**European Journal of Human Genetics (2000) 8,** 545–551

www.nature.com/ejhg

ARTICLE

# Pedigree tests of transmission disequilibrium

Gonçalo R Abecasis, William OC Cookson and Lon R Cardon

*Wellcome Trust Center for Human Genetics, University of Oxford, UK*

**High-resolution mapping is essential for the positional cloning of complex disease genes. In outbred populations, linkage disequilibrium is expected to extend for short distances and could provide a powerful fine-mapping tool. Current family-based association tests use nuclear family members to define allelic transmission and controls, but ignore other types of relatives. Here we construct a general approach for scoring allelic transmission that accommodates families of any size and uses all available genotypic information. Family data allows for the construction of an expected genotype for every non-founder, and orthogonal deviates from this expectation are a measure of allelic transmission. These allelic transmission scores can be used to extend previously described tests of linkage disequilibrium for dichotomous or quantitative traits. Some of these tests are illustrated, together with a permutation framework for estimating exact significance levels. Simulation studies are used to investigate power and error rates of the approach. As a practical application, the method is used to investigate the relationship between circulating angiotensin-1 converting enzyme (ACE) levels and polymorphisms in the *ACE* gene using previously published data.** *European Journal of Human Genetics* (2000) **8**, 545–551.

**Keywords:** Linkage disequilibrium; gene mapping; SNP; quantitative trait; ACE; large pedigrees

## Introduction

Disease genes can be identified on the basis of their location, even when the underlying biochemical pathways are unknown.[1] However, in complex disease, confidence intervals surrounding the highest linkage typically extend several centimorgans[2,3] and are too broad for positional cloning. Although the additional information in large pedigrees[4–6] may provide more power than small nuclear families,[7,8] the resolution of linkage analyses remains limited even in extremely large pedigrees.[2]

Increasingly large numbers of single-nucleotide polymorphisms (SNPs) are available in public and private databases[9] and high-throughput methods for their genotyping are emerging.[10,11] While it is unlikely that additional markers will improve the resolution of traditional linkage analyses, linkage disequilibrium mapping should have greater resolution and will benefit from dense SNP maps.[12]

Family-based association tests can distinguish disequilibrium from other types of association due to population substructure. Previously described methods use parents[13–17] or siblings[13,18–20] to construct controls that are robust to stratification and are applicable to dichotomous or quantitative traits.

Intuitively, larger pedigrees include more information on population substructure than nuclear families, but current methods of scoring allelic transmission consider only parents and siblings and ignore other relatives. Linkage disequilibrium tests that accommodate not only nuclear family data, but also other relatives, will be important for refining gene location in many current studies.

Here, we describe a fast, computationally efficient method of scoring allelic transmission in extended pedigrees that considers not only parents and siblings but also all available ancestors. The method can be used in tests of linkage disequilibrium for quantitative or qualitative traits. As an example, power and type I error rates of some of these tests

**Correspondence: Gonçalo Abecasis, Wellcome Trust Center for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, United Kingdom. Tel: +44 1865 287 597; Fax: +44 1865 287 650; E-mail: goncalo@well.ox.ac.uk**
**Received 20 December 1999; revised 2 March 2000; accepted 9 March 2000**

are explored by simulation. Also, using a previously published data set,[21,22] which includes several extended pedigrees, we investigate the relationship between polymorphisms in the *ACE* gene and circulating ACE levels.

## Methods

### Scoring allelic transmission

In the present method, we assume no segregation distortion (meiotic drive), so that heterozygous parents are equally likely to transmit either allele to their offspring. In the presence of population stratification, the distribution of genotypes (and phenotypes) in each sub-population may be different. Tests based on allelic transmission should allow for the most severe form of population admixture where each individual is drawn from a different sub-population. Typically, family members are used to construct an expected genotype for each individual, and deviations from this expectation (due to excess transmission of a particular marker allele) are measured.[23] Under the null hypothesis of no linkage disequilibrium, family data allows appropriately constructed expectations to be unbiased in all sub-populations. Corresponding deviates are equally likely to indicate a surplus or deficit of transmission for any allele, whatever the genetic make-up of each sub-population. Therefore, deviates can be used to construct tests of linkage disequilibrium.

Consider a candidate di-allelic marker, M, with alleles arbitrarily designated as '1' (with frequency $p$) and '2' (with frequency $q = 1-p$). Given a set of $N$ families, define the marker phenotype $m_{ij}$ and the genotype score $g_{ij}$ for the $j^{th}$ individual in the $i^{th}$ family as

$m_{ij}$ = number of '1' alleles at locus M, and

$g_{ij} = m_{ij} - 1$.

