hospital, Hills Road, Cambridge CB2 2XY UK.
E-mail: john.todd@cimr.cam.ac.uk
* Current address: Department of Immunohematology and Bloodbank, Leiden University Medical Center, PO Box 9600 Leiden, The Netherlands.

Software for haplotyping and association analysis (T_{sp}, ETDT, conditional ETDT) can be obtained by anonymous ftp from ftp://ftp-gene.cimr.cam.ac.uk or by contacting F.D. (frank@gene.cimr.cam.ac.uk).

disequilibrium with a candidate disease gene. On haplotypes containing all the variants that determine susceptibility, the frequency of neutral alleles at an additional locus is expected to be equal in cases and controls. This intuitive assumption has been used in several studies in which cases and controls were matched for high risk genotypes at a candidate locus or loci, and were thus expected to have equal frequencies of neutral marker alleles at a secondary marker locus, even when the marker was in linkage disequilibrium with the candidate disease genes (McGarry *et al.* 1997; Moghaddam *et al.* 1998; Singal *et al.* 1998; Lie *et al.* 1999*a*). A test based on this principle, the Haplotype Method (HM), was developed and applied to amino acid variations in HLA class II *DQA1-DQB1* genes in type 1 diabetes and rheumatoid arthritis (Valdes & Thomson, 1997; Valdes *et al.* 1997). The Haplotype Method is a test for homogeneity of relative allele frequencies at a secondary locus on haplotypes identical for alleles at a candidate locus. Application of the haplotype method to cases and affected family-based (AFBAC) controls (Thomson, 1995) prevents possible bias in 'classical' case control studies due to population stratification, but requires random mating and Hardy–Weinberg equilibrium (HWE) in the parents of cases. These assumptions are not required by the Transmission/Disequilibrium Test (TDT) (Spielman *et al.* 1993) version of the Homozygous Parent Test (HPT) (Robinson *et al.* 1993; Lie *et al.* 1999*b*). The HPT is a TDT, which counts transmissions of marker alleles only from parents homozygous at the candidate disease locus. Unfortunately, this restriction may lead to considerable loss of information when the candidate locus has more than two alleles. For example, when HPT is applied to the HLA class II genes in type 1 diabetes, only 10–15% of the parents are homozygous for *DR/DQ* haplotypes, and therefore very large clinical resources are required (Lie *et al.* 1999*b*).

The restriction to homozygous parents is not made in the MASC method, nor in the logistic model described by Thomas *et al.* (Clerget-Darpoux *et al.* 1988; Thomas *et al.* 1995). The MASC method has been applied to test for interactive effects and two locus disease models by a goodness-of-fit test (Dizier *et al.* 1994). Thomas *et al.* used an empirical Bayes model within a conditional logistic regression modelling of the disease risk as described by Self *et al.*, to test for effects at HLA class I alleles additional to *DRB1* on the susceptibility of type 1 diabetes (Self *et al.* 1991; Langholz *et al.* 1995; Thomas *et al.* 1995).

Logistic regression has also been used to extend the TDT to multi-allelic markers, giving the Extended TDT (ETDT) (Sham & Curtis, 1995). Here, the ETDT is adapted to test for an effect at a secondary locus or marker conditional on the association of a candidate disease locus in case-parent triads. Considering gametic haplotypes of a candidate (or established) disease locus, and a neutral marker, it is expected that haplotypes with identical alleles at the candidate disease locus, but different alleles at the marker, have equal transmission probabilities. ETDT transmission probabilities can be estimated and, in our adaptation, can be tested for equality using a likelihood-ratio test, which we call Conditional ETDT (CETDT). A significant difference in transmission of haplotypes identical at the candidate locus, but different at the secondary locus, provides evidence for the involvement of either the secondary locus or a locus in linkage disequilibrium with it. In the next section we recall the mathematical model of the ETDT and describe its adaptation into the CETDT. We then use simulated data to investigate the effect of several inheritance models on the power of the CETDT compared to the HPT.

