Scientific
Research
Publishing

Haplotype Frequency Comparison for Case-Parents Data

Caixia Li[1], Peixing Li[1,2]*

[1]School of Mathematics, Sun Yat-sen University, Guangzhou, China
[2]Guangdong Province Key Laboratory of Computational Science, Sun Yat-sen University, Guangzhou, China
Email: *lnslpx@mail.sysu.edu.cn

Abstract

For case-parents data, the information from offspring can be used to reduce the uncertainty of parents' haplotype. In this article we develop likelihood ratio test to compare haplotype frequencies in transmitted and non-transmitted group. The maximum likelihood estimate of the haplotype frequencies for the family data is obtained via expectation-maximization (EM) algorithm. Our proposed method can handle the uncertainty of haplotypes and missing data. The simulations show that the method is more powerful to test association between haplotype and traits than TRANSMIT. We also demonstrated the method to detect the association between Megsin gene and immunoglobulin A nephropathy.

Keywords

Haplotype, Transmission Disequilibrium Test, Likelihood Ratio Test, Expectation-Maximization Algorithm

1. Introduction

In association studies, to avoid false positive results caused by population stratification, family-based tests of association are often used for fine mapping of a disease susceptibility locus. Transmission disequilibrium test (TDT) proposed by Spielman is an association test for case-parents triad data [1] [2]. The classic TDT is McNemar test on a 2-way transmission/non-transmission table.

Multiple linked markers can provide more polymorphism information than single marker (especially SNPs). However, haplotype phase is often uncertain for multi-locus genotype. There may be several haplotype pairs compatible with observed genotype. Many haplotype reconstruction algorithms are developed, e.g. expectation-maximization (EM) algorithm [3], pseudo MCMC [4], Bayesian

haplotype inference [5] for population data of unrelated individuals.

TDT has been extended for tightly linked marker loci. Zhao *et al.* (2000) [6] performed TDT via two steps: constructing the underlying haplotype first, and then constructing the 2-way transmission/non-transmission table in TDT by assigning a weight to each possible phase. Based on conditional likelihood, Clayton (1999) proposed a score test with program TRANSMIT [7]. The TDT-type methods were widely used in medical studies and GWAS studies [8] [9].

In this paper, based on full likelihood, we proposed a likelihood ratio test integrating haplotype construction for case-parent or case-parents data.

2. Method

The key idea is from classical case-control association study. Association between trait and marker yields some allele or haplotypes are found more often in case group than control group. In case-parents data, the allele or haplotypes of the case are transmitted from the parents. And so, we can test whether a particular allele or haplotype exists more often in transmitted (case) than in non-transmitted (control).

For tightly linked marker loci, we treat haplotypes as extended alleles and use the transmission information to reduce the phase uncertainty.

Let H denote the number of haplotypes for I tightly linked loci. The set of all possible $H(H+1)/2$ genotypes (haplotype pairs) is $G = \{1/1, 1/2, \cdots, 1/H, 2/2, 2/3, \cdots, 2/H, \cdots, (H-1)/H, H/H\}$.

For a case-parents trios, the genotypes for father, mother and affected child are denoted by g_f, g_m, g_c respectively. Let $F = (F_1, F_2, \cdots, F_H)$ and $F^* = (F_1^*, F_2^*, \cdots, F_H^*)$ denote haplotype frequency for transmitted and non-transmitted group, and $R_{ij} = P(Affected \mid i/j)/P(Affected \mid H/H)$ denote relative risk for genotype i/j reference to H/H. Under multiplicative haplotype risk $R_{ij} = R_i R_j$, where $R_i = \sqrt{R_{ii}}$ is relative risk for haplotype i reference to H. Under assumption of Hardy-Weinberg equilibrium, the probability of the father transmitted haplotype i and non-transmitted j, whereas the mother transmitted haplotype k and non-transmitted j to the child conditional on that the child is affected, is

$$P\left(g_f = i/j, g_m = k/l, g_c = i/k \mid A\right) = F_i^* F_j F_k^* F_l \tag{1}$$

where A means the child is affected, and $F_i^* = R_i F_i \Big/ \sum_{j=1}^{H} R_j F_j$ is regarded as frequency of haplotype i for disease population.