Let $M_{ij}$ and $F_{ij}$ represent specific indexes for the male and female parents of the $j^{th}$ individual in the $i^{th}$ family. If an individual has no observed ancestors in the pedigree, leave $M_{ij}$ and $F_{ij}$ undefined. For convenience, define the following sets for each family $i$:

(i) the set of genotyped individuals as $G_i = \{k | g_{ik}$ is known$\}$;

(ii) the set of founders, which includes all individuals with no observed ancestors in the pedigree, as $A_i = \{k | \{M_{ik}, F_{ik}\} = \emptyset\}$;

(iii) The set of full siblings for each non-founder, $j \in A_i$, as $S_{ij} = \{k | M_{ik} = M_{ij}\} \cap \{k | F_{ik} = F_{ij}\}$.

Using standard set notation, the number of founders in a pedigree is $|A_i|$ and the number of individuals in a sibship is $|S_{ij}|$, of which $|S_{ij} \cap G_i|$ are genotyped.

In nuclear families, define allelic transmission in terms of an expected genotype $b_{ij}$ and deviate $w_{ij}$ where

$$b_{ij} = b_i = \begin{cases} \dfrac{g_{F_{ij}} + g_{M_{ij}}}{2} & \text{if parental genotypes are known} \\ \dfrac{\sum_{k \in S_{ij} \cap G_i} g_{ik}}{|S_{ij} \cap G_i|} & \text{otherwise} \end{cases}$$

and $w_{ij} = g_{ij} - b_{ij}$.[13]

Positive values of $w_{ij}$ indicate excess transmission of allele '1', while negative values indicate excess transmission of allele '2'. Whenever both parents are homozygous at the marker locus, $w_{ij} = 0$, so that $w_{ij} \neq 0$ implies that at least one parent is heterozygous. Also, in the absence of segregation distortion or selection, $E(w_{ij}) = 0$.

Although the previous definition could be used in extended families, it would ignore the information available from relatives other than parents or siblings. We will now present a more general algorithm for defining $b_{ij}$ that considers all available information (although, our interest remains focused on $w_{ij} = g_{ij} - b_{ij}$ as a measure of allelic transmission).

Traverse pedigree $i$, for $j = 1 \ldots n_i$ so that each individual $j$ is preceded by all his ancestors:

(1) If $j$ is a genotyped founder, $j \in (A_i \cap G_i)$, assign $b_{ij} = g_{ij}$ and $w_{ij} = 0$. Proceed to individual $j + 1$.

(2) If $b_{M_{ij}}$ and $b_{F_{ij}}$ are defined, assign

$$b_{ij} = \frac{b_{M_{ij}} + b_{F_{ij}}}{2} \text{ and } w_{ij} = g_{ij} - b_{ij}.$$

In a pedigree where all founders are typed, $A_i \subset G_i$, this is equivalent to

$$b_{ij} = \sum_{k \in A_i} 2\varphi_{ijk}\, g_{ik}$$

for every individual (where $\varphi_{ijk}$ is the kinship coefficient between individuals $j$ and $k$ in family $i$). Proceed to individual $j + 1$.

(3) Otherwise, if $j$ is genotyped, assign

$$b_{ij} = \frac{\sum_{k \in S_{ij} \cap G_i} g_{ik}}{|S_{ij} \cap G_i|} \text{ and } w_{ij} = g_{ij} - b_{ij}.$$

Note that, when no genotyped siblings are available, $b_{ij}$ and $w_{ij} = 0$. Proceed to individual $j + 1$.

(4) Finally if $j$ is not genotyped and $b_{F_{ij}}$ or $b_{M_{ij}}$ are undefined, leave $b_{ij}$ and $w_{ij}$ undefined. Proceed to individual $j + 1$.

This definition uses as many ancestral chromosomes as possible to define each $b_{ij}$, but makes no attempt to infer missing genotypes, as this is fraught with pitfalls.[24] Note that $w_{ij} \neq 0$ implies that individual $j$ has at least one heterozygous ancestor. The algorithm can accommodate pedigrees of any practical size but requires negligible computing resources.

### Transmission disequilibrium tests

Using this general approach for scoring allelic transmission, we now show how the resulting $w_{ij}$ values may be used in different measures of linkage disequilibrium.

### Discrete traits

If a single offspring per family is considered, the sum of all $w_{ij}$ has mean zero and variance $\sum_{ij} w_{ij}^2$.