Likelihood model for case-parent triads

Consider a simplex family with genotype information for both parents and a single affected offspring. Four genotype configurations are possible for the offspring. Likelihood methods for these case-parent triads have been described previously (Self *et al.* 1991; Schaid & Sommer, 1993; Schaid, 1996). The probability of transmitting genotype c to case n, conditional on

parental genotypes f and m and affection status A, is given by Bayes' theorem as:

$$\Pr(c|f,m,A) = \frac{\Pr(A|c,f,m)\Pr(c|f,m)\Pr(f,m)}{\sum_{g\in G_n}\Pr(A|g,f,m)\Pr(g|f,m)\Pr(f,m)}, \quad (1)$$

where G_n is the set of possible offspring genotypes from the parents. The mating type frequency $\Pr(f,m)$ cancels, and the $\Pr(g|f,m)$ are the mendelian probabilities of case n having genotype g given that its parents have genotypes f and m. Therefore, equation (1) is determined by $\Pr(A|g,f,m)$, that is the probabilities of affection given genotype g and parental genotypes f and m. It can be assumed that $\Pr(A|g,f,m)$ depends only on the relative risk for disease of genotype g, which we denote by R_g, and which can be specified relative to an arbitrary 'baseline' genotype g_0. Furthermore, under a multiplicative inheritance model (Risch, 1990), the relative risk of the genotype consisting of alleles i and k is the product of the relative risks of allele i and k ($R_{ik} = R_i\ R_k$), so that for parents with genotypes ij and kl that transmit alleles i and k, equation (1) simplifies to:

$$\Pr(ik|ij,kl,A) = \frac{R_i}{R_i+R_j}\cdot\frac{R_k}{R_k+R_l} \quad (2)$$

If we write $\Pr(i|ij,A) = R_i/(R_i+R_j)$ for the probability of transmission of allele i from a parent with genotype ij, then under multiplicative inheritance at the disease locus,

$$\Pr(ik|ij,kl,A) = \Pr(i|ij,A)\,.\,\Pr(k|kl,A) \quad (3)$$

so that the parental transmissions may be regarded as independent in this case. Furthermore the parental transmissions are also independent under the null hypothesis of no association ($R_i = R_j$ for all i,j), regardless of the inheritance model.

The log-likelihood for the ETDT (Sham & Curtis, 1995) is

$$L = \sum_{i,j} n_{ij}\ln(R_i/(R_i+R_j)), \quad (4)$$

where n_{ij} is the number of parents with genotype ij that transmit allele i, and i and j range over all the alleles at the locus. The likelihood is maximised over all R_i, giving log-likelihood L_1, and compared with the likelihood under the null hypothesis of no association (all R_i equal, log-likelihood L_0). The likelihood-ratio statistic $2(L_1-L_0)$ has asymptotically the χ^2 distribution with degrees of freedom equal to the number of alleles minus one.

Transmission probabilities for two disease genes

We now consider two disease genes, denoted by D and S, and transmissions of gametic haplotypes. The conditional transmission probability for two haplotypes of these loci has the same form as equation (1). For one parent with genotypes i_Dj_D and i_Sj_S, and the other parent with genotypes k_Dl_D and k_Sl_S, the conditional probability of the offspring having genotypes i_Dk_D and i_Sk_S is

$$\Pr(i_Dk_D,i_Sk_S|i_Dj_D,i_Sj_S,k_Dl_D,k_Sl_S,A) = \frac{R_{i_Dk_Di_Sk_S}}{R_{i_Dk_Di_Sk_S}+R_{i_Dl_Di_Sl_S}+R_{j_Dk_Dk_Sk_S}+R_{j_Dl_Dj_Sl_S}} \quad (5)$$

if the parental haplotypes are i_D-i_S, j_D-j_S, k_D-k_S and l_D-l_S, otherwise this probability is zero. For independence of the parental transmissions, we require

$$\begin{aligned} R_{i_Dk_Di_Sk_S} &= R_{i_Di_S}R_{k_Dk_S} \\ R_{i_Dl_Di_Sl_S} &= R_{i_Di_S}R_{l_Dl_S} \\ R_{j_Dk_Dj_Sk_S} &= R_{j_Dj_S}R_{k_Dk_S} \\ R_{j_Dl_Dj_Sl_S} &= R_{j_Dj_S}R_{l_Dl_S} \end{aligned} \quad (6)$$

In general this only holds under the following conditions:

1. No disease association at locus D nor at locus S;
2. Association of locus D under a multiplicative inheritance model, and no association at locus S;
3. No association at locus D, and association of locus S under a multiplicative inheritance model;
4. Association of both locus D and locus S under a multiplicative inheritance model at each locus and a multiplicative model for the interaction between the loci (Risch, 1990).