However, we only observed the genotype of each locus. The haplotype phase of the I tightly linked loci is often uncertainty, especial for missing parental genotype data. So there exist ambiguities to decide which haplotype is transmitted or not transmitted from the parent. Then

$$P\left(g_f, g_m, g_c \mid A\right) = \left[\sum_{(i,j,k,l) \in \tilde{G}} F_{i_r}^* F_{j_r} F_{k_r}^* F_{l_r}\right], \tag{2}$$

where $\tilde{G}$ is the set of haplotype groups (i, j, k, l) which haplotype pairs

(i, j), (k, l), and (i, k) are compatible with g_f, g_m and g_c, respectively. Here, missing parental genotype is allowable.

Suppose there are N case-parents trios, and then there are $2N$ parents in all. The genotypes for the r-th trios are g_{rf}, g_{rm}, g_{rc}. The log-likelihood

$$\ln L(F, F^*) = \sum_{r=1}^{N} \ln P(g_{rf}, g_{rm}, g_{rc} \mid A) = \sum_{r=1}^{N} \ln \left[\sum_{(i_r, j_r, k_r, l_r) \in \tilde{G}_r} F_{i_r}^* F_{j_r} F_{k_r}^* F_{l_r} \right]. \quad (3)$$

It is difficult to find the maximum likelihood estimate (MLE) $(\hat{F}, \hat{F}^*)$ directly. We employ expectation-maximization (EM) algorithm to estimate the haplotype frequencies $(\hat{F}, \hat{F}^*)$, by treating underlying haplotype pairs as "missing data". The complete-data log-likelihood after adding missing data $\{(i_1, j_1, k_1, l_1), \cdots, (i_N, j_N, k_N, l_N)\}$ is

$$\ln L_c(F, F^*) = \sum_{r=1}^{N} \ln \left(F_{i_r}^* F_{j_r} F_{k_r}^* F_{l_r} \right). \quad (4)$$

We can show that the expected complete-data log-likelihood $Q(F, F^*)$ in E (expectation) step (see **Appendix**)

$$\ln Q(F, F^*) = E\left[\ln L_c(F, F^*) \right] = \sum_{r=1}^{N} \sum_{i,j,k,l=1}^{M} w_r(i, j, k, l) \ln \left(F_i^* F_j F_k^* F_l \right), \quad (5)$$

where

$$w_r(i, j, k, l) = \begin{cases} \dfrac{F_i^* F_j F_k^* F_l}{\displaystyle\sum_{(i', j', k', l') \in \tilde{G}_r} F_{i'}^* F_{j'} F_{k'}^* F_{l'}}, & (i, j, k, l) \in \tilde{G}_r, \\[4mm] 0, & (i, j, k, l) \notin \tilde{G}_r. \end{cases}$$

An iterative procedure can be used to find the MLE via EM algorithm. Given the current estimates $\left(F^{(t)}, F^{*(t)} \right)$, the estimates in the next step:

$$\begin{cases} F_h^{(t+1)} = \dfrac{1}{2N} \displaystyle\sum_{r=1}^{N} \sum_{i,j,k,l=1}^{H} \left[w_r^{(t)}(i, h, k, l) + w_r^{(t)}(i, j, k, h) \right], \\[4mm] F_h^{*(t+1)} = \dfrac{1}{2N} \displaystyle\sum_{r=1}^{N} \sum_{i,j,k,l=1}^{H} \left[w_r^{(t)}(h, j, k, l) + w_r^{(t)}(i, j, h, l) \right], \end{cases} \quad (6)$$

$h = 1, 2, \cdots, H$. Under the null hypothesis $F = F^*$, we have

$$\begin{aligned} F_h^{(t+1)} &= F_h^{*(t+1)} \\ &= \dfrac{1}{4N} \sum_{r=1}^{N} \sum_{i,j,k,l=1}^{H} \left[w_r^{(t)}(h, j, k, l) + w_r^{(t)}(i, h, k, l) + w_r^{(t)}(i, j, h, l) + w_r^{(t)}(i, j, k, h) \right] \end{aligned} \quad (7)$$

Let $L_0(\hat{F})$ denote the Likelihood under the null hypothesis $F = F^*$ or $R = (1, 1, \cdots, 1)$. The likelihood ratio statistic

$$\Lambda = 2 \ln L(\hat{F}, \hat{F}^*) - 2 \ln L_0(\hat{F}) \quad (8)$$

follows an asymptotic χ^2 distribution with H-1 degrees of freedom (df) when the null hypothesis is true.