Thus, the statistic

$$T_{TDT} = \frac{\sum_{ij} w_{ij}}{\sqrt{\sum_{ij} w_{ij}^2}}$$

is asymptotically distributed as standard normal in the absence of linkage disequilibrium, and is analogous to the widely-used TDT test of Spielman *et al*.[16]

If multiple affected offspring are considered in each family, transmissions to family members are not independent and the set of transmissions from each parent should be treated as unit.[25] In extended families it is not practical to separate transmission from each founder, but transmissions in each family may be treated as independent sets. In this case, the statistic

$$T_{affected} = \frac{\sum_i \sum_j w_{ij}}{\sqrt{\sum_i \left(\sum_j w_{ij}\right)^2}}$$

is asymptotically distributed as standard normal in the absence of linkage disequilibrium (see Martin *et al*.[25]).

If affected and unaffected offspring are considered in each family, the statistic

$$T_{all} = \frac{\sum_i \sum_j a_{ij} w_{ij}}{\sqrt{\sum_i \left(\sum_j a_{ij} w_{ij}\right)^2}},$$

is asymptotically distributed as standard normal in the absence of linkage disequilibrium and includes information on all family members ($a_{ij}$ is an indicator variable for affection status, defined as $a_{ij} = 1$ for affected individuals, $a_{ij} = -1$ for unaffected individuals and $a_{ij} = 0$ otherwise. Other definitions for $a_{ij}$, such as asymmetric weights for affected and unaffected individuals could be used. Rabinowitz[17] considers weights based on quantitative phenotypes).

### Quantitative traits

Straightforward linear models express expected phenotype scores, $\mu_{ij}$, as a function of the overall population mean, $\mu$, and the marker genotype. Allison[14] (TDTQ5) suggested a linear model with indicator variables for each mating type as controls for population stratification. Abecasis *et al*[13] showed a more parsimonious model can be defined using expected genotype scores, $b_{ij}$, and corresponding deviates, $w_{ij}$:

$$\mu_{ij} = \mu + \beta_b b_i + \beta_w w_{ij}.$$

If a single offspring per family is considered, define $R^2_{\beta_w=0}$ as the residual sum of squares when the model is fitted with $\beta_w = 0$ and $R^2$ as the residual sum of squares when the model is fitted with no constraints. Then the goodness of fit statistic

$$F_{QTL} = \frac{R^2 - R^2_{\beta_w=0}}{(1 - R^2_{\beta_w=0})/(K-3)}$$

is distributed as $F$ with one and K – 3 degree of freedom in the absence of linkage disequilibrium.

Simple linear regression approaches, such as this one, do not account for familial correlations, and are only appropriate when a single offspring is considered in each family.[26] When multiple offspring per family are considered, variance component models offer a framework for describing family data.[27] In this case, $\Omega_i$, a $n_i$ x $n_i$ matrix of expected variances and covariances for family $i$, may be specified according to the classical biometrical model.[28]

$$\Omega_{ijk} = \begin{cases} \sigma_a^2 + \sigma_g^2 + \sigma_e^2 & \text{if } j = k \\ \pi_{ijk}\sigma_a^2 + 2\varphi_{ijk}\sigma_g^2 & \text{if } j \neq k \end{cases}$$

where $\sigma_a^2$ is the additive genetic variance of the QTL, $\sigma_g^2$ is the variance attributable to polygenes, $\sigma_e^2$ is the residual environmental variance, and $\pi_{ijk}$ is the proportion of alleles shared identical-by-descent (IBD) at the marker locus between individuals $j$ and $k$ in family $i$.

Note that the random effect parameters ($\sigma_a^2$, $\sigma_g^2$ and $\sigma_e^2$) account for familiarity and linkage effects, while stratification and linkage disequilibrium are modeled in the fixed effect parameters ($\beta_b$ and $\beta_w$). The likelihood of the data can be expressed in terms of the observed phenotypes $\mathbf{y}_i$ and the random effects in $\Omega_i$ and the linear model as

$$L = \prod (2\pi)^{-n_i/2} |\Omega_i|^{-1/2} e^{-1/2[(y_i-\mu_i)'\Omega_i^{-1}(y_i-\mu_i)]}.$$

Define $L_o$ as the maximum likelihood of the data when $\beta_w = 0$ and $L_1$ as the maximum likelihood of the data when there are no constraints on the parameters. Then, under the assumption of multivariate normality,

$$\chi^2_{QTL} = 2\ ln\ (L_1/L_o)$$

is asymptotically distributed as chi-squared with one degree of freedom.[13,19]

### Permutation framework

Relying on asymptotic theory for estimating significance levels is undesirable when small samples are considered or when a quantitative trait is not distributed as multivariate

normal.[27,29] As described above, the pattern of allelic transmission in each family can be expressed as $\mathbf{w}_i = [w_{i1}, w_{i2}, \ldots]$. A consequence of the assumption of Mendelian segregation is that, in the absence of linkage disequilibrium, the observed pattern of transmission ($\mathbf{w}_i$) and its mirror image ($-\mathbf{w}_i$) are equally likely. Construct the random permutation $P_r$ of any set of $N$ families by replacing each $\mathbf{w}_i$ with itself or $-\mathbf{w}_i$ with equal probability, so that for any given data set there are $2^N$ different permutations of the data. In the absence of linkage disequilibrium, the distributions of $T_{TDT}$, $T_{aff}$, $T_{all}$, $F_{QTL}$ and $\chi^2_{QTL}$ can be estimated by sampling a large number of these permutations at random.