The ETDT log-likelihood may be constructed for haplotypes as in equation (4), using (for example) the count of parents with genotype $i_D j_D$ and $i_S j_S$ that transmit haplotype $i_D i_S$, and can be maximized over all haplotype relative risks.

Note that because the transmission probability is conditional upon deduction of the four gametic haplotypes, we may ignore the possibility of recombination between D and S in this analysis. Furthermore, by conditioning on the four gametic haplotypes no HWE has to hold in parental genotypes.

Conditional ETDT (CETDT)

We now let locus D represent a candidate or previously associated locus for the primary effect in a region associated with disease, which we term the *conditioning locus*. Locus S represents a candidate locus for a further effect, which we term the *test locus*. We test for a further effect of S independent of locus D. Therefore, under the null hypothesis all haplotypes with identical alleles at locus D have equal relative risks (transmission probabilities). The maximum log-likelihood L_1 is again obtained by maximising (4) over all haplotype relative risks, but the null log-likelihood L_0 is now obtained by maximising (4) under the constraint that

$$R_{i_D i_S} = R_{i_D j_S}$$

for each allele i_D of D, i_S and j_S of S.

As a multi-allelic test for a further effect of locus S conditional on locus D, the Conditional ETDT statistic $2(L_1\text{-}L_0)$ has asymptotically the χ^2 distribution with degrees of freedom equal to the number of D-S haplotypes minus the number of D alleles. Because we have considered the parental trqansmissions independently, the CETDT statistic follows the χ^2 distribution only when the inheritance model of the conditioning locus is multiplicative, or if it has no disease association.

An alternative to the global test is to test whether two specific haplotypes have different transmission probabilities, for example $D1$-$S1$ versus $D1$-$S2$. In this case the null hypothesis is that the two haplotypes have the same relative risk, so we obtain L_0 by maximising (4) under the constraint that $R_{1\text{–}1} = R_{1\text{–}2}$. The test statistic $2(L_1 - L_0)$ now has asymptotically the χ^2 distribution with one degree of freedom. We term this test the pairwise ETDT (PETDT).

Linkage disequilibrium between the disease alleles of D and S is expected to reduce the power of these tests, as we demonstrate in our simulations. Furthermore, CETDT requires that both transmitted and non-transmitted two-locus haplotypes are determined, which is not needed for the HPT. The haplotyping requirement will result in discarding those families where haplotyping is ambiguous. For example, when both parents are heterozygous for identical alleles at one locus, one can only determine haplotypes when the sibling is homozygous. Inclusion of such families may lead to an increased type 1 error as we have described elsewhere (Dudbridge *et al*. 2000). Therefore, those families are discarded using CETDT, but included in the HPT when the parents are homozygous for identical alleles at the candidate disease locus.

With highly polymorphic systems such as the HLA class II genes some haplotypes may be very infrequent leading to sparse cell counts and reduced power. Grouping of haplotypes with counts less than a certain threshold can partially solve this problem, but this is only valid when haplotypes with identical alleles at the conditioning locus are grouped together. However, grouping haplotypes will also reduce power when haplotypes with opposing effects are grouped. Furthermore, the data may remain sparse even after grouping, and in this case it may be preferable to exclude rare haplotypes from the analysis. Finally, sparse cell counts can also lead to inaccurate approximation of the χ^2-distribution. Therefore, our program has the ability to calculate empirical p-values, which alleviates these problems.