According to the relationship between $R = (R_1, R_2, \cdots, R_H)$ and (F, F^*)

$$\frac{F_H^*}{F_H} = \frac{R_H}{\sum_{j=1}^{H} R_j F_j} = \frac{1}{\sum_{j=1}^{H} R_j F_j}, \quad \frac{F_i^*}{F_i} = \frac{R_i}{\sum_{j=1}^{H} R_j F_j}, \quad i = 1, 2, \cdots, H-1,$$

We can get

$$R_i = \frac{F_i^*/F_i}{F_H^*/F_H}, \quad i = 1, 2, \cdots, H. \tag{9}$$

The maximum likelihood estimator (MLE) of the relative risk R_i for haplotype i relative to haplotype H is therefore given as

$$\hat{R}_i = \frac{\hat{F}_i^*/\hat{F}_i}{\hat{F}_H^*/\hat{F}_H}. \tag{10}$$

3. Application

We apply our method to the published data which was used for family-based association analysis for immunoglobulin A nephropathy (IgAN) [10]. Two tightly linked loci C2093T and C2180T, located in the 3' untranslated region (UTR) in Megsin gene, were studied. This dataset contains 232 families with an affected child were entered into analysis (Table 1). There are missing data since the genotyping information of some parents are not available. Out of 232 nuclear families, only 125 (53.9%) families are complete case-parents trios for both loci, and 107 (46.1%) families are case-parent families for the locus C2093T or C2180T, which including 25 C2180T single parent families (SPF), 26 C2093T SPF, and 56 C2093T & C2180T SPF.

There exist 4 haplotypes for Megsin C2093T-C2180T, CC, CT, TC and TT, coded as 1, 2, 3 and 4. Give initial value $F^{(0)} = F^{*(0)} = (0.25, 0.25, 0.25, 0.25)$ and precision $\varepsilon = 10^{-6}$. The haplotype frequency estimates after the iterative procedure stops

$$\hat{F} = (0.078118, 0.508773, 0.245039, 0.168070),$$
$$\hat{F}^* = (0.055631, 0.638335, 0.181438, 0.124596).$$

The log-likelihood $\ln L(\hat{F}, \hat{F}^*) = -904.475$. For the null model, the haplotype frequency estimation

$$\hat{F} = \hat{F}^* = (0.065499, 0.579534, 0.211046, 0.143921),$$

and log-likelihood $\ln L_0 = -911.554$. So likelihood ratio statistic $\Lambda = 2 \times (911.554 - 904.475) = 14.158$, df = 3, $P = 0.0027$. The results show the significant difference between F and F^*, the transmitted haplotype frequencies

Table 1. 232 families entered into analysis.

C2093T trios + C2180T trios	C2093T trios + C2180T SPF	C2093T SPF + C2180T trios	C2093T SPF + C2180T SPF	Total
125	25	26	56	232

SPF: single parent family.

and non-transmitted haplotype frequencies. $\hat{F}_2^*$ is much higher than $\hat{F}_2$. That means haplotype 2 (2093C-2180T) is over-transmitted from parents to cases. In addition, reference to haplotype 4 (2093T-2180T), the estimated relative haplotype risk $\hat{R} = (0.961, 1.692, 0.999, 1)$.

4. Simulations

In our simulations, two tightly linked single nucleotide polymorphism (SNP) marker genotype data for 100 case-parents trios are generated. In a similar way in Morris *et al.* (1997) [11], samples are generated for simulation using the transmission model provided by Bickeboller *et al.* (1995) [12].

$$
\begin{aligned}
& P\left(g_c = si/uk, g_f = si/tj, g_m = uk/vl \mid A\right) \\
&= \frac{P\left(g_c = si/uk, g_f = si/tj, g_m = uk/vl, A\right)}{P(A)} \\
&= \frac{P\left(A \mid g_c = si/uk, g_f = si/tj, g_m = uk/vl\right) P\left(g_c = si/uk, g_f = si/tj, g_m = uk/vl\right)}{K_p} \\
&= \frac{P\left(A \mid g_c = s/u\right)}{K_p} h_{si} h_{tj} h_{uk} h_{vl} \\
&= \left(\frac{f_{su}}{K_p} h_{si} h_{uk}\right) \cdot \left(h_{tj} h_{vl}\right),
\end{aligned}
$$

where $s, t, u, v \in \{D, d\}$, $i, j, k, l \in \{1, 2, \cdots, H\}$, $f_{su} = P(A \mid s/u)$ is penetrance for genotype s/u, and $K_p = P(A)$ is prevalence.