### Simulations

For power and error rate assessments, data were simulated in two general types of extended pedigrees: (1) a small three-generation pedigree with all individuals genotyped (Figure 1A), and (2) a larger three-generation pedigree where no grandparental genotypes are available (Figure 1B). Fifty pedigrees were simulated 1000 times.

For examining the contribution of different types of relatives to power, three family configurations were examined: (1) sib pairs, (2) sib pairs with parents, and (3) sib pairs with parents and grandparents. Sets of 200, 400 or 800 pedigrees were simulated 1000 times.

Briefly, trait values were constructed as the sum of polygenic (with variance $\sigma^2_g$) and environmental ($\sigma^2_e$) effects, assigned independently from a normal distribution with mean zero, and a major gene effect ($\sigma^2_a$), generated by an additive di-allelic trait locus, Q. A di-allelic marker locus, M, was simulated at a very small recombination fraction $\theta$. Since linkage disequilibrium is only expected at short distances in outbred populations[30–32], we consider $\theta = 0$ unless noted otherwise. The trait and marker loci allele frequencies were assumed to be $p_Q = q_Q = p_m = q_m = \frac{1}{2}$.

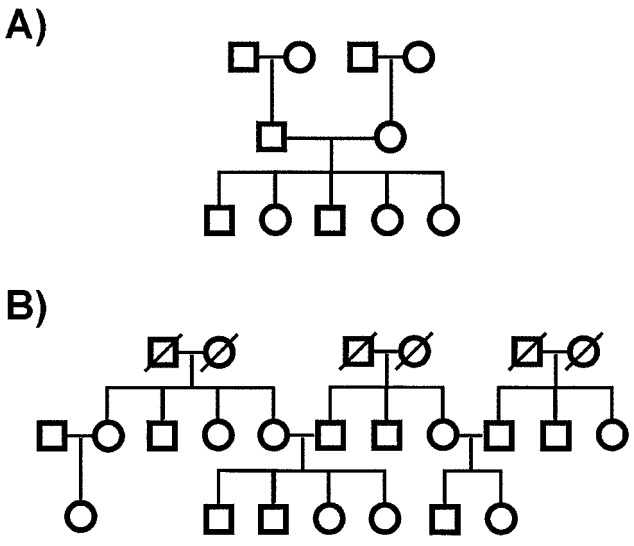


**Figure 1 Pedigrees used in simulations**. In these pedigrees, transmission can be scored using not only information from parents and siblings, but also from other types of relatives such as grandparents and uncles.

Linkage disequilibrium between the trait and marker loci was introduced in the founder chromosomes. Disequilibrium was modeled in the usual fashion as $D = p_{mQ} - p_m p_Q$ ($p_{mQ}$ is the frequency of the haplotype with allele '1' at both the marker and trait loci, so that $D_{max} = min(p_m p_Q) - p_m p_Q$, and the standardised equilibrium coefficient $D'$ is $D/D_{max}$.[33]

Where noted, population admixture was generated by mixing families drawn from two populations (A and B) with different phenotypic means ($\mu_A$ and $\mu_B$) and marker allele frequencies ($p_A = 0.7$ and $p_B = 0.3$) in equal sampling proportions. $\mu_A$ and $\mu_B$ were selected such that admixture accounted for 20% of the total phenotypic variance in the combined population, that is

$$\frac{(\mu_A - \mu_B)^2}{4\sigma^2} = 0.20.$$

For discrete trait analysis, individuals were arbitrarily assigned a status of affected (unaffected) when the simulated quantitative trait score was above (below) the mean. When analyzing simulated data sets we assumed full identity-by-descent (IBD) information at the marker locus. In practical settings, multipoint methods should be able to extract this information when many markers are screened in a dense map.

### Angiotensin converting enzyme data

The data consist of 553 individuals in 69 British extended families with no inbreeding. Pedigrees range in size from two to three generations, including from four to 18 individuals each. Genotypes are available for approximately 50% of all founders. Circulating ACE levels were measured for 405 individuals and standardised separately for males and females. Ten di-allelic polymorphisms in the ACE gene were genotyped.[21]

SimWalk2[34] was used to estimate multipoint IBD at each marker. After scoring allelic transmission using the method described here, evidence for linkage disequilibrium was evaluated with the $\chi^2_{qtl}$ statistic. Evidence for linkage and complete linkage disequilibrium was evaluated as previously described,[13,19] and lod scores were calculated as $\chi^2/(2 \ln 10)$.