Simulations

Genotypes for loci D and S were simulated for simplex families with one affected offspring. Allele 1 of both D and S was designated the

Table 1. *Effect of number of alleles at locus D**

Test	#Allele D	Power $\alpha = 0.01$	$\alpha = 0.001$	-log(p) (S.E.)
HPT	2	1	1	10.8 (2.5)
CETDT	2	1	1	9.0 (2.7)
HPT	3	0.83	0.45	3.3 (1.3)
CETDT	3	1	0.99	8.6 (2.7)
HPT	4	0.34	0.10	1.7 (0.9)
CETDT	4	1	0.93	8.7 (2.7)

* Case-parent triads were simulated under a multiplicative inheritance model for the conditioning (D) and test (S) locus, as well as a multiplicative interaction between the loci. Disease alleles at either D or S are sufficient for development of disease. Power is expressed as the fraction of replicates that produce a p-value smaller than α, and the mean -$\log_{10}$(p-value) is also shown with its standard error (S.E.).

Table 2. *Effect of inheritance model**

Test	Power $\alpha = 0.01$	$\alpha = 0.001$	-log(p) (S.E.)	Model
HPT	0.37	0.12	1.8 (1.0)	Nec
CETDT	0.95	0.85	4.9 (2.0)	Nec
HPT	0.58	0.24	2.3 (1.1)	Dom
CETDT	1	1	8.6 (2.6)	Dom
HPT	0.57	0.24	2.3 (1.1)	Rec
CETDT	1	1	8.6 (2.6)	Rec

* Case-parent triads were simulated as in table 1, but with different inheritance models for locus S. Model 'Nec' indicates that the disease allele of locus S is not sufficient for development of disease, model 'Dom' indicates a dominant inheritance at locus S, and 'Rec' indicates a recessive inheritance at locus S. Power is expressed as the fraction of replicates that produce a p-value smaller than α, and the mean -$\log_{10}$(p-value) is also shown with its standard error (S.E.).

Table 3. *Effect of linkage disequilibrium**

Test	K_1	Power $\alpha = 0.01$	$\alpha = 0.001$	-log(p) (S.E.)
HPT	0.5	0.49	0.15	2.0 (0.9)
CETDT	0.5	0.86	0.64	3.8 (1.8)
HPT	0.8	0.02	0	1.0 (0.5)
CETDT	0.8	0.30	0.1	1.6 (1.1
HPT	1	0	0	0
CETDT	1	0	0	0

* Case-parent triads were simulated under a multiplicative model as in Table 1, and with linkage disequilibrium between allele 1 of both the conditioning locus and the test locus. Linkage disequilibrium is quantified by the parameter K_1, following Thomson, 1983. Power is expressed as the fraction of replicates that produce a p-value smaller than α, and the mean -$\log_{10}$(p-value) is also shown with its standard error (S.E.).

'disease' variant with relative risk of 10 for allele 1 of D, and 4 for allele 1 of S; all other alleles were given relative risk of 1. The population frequency of both disease variants (q_D and q_S) was 0.1, and the phenocopy frequency was 0.0005. In the simulations where locus D had more than 2 alleles, the non-disease alleles had equal frequencies. Linkage disequilibrium was simulated between allele 1 of both loci with the parameter K_1 quantifying the frequency of allele 1 at locus S on haplotypes with allele 1 at locus D (Thomson, 1983). Complete linkage disequilibrium existed when K_1 was either 1 or 0, and equilibrium when $K_1 = q_D = q_S$.

Tables 1–3 show the results for different inheritance and interaction models, as well as for different numbers of alleles of the conditioning locus D. Table 1 shows the effect of increasing numbers of alleles at the conditioning locus D. Haplotyping results in loss of power for CETDT, since some families have to be discarded when haplotyping is ambiguous. This effect is most prominent when the simulation is performed with D having only two alleles, in which case HPT has comparable power to CETDT. We observe that the power of HPT decreases dramatically with increasing number of alleles at D, due to increasingly lower numbers of homozygous parents (informative parents in the HPT).

Different inheritance models for the two loci gave reduced power for all methods, as expected, with the most dramatic decrease for a necessity model, in which the disease allele at locus S is not sufficient for development of disease (Table 2). Similarly, linkage disequilibrium between the loci decreases the power of all methods (Table 3). However, CETDT remained consistently more powerful than HPT.