The frequencies of mutant disease allele D and normal allele d in a disease locus are denoted by $q_1 = q$ and $q_2 = 1 - q$. Under Hardy-Weinberg Equilibrium, the population prevalence is therefore $K_p = q^2 f_{DD} + 2q(1-q) f_{Dd} + (1-q)^2 f_{dd}$. We assume that there is no recombination between disease and marker locus. Linkage diseqilibrium parameter to detect association between disease locus and marker locus is defined as in Sham (1995) [13]

$$
e_{si} = \frac{h_{si}}{F_i q_s}, \quad s = D, d; \ i = 1, 2, \cdots, H,
$$

where h_{si} is frequency of disease-marker haplotype *si*. The $\{e_{si}\}$ satisfies $\sum_{i=1}^{H} F_i e_{si} = 1$ and $\sum_{s=1}^{2} q_s e_{si} = 1$. All $e_{si} = 1 (s = D, d; i = 1, 2, 3, 4)$ implies linkage equilibrium. $e_{Di} > 1$ means marker haplotype i is positively associated with the disease, and $e_{Di} < 1$ means marker haplotype i is negatively associated. In our simulations, the LD pattern is given as

$$
1 \le e_{D1} = e \le 1/F_1, \quad 0 \le e_{Dk} = \frac{1 - F_1 e}{1 - F_1} \le 1, \ k = 2, 3, \cdots, H
$$

i.e. marker haplotype 1 is positively associated with the disease, and the other haplotypes are equally negatively associated, if there is association. In addition, the marker haplotypes are assumed to be equally frequent, and $1 \le e_{D1} = e \le 4$,

$0 \leq e_{D2} = e_{D3} = e_{D4} = \dfrac{4-e}{3} \leq 1$. Three classical genetic models, recessive, dominant and common models, specified heredity modes are considered. These models are shown in **Table 2**.

For each genetic model, 5000 replicated samples were generated to evaluate the distribution of test statistics in the case of no association, *i.e. e* = 1. The quantile-quantile (QQ) plots for test statistic Λ from 5000 samples under the null hypothesis for the three genetic models are showed in **Figures 1(a)-(c)**. The

Table 2. Genetic models for simulation study.

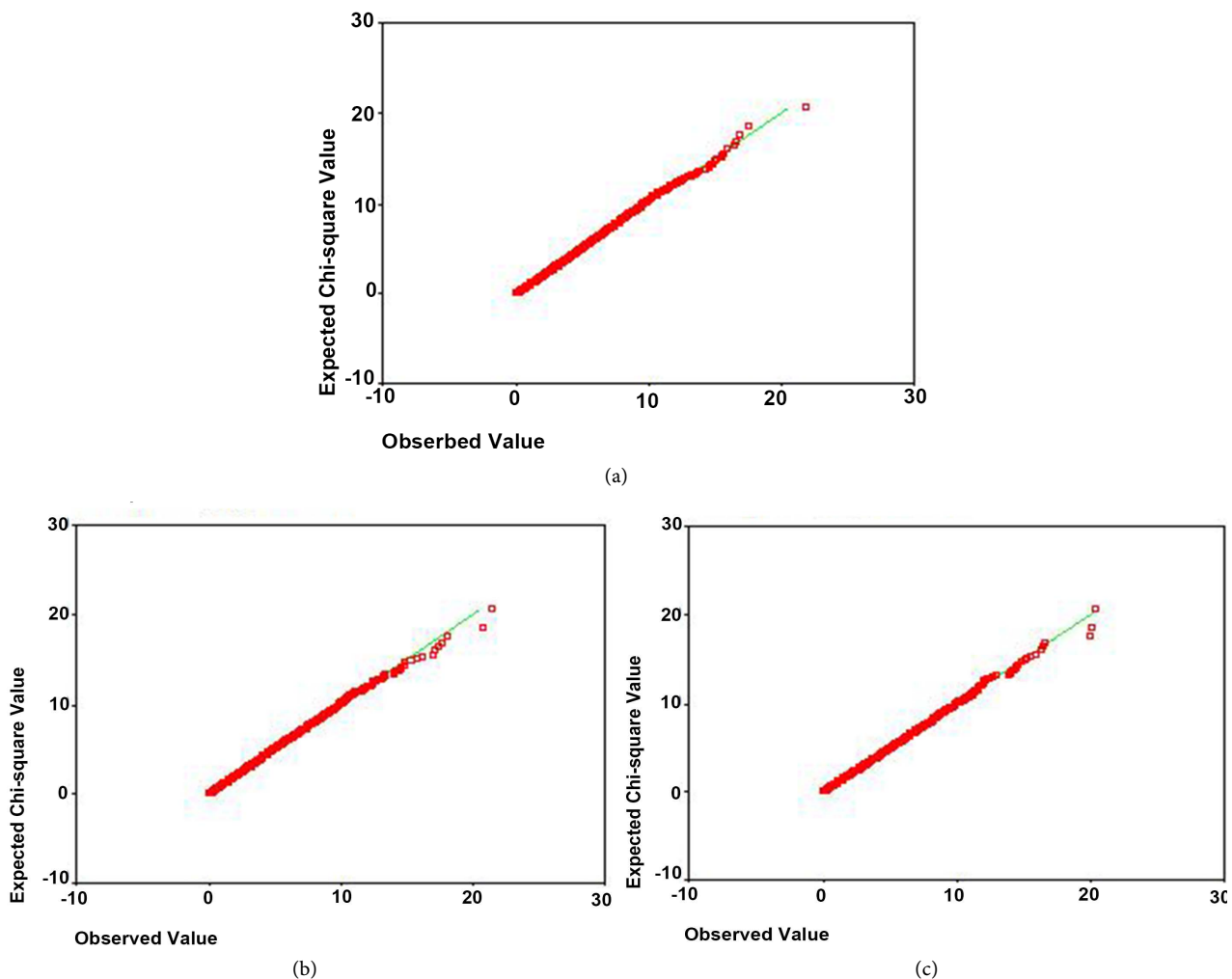| Heredity mode	q	Penetrances		
		f_{DD}	f_{Dd}	f_{dd}
Recessive	0.0001	1	0	0
Dominant	0.0001	1	1	0
Common	0.2	0.02	0.005	0.001