## Results

### Simulations

In the absence of linkage disequilibrium, the Type I error rate for $\chi^2_{qtl}$, $T_{all}$ and $T_{aff}$ is compatible with both nominal and empirical (estimated from 1000 permutations) significance levels (Table 1). In contrast, the error rates for the $T_{TDT}$ and $F_{QTL}$ statistics, which assume independent observations, are very high when nominal significance levels are used. We

recommend exact significance levels estimated from permutations, as these should guarantee accurate results even in smaller datasets or with non-normal data.

It is important to note that the additional information available in the larger pedigree always provides more power in the presence of linkage disequilibrium (Table 2). In simulated data, $\chi^2_{qtl}$, which considers all the information available in a continuous distribution, provided more power than any other statistic. The $F_{qtl}$ statistic, which ignores information due to linkage and familiality, was slightly less powerful. The $T_{aff}$ and $T_{TDT}$ statistics, which consider only a small proportion of all available individuals (the affected descendants of heterozygous individuals), provided little power in small pedigrees.

As expected, power is very sensitive to the degree of linkage disequilibrium (Table 2). While all statistics provide reasonable power when the trait and marker locus are in perfect disequilibrium (D' = 1.00), almost no power was available for low levels of disequilibrium (D' = 0.25). In the absence of linkage disequilibrium (D' = 0), power estimated by simulation closely approximates the expected 0.01 error rate. These methods provide specific tests of linkage disequilibrium even in multigenerational pedigrees.

Information on additional relatives, such as parents or grandparents always increases power for the $\chi^2_{qtl}$ or $T_{all}$ statistics, but parental genotypes can decrease power for the $T_{aff}$ statistic (Table 3, along each row). The result is counterintuitive, but it is known that scoring of transmission in discordant pairs can be more efficient than the TDT for traits with high prevalence.[35]

For a fixed genotyping effort, simple sib-pair families always provide more power than other family configurations (Table 3). Although information from all individuals is used in scoring transmission, parental and grandparental phenotypes do not contribute to the linkage disequilibrium statistics, so this is not surprising. Note that, for the $\chi^2_{qtl}$ and $T_{all}$

**Table 1** Error rates are reported as the proportion of simulations exceeding the 0.05 significance level

| | *Error rates (5% significance level)* | | | | | |
|---|---|---|---|---|---|---|
| | *Nominal* | | | *Empirical* | | |
| | *Linkage* | *Familiality* | *Stratification* | *Linkage* | *Familiality* | *Stratification* |
| Small 3-generation pedigree | | | | | | |
| $\chi^2_{qtl}$ | 5.6 | 4.4 | 5.8 | 6.0 | 4.6 | 5.6 |
| $F_{qtl}$ | 14.5 | 12.1 | 8.5 | 4.5 | 4.6 | 5.6 |
| $T_{all}$ | 5.2 | 4.4 | 4.0 | 6.1 | 4.9 | 4.7 |
| $T_{aff}$ | 4.5 | 4.7 | 4.9 | 5.0 | 5.5 | 5.1 |
| $T_{TDT}$ | 20.1 | 20.3 | 19.5 | 4.8 | 5.4 | 5.0 |
| Large 3-generation pedigree | | | | | | |
| $\chi^2_{qtl}$ | 4.8 | 4.9 | 4.7 | 4.3 | 5.3 | 6.2 |
| $F_{qtl}$ | 11.0 | 5.2 | 6.1 | 4.4 | 4.7 | 5.5 |
| $T_{all}$ | 4.9 | 4.2 | 4.8 | 5.5 | 4.4 | 5.5 |
| $T_{aff}$ | 4.3 | 4.0 | 4.6 | 4.8 | 4.3 | 4.9 |
| $T_{TDT}$ | 4.9 | 3.2 | 4.0 | 5.1 | 4.8 | 5.4 |

Datasets with 50 pedigrees (Figure 1) were simulated, under the null hypothesis of no disequilibrium, in the presence of large linkage ($\sigma^2_a$ = 0.5, θ = 0) or familial ($\sigma^2_g$ = 0.5) effects, or population stratification (see text). Empirical significance levels were estimated from 1000 permutations of each dataset (see text).