For all models type 1 error was estimated by setting the RR of locus S to 1, and was found to be close to the nominal value, provided that rare haplotypes were excluded or grouped as described in the previous section.

discussion

We have adapted the ETDT to detect further disease loci linked or unlinked to an established or suspected disease gene. This study illustrates that the Conditional ETDT is more powerful than the Homozygous Parent Test, but only when the candidate locus has more than two alleles. The lower power of HPT is caused mainly by its limitation to parents homozygous for the disease locus. The CETDT overcomes this limitation and as a result, family resources are more efficiently used, making it more feasible to detect low penetrance disease genes or population-specific associations. When the candidate locus is biallelic, the benefit of including heterozygous parents is counteracted by the loss of families due to haplotyping ambiguities.

We expect that other methods for detecting an effect of a second locus that do not have the restriction to homozygous parents, such as the Haplotype Method and the MASC method, have similarly increased power over the HPT. However, compared to HM, CETDT does not require Hardy–Weinberg equilibrium in parental genotypes, random mating, nor knowledge of the ascertainment scheme. Like other TDT-based methods, CETDT is not affected by recent population stratification. The MASC method has been used to perform a goodness-of-fit test on a two-locus model, but requires specification or estimation of model parameters such as gene frequencies and linkage disequilibrium parameters (Dizier *et al*. 1994). However, the CETDT is based on the elegant ETDT model, which considers only the haplotype relative risks and has no nuisance parameters. When the genetic model is specified accurately, the MASC method may give more robust estimates of haplotype relative risks, and can be easily adapted to different inheritance models at the conditioning locus. Furthermore, the MASC method can be used with extended families and can integrate information on segregation and linkage. The model described by Thomas *et al*. calculates main effects at each locus, haplotype effects, and combined effects, rather than an overall statistic for association (Thomas *et al*. 1995; Langholz *et al*. 1995). This test can be used, like the MASC method, for a more in depth quantification of the two locus model, whereas the CETDT can be used as a straightforward test for association conditional on a candidate locus, which requires fewer parameters.

The relative risk estimated by ETDT is a measure of disease predisposition that is more accurate than TDT transmission ratio, because it is robust to Hardy–Weinberg disequilibrium in the parents. TDT transmission ratios are calculated from marginal transmission counts and do not condition on parental genotype. Thus, if a neutral haplotype always occurs in parents also carrying a susceptible haplotype, its transmission ratio will be reduced and it will seem to be protective. However, the ETDT estimates haplotype relative risks based on a complete table of transmission and the estimates are more robust to this kind of distortion.

However, the estimates of haplotype relative risks are influenced by recombination between marker and disease locus, LD, and discrepancy between frequency of disease alleles and frequency of marker alleles (Schaid, 1996). When conditioning does not take place on a candidate locus itself, but on a nearby marker, it is expected that the estimate of relative risk will be affected. As with the Haplotype Method, testing nearby markers conditional on the most associated variant can indicate whether or not the disease determinants have been found (Valdes & Thomson, 1997). Similarly, when a marker closely linked to the test locus is used, the estimation of the relative risks is also influenced by recombination, LD and allele frequency differences, and results in reduced power to detect an effect of the test locus.

Both HPT and the CETDT presented here are most powerful when a multiplicative disease model holds at the test locus, and when the interaction between the conditioning and test loci is multiplicative. Potentially, further advances may be made by application of genotype relative risk methods, which may allow modelling of complex interaction of different

disease alleles, or different disease loci (Self *et al.* 1991; Schaid, 1996; Dudbridge *et al.* 2000).

In the accompanying paper we apply the CETDT to experimental data from the analysis of the *IDDM1* locus in the MHC in type 1 diabetic families in order to decipher the contributions of various MHC loci to the *IDDM1* effect.

This study was supported by the Wellcome Trust, and by a TMR Marie Curie Research Training Grant of the European Commission.

REFERENCES

Beck, S. & Trowsdale, J. (1999) Sequence organisation of the class II region of the human MHC. *Immunol. Rev.* **167**, 201–210.

Carrington, M. (1999) Recombination within the human MHC. *Immunological Reviews* **167**, 245–256.

Clerget-Darpoux, F., *et al.* (1988) A new method to test genetic models in HLA associated diseases: the MASC method. *Ann. Hum. Genet.* **52**, 247–258.