(a)

(b)

(c)

Figure 1. QQ plots for test statistic Λ from 5000 samples. Red: expected quantiles of distribution $\chi^2(3)$ versus observed quantiles; Green: the line *y* = *x*. (a) Recessive Model; (b) Dominant Model; (c) Common Model.

plots show that the scatters of the observed quantiles and the expected quantiles of distribution $\chi^2(3)$ are very close to the line $y = x$.

In addition, 1000 replicated samples were generated for statistical power analysis. The statistical significance level $\alpha = 0.05$ or 0.01. The empirical power for our proposed method, comparison with TRANSMIT, are summarized in Table 3, under different association level e for three disease model.

Table 3. Comparisons of empirical power.

	Recessive		Dominant		Common	
e	Proposed	TRANSMIT	Proposed	TRANSMIT	Proposed	TRANSMIT
$\alpha = 0.05$:						
1.00	0.0440	0.0434	0.0474	0.0462	0.0514	0.0486
1.25	0.184	0.173	0.073	0.071	0.058	0.052
1.50	0.557	0.539	0.187	0.179	0.134	0.132
1.75	0.914	0.912	0.348	0.347	0.250	0.240
2.00	0.994	0.990	0.570	0.557	0.430	0.410
2.25	1	1	0.777	0.765	0.607	0.596
2.50	1	1	0.893	0.885	0.762	0.746
2.75	1	1	0.978	0.978	0.881	0.873
3.00	1	1	0.998	0.997	0.952	0.949
3.25	1	1	1	1	0.989	0.987
3.50	1	1	1	1	0.998	0.996
3.75	1	1	1	1	1	1
4.00	1	1	1	1	1	1
$\alpha = 0.01$:						
1.00	0.0086	0.0082	0.0092	0.0082	0.0100	0.0094
1.25	0.055	0.049	0.019	0.015	0.015	0.015
1.50	0.315	0.300	0.055	0.050	0.043	0.037
1.75	0.790	0.775	0.160	0.147	0.086	0.082
2.00	0.966	0.963	0.321	0.304	0.221	0.204
2.25	1	0.999	0.573	0.548	0.394	0.363
2.50	1	1	0.749	0.741	0.545	0.525
2.75	1	1	0.916	0.906	0.720	0.705
3.00	1	1	0.977	0.976	0.879	0.868
3.25	1	1	0.994	0.991	0.946	0.936
3.50	1	1	1	1	0.985	0.979
3.75	1	1	1	1	0.993	0.992
4.00	1	1	1	1	0.997	0.996

5. Discussion

Haplotype frequencies are usually estimated when haplotypes are reconstructed or linkage disequilibrium is tested. For tightly linked loci, the likelihood as a function of haplotype frequencies in transmitted and non-transmitted group was given for case-parents data. We estimate the MLEs of the haplotype frequencies via an EM algorithm. The results showed that haplotype frequencies could be estimated using a simple iterative procedure. The likelihood ratio