**Table 2** Power is reported as the proportion of simulations exceeding the 0.01 empirical significance level

| | *Power (1% empirical significance level)* | | | | |
|---|---|---|---|---|---|
| *D'* | *0.00* | *0.25* | *0.50* | *0.75* | *1.00* |
| Small 3-generation pedigree | | | | | |
| $\chi^2_{qtl}$ | 0.6 | 7.8 | 33.5 | 76.5 | 98.4 |
| $F_{qtl}$ | 0.7 | 6.2 | 22.2 | 60.1 | 92.1 |
| $T_{all}$ | 1.2 | 4.4 | 14.3 | 42.7 | 76.2 |
| $T_{aff}$ | 0.7 | 1.7 | 4.1 | 7.5 | 15.8 |
| $T_{TDT}$ | 0.7 | 2.5 | 4.5 | 10.1 | 21.0 |
| Large 3-generation pedigree | | | | | |
| $\chi^2_{qtl}$ | 0.6 | 16.4 | 73.3 | 99.1 | 100.0 |
| $F_{qtl}$ | 0.9 | 12.0 | 64.7 | 96.7 | 100.0 |
| $T_{all}$ | 0.6 | 7.4 | 37.9 | 77.9 | 96.2 |
| $T_{aff}$ | 1.3 | 5.4 | 20.7 | 53.6 | 85.2 |
| $T_{TDT}$ | 0.9 | 5.2 | 23.4 | 58.0 | 88.5 |

Datasets with 50 pedigrees (Figure 1) were simulated, for varying levels of disequilibrium between trait and marker loci. The major locus had a small effect ($\sigma^2_a$ = 0.1, θ = 0) and background familial effects ($\sigma^2_g$ = 0.5) were present. Empirical significance levels were estimated from 1000 permutations of each dataset (see text).

**Table 3** Family Structure. Power is reported as the proportion of simulations exceeding the 0.0001 nominal significance level

| | | *Power at 0.0001 significance level (number of genotypes)* | | | | | |
|---|---|---|---|---|---|---|---|
| | *No of Families* | *Sib pair* | | *Sib pair with parents* | | *Sib pair, parents and grandparents* | |
| $\chi^2_{qtl}$ | 200 | 13.3 | (200) | 33.8 | (400) | 90.2 | (800) |
| | 400 | 52.9 | (400) | 86.2 | (800) | 100.0 | (1600) |
| | 800 | 96.1 | (800) | 100.0 | (1600) | 100.0 | (3200) |
| $T_{all}$ | 200 | 1.3 | (200) | 7.2 | (400) | 44.4 | (800) |
| | 400 | 13.1 | (400) | 45.8 | (800) | 94.5 | (1600) |
| | 800 | 60.5 | (800) | 94.5 | (1600) | 100.0 | (3200) |
| $T_{TDT}$ | 200 | 1.2 | (200) | 1.8 | (400) | 4.5 | (800) |
| | 400 | 12.8 | (400) | 10.7 | (800) | 26.1 | (1600) |
| | 800 | 60.7 | (800) | 49.9 | (1600) | 78.6 | (3200) |

Different family configurations and sample sizes were examined, and the total number of individuals genotyped in each case is given in (parenthesis). A high level of disequilibrium between trait and marker loci (D' = 0.75) was introduced. The major locus had a small effect ($\sigma^2_a$ = 0.1) and was tightly linked to the marker (θ = 0.0001). Background familial effects ($\sigma^2_g$ = 0.5) were present.

**Table 4** Transmission disequilibrium analysis of the ACE gene. The dataset has been previously published.[21] Using multipoint IBD estimates, evidence for linkage was evaluated using a variance components model. After scoring allelic transmission, evidence for association and additional effects were evaluated.[13,19] In each case, parameter estimates and lod scores are presented

| *Marker* | *Position* | *Linkage* | | *Association* | | *Additional effects* | |
|---|---|---|---|---|---|---|---|
| *Keavney* et al[21] | *Rieder* et al[39] | *lod* | $\sigma^2_a$ | *lod* | $\beta_w$ | *lod* | $\sigma^2_a$ |
| T-5491C | –2851 | 7.19 | 0.60 | 9.86 | 0.89 | 1.12 | 0.23 |
| A-5466C | –2826 | 7.19 | 0.60 | 9.04 | 0.84 | 0.81 | 0.20 |
| T-3892C | –1252 | 7.19 | 0.60 | 12.49 | 0.86 | 0.90 | 0.17 |
| A-240T | 2400 | 7.18 | 0.60 | 10.81 | 0.91 | 0.35 | 0.13 |
| T-93C | 2547 | 7.18 | 0.60 | 10.93 | 0.92 | 0.38 | 0.14 |
| T-1237C | 8128 | 7.18 | 0.60 | 11.52 | 0.82 | 0.38 | 0.10 |
| G2215A | 12257 | 7.19 | 0.60 | 14.91 | 0.93 | 0.04 | 0.03 |
| I/D | 14094 | 7.19 | 0.60 | 15.76 | 0.90 | 0.00 | 0.01 |
| G2350A | 14521 | 7.19 | 0.60 | 14.40 | 0.93 | 0.00 | 0.01 |
| 4656(CT)3/2 | 23945 | 7.23 | 0.60 | 14.22 | 0.89 | 0.25 | 0.08 |

statistics, the efficiency of the sib pair with parents and the sib pair, parents and grandparents designs are roughly equivalent.