Cucca, F. & Todd, J. A. (1996) HLA susceptibility to type 1 diabetes: methods and mechanisms. In: *HLA and MHC: genes, molecules and function* (eds. Browning MJ, McMichael AJ), pp. 23–60. Oxford: BIOS Scientific Publishers Ltd.

Dizier, M. H. *et al* (1994) Interactive effect of two candidate genes in a disease: extension of the marker-association-segregation χ^2 method. *Am. J. Hum. Genet.* **55**, 1042–1049.

Dudbridge, F., Koeleman, B. P. C., Todd, J. A. & Clayton, D. G. (2000) Unbiased application of the Transmission/Disequilibrium Test to multilocus haplotypes. *Am. J. Hum. Genet.* **66**, 2009–2012.

Langholz, B., *et al.* (1995) Variation in HLA-associated risks of childhood insulin-dependent diabetes in the Finnish population: I. Allele effects at A, B, and DR loci. DiMe Study Group. Childhood Diabetes in Finland. *Genet. Epidemiol.* **12**, 441–453

Lie, B. A., *et al.* (1999*a*) HLA associations in insulin-dependent diabetes mellitus: no independent association to particular DP genes. *Hum. Immunol.* **55**, 170–175.

Lie, B. A., *et al.* (1999*b*). Predisposition to type 1 diabetes, linked to the HLA complex (IDDM1), includes at least one non-class II gene. *Am. J. Hum. Genet.* **64**, 793–800.

McGarry, F., *et al.* (1999) The -308.1 polymorphism in the promoter region of the tumor necrosis factor gene is associated with ankylosing spondylitis independent of HLA-B27. *J. Rheumatol.* **26**, 1110–1116.

Moghaddam, P., *et al.* (1998). Genetic structure of IDDM1: two separate regions in the major histocompatibility complex contribute to susceptibility or protection. Belgian Diabetes Registry. *Diabetes* **47**, 263–269.

Risch, N. (1990). Linkage strategies for genetically complex traits. I. Multilocus models. *Am. J. Hum. Genet.* **46**, 222–228.

Robinson, W. P., *et al.* (1993). Homozygous parent affected sib pair method for detecting disease predisposing variants: application to insulin dependent diabetes mellitus. *Genet. Epidemiol.* **10**, 273–288.

Schaid, D. J. & Sommer, S. S. (1993). Genotype relative risks: methods for design and analysis of candidate-gene association studies. *Am. J. Hum. Genet.* **55**, 1114–1126.

Schaid, D. J. (1996). General score tests for associations of genetic markers with disease using cases and their parents. *Genet. Epidemiol.* **13**, 423–449.

Self, S. G., *et al.* (1991). On estimating HLA/disease association with application to a study of aplastic anemia. *Biometrics.* **47**, 53–61.

Sham, P. C. & Curtis, D. (1995). An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Ann. Hum. Genet.* **59**, 323–36

Singal, D. P., *et al.* (1998). D6S273 microsatellite polymorphism and susceptibility to rheumatoid arthritis. *Tissue Antigens* **52**, 353–358.

Spielman, R. S., *et al.* (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52**, 506–516.

Thomson, G. (1983) Investigation of the mode of inheritance of the HLA associated diseases by the method of antigen genotype frequencies among disease individuals. *Tissue Antigens* **21**, 81–104.

Thomas, D., *et al.* (1995) Variation in HLA-associated risks of childhood insulin-dependent diabetes in the Finnish population: II. Haplotype effects. DiMe Study Group. Childhood Diabetes in Finland. *Genet. Epidemiol.* **12**, 455–466.

Thomson, G. (1995) Mapping disease genes: family-based association studies. *Am. J. Hum. Genet.* **57**, 487–498.

Valdes, A. M. & Thomson, G. (1997) Detecting disease-predisposing variants: the haplotype method. *Am. J. Hum. Genet.* **60**, 703–716.

Valdes, A. M., *et al.* (1997) HLA class II DR-DQ amino acids and insulin-dependent diabetes mellitus: application of the haplotype method. *Am. J. Hum. Genet.* **60**, 717–728.