### ACE data

The ACE locus is strongly linked to ACE levels (Table 4, Linkage columns, lod ≥ 7.18) and all the polymorphisms examined by Keavney *et al*[21] show strong evidence for association when the $\chi^2_{qtl}$ statistic is used (Table 4, Association columns, lod ≥ 9.86). Evidence for association is strongest at the I/D polymorphism (lod = 15.76, which was not exceeded in 1 million permutations of the data). The evidence for additional linked factors when association is included in the model (Table 4, Additional effects columns) can be used to determine whether a polymorphism is in complete disequilibrium with the trait alleles.[19,36] In Table 4, there is no evidence for additional linked effects when markers G2215A, I/D or G2350 are considered (lod ≤ 0.05). The data suggest these polymorphisms are in complete disequilibrium with the trait alleles (and could be the trait alleles themselves).

## Discussion

We have shown that allelic transmission can be scored in extended pedigrees, incorporating information not only from parents and siblings but also from other ancestors. These allelic transmission scores can be used to construct tests of linkage disequilibrium in general pedigrees and may be useful in refining the location of complex disease genes. As it is notoriously difficult to replicate linkage findings in complex disease,[37] the possibility of using the same family data sets to establish an original linkage by allele-sharing analysis and fine mapping by association analysis seems attractive.

By simulation we have shown that tests based on these transmission scores are unbiased in presence of familiality, population stratification or mere linkage, and are thus tests of linkage disequilibrium. A permutation framework for examining non-normal quantitative trait data or small datasets is also illustrated. Since transmissions to each family are considered (and permuted) as a whole, the test may be conservative in small datasets. It would be desirable to consider transmission from each founder separately, but in practice this is difficult.[25]

The method was applied to a published dataset of extended pedigrees, and our results agree with the comprehensive haplotype analysis of Keavney *et al*[21] and Farrall *et al*.[22] The original haplotype analysis shows the trait alleles are located in a haplotype defined by polymorphisms T1237G, G2215A, I/D, G2350A and 4656CT(3/2). The present transmission disequilibrium analysis suggests that the trait alleles are indistinguishable from polymorphisms G2215A, I/D and G2350A.

Proceeding from gene localisation to a broad chromosomal segment to localisation to a smaller segment and gene identification is a daunting task. This approach and others currently under development[38] extend the usefulness of linkage disequilibrium mapping to large pedigrees. Further research, including a comparison of these approaches in terms of speed, population assumptions and parameter interpretations, would be helpful to investigators.

### Acknowledgements

*GRA is supported by a Wellcome Trust Prize Studentship. WOCC is a Wellcome Trust Senior Clinical Research Fellow. We would like to thank Dr Martin Farrall for making the ACE data available. We also thank Dr Richard Mott for clarifying our notation.*

### References

1 Collins FS: Positional cloning: let's not call it reverse anymore. *Nat Genet* 1992; **1**: 3–6.
2 Darvasi A, Soller M: A simple method to calculate the resolving power and confidence interval of QTL map location. *Behav Genet* 1997; **27**: 125–132.
3 Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS: Replication of linkage studies in complex traits: an examination of variation in location estimates. *Am J Hum Genet* 1999; **65**: 876–884.
4 Whittemore AS, Halpern J: A class of tests for linkage using affected pedigree members. *Biometrics* 1994; **50**: 118–127.
5 Morton NE: Sequential tests for the detection of linkage. *Am J Hum Genet* 1955; **7**: 277–318.
6 Elston RC, Stewart J: A general model for the genetic analysis of pedigree data. *Hum Hered* 1971; **21**: 523–542.

7 Wijsman EM, Amos CI: Genetic analysis of simulated oligogenic traits in nuclear and extended pedigrees: summary of GAW10 contributions. *Genet Epidemiol* 1997; **14**: 719–735.
8 Badner JA, Gershon ES, Goldin LR: Optimal ascertainment strategies to detect linkage to common disease alleles. *Am J Hum Genet* 1998; **63**: 880–888.
9 Collins FS, Guyer MS, Chakravarti A: Variations on a theme: cataloguing human DNA sequence variation. *Science* 1997; **278**: 1580–1581.
10 Lander E: Array of hope. *Nat Genet* 1999; **21**: 3–4.
11 Chakravarti A: It's raining SNPs, hallelujah? *Nat Genet* 1999; **19**: 216–217.
12 Risch N, Merikangas K: The future of genetic studies of complex human diseases. *Science* 1996; **273**: 1516–1517.
13 Abecasis GR, Cardon LR, Cookson WOC: A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 2000; **66**: 279–292.
14 Allison DB: Transmission-disequilibrium tests for quantitative traits. *Am J Hum Genet* 1997; **60**: 676–690.
15 Weinberg CR, Wilcox AJ, Lie RT: A log-linear approach to case-parent-triad data: assessing the effects of genes that act either directly or through maternal effects and that may be subject to parental imprinting. *Am J Hum Genet* 1998; **62**: 969–978.
16 Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993; **52**: 506–516.
17 Rabinowitz D: A transmission disequilibrium test for quantitative trait loci. *Hum Hered* 1997; **47**: 342–350.
18 Curtis D: Use of siblings as controls in case-control association studies. *Ann Hum Genet* 1997; **61**: 319–333.
19 Fulker DW, Cherny SS, Sham PC, Hewitt JK: Combined linkage and association analysis for quantitative traits. *Am J Hum Genet* 1999; **64**: 259–267.
20 Allison DB, Heo M, Kaplan N, Martin ER: Sibling-based tests of linkage and association for quantitative traits. *Am J Hum Genet* 1999; **64**: 1754–1763.
21 Keavney B, McKenzie CA, Connell JM *et al*: Measured haplotype analysis of the angiotensin-I converting enzyme gene. *Hum Mol Genet* 1998; **7**: 1745–1751.
22 Farrall M, Keavney B, McKenzie C, Delepine M, Matsuda F, Lathrop GM: Fine-mapping of an ancestral recombination breakpoint in DCP1. *Nat Genet* 1999; **23**: 270–271.
23 Thomson G: HLA disease associations: models for insulin dependent diabetes mellitus and the study of complex human genetic disorders. *Annu Rev Genet* 1988; **22**: 31–50.
24 Curtis D, Sham PC: A note on the application of the transmission disequilibrium test when a parent is missing. *Am J Hum Genet* 1995; **56**: 811–812.
25 Martin ER, Kaplan NL, Weir BS: Tests for linkage and association in nuclear families. *Am J Hum Genet* 1997; **61**: 439–448.
26 George VT, Elston RC: Testing of association between polymorphic markers and quantitative traits in pedigrees. *Genet Epidemiol* 1987; **4**: 193–201.
27 Hopper JL, Mathews JD: Extensions of multivariate normal models for pedigree analysis. *Ann Hum Genet* 1982; **46**: 373–383.
28 Falconer DS: *Introduction to Quantitative Genetics*. Longman Scientific and Technical: London, 1989.
29 Allison DB, Neale MC, Zannolli R, Schork NJ, Amos CI, Blangero J: Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure. *Am J Hum Genet* 1999; **65**: 531–544.
30 Moffatt MF, Traherne JA, Abecasis GR, Cookson WOC: Single nucleotide polymorphisms and linkage disequilibrium with the TCR alpha/delta locus. *Hum Mol Genet* 2000; **9**: 1011–1019.
31 Huttley GA, Smith MW, Carrington M, O'Brien SJ: A scan for linkage disequilibrium across the human genome. *Genetics* 1999; **152**: 1711–1722.
32 Nickerson DA, Taylor SL, Weiss KM *et al*: DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat Genet* 1998; **19**: 233–240.
33 Lewontin RC, Kojima K: The evolutionary dynamics of complex polymorphisms. *Evolution* 1960; **14**: 450–472.
34 Sobel E, Lange K: Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am J Hum Genet* 1996; **58**: 1323–1337.
35 Horvath S, Laird NM: A discordant-sibship test for disequilibrium and linkage: no need for parental data. *Am J Hum Genet* 1998; **63**: 1886–1897.
36 Cardon LR, Abecasis GR: Some properties of a variance components model for fine-mapping quantitative trait loci. *Behav Genet* 2000; (in press).
37 Suarez BK, Hampe CL, Eerdewegh PV: Problems of replicating linkage claims in psychiatry. In: Gershon ES, Cloninger CR (eds). *Genetic Approaches to Mental Disorders*. American Psychiatric Press: Washington, 1994, 23–46.
38 Rabinowitz D: Adjusting for population stratification and admixture when testing non-parametric hypotheses in statistical genetics. *J Am Stat Assoc* 2000; (submitted).
39 Rieder MJ, Taylor SL, Clark AG, Nickerson DA: Sequence variation in the human angiotensin converting enzyme. *Nat Genet* 1999; **22**: 59–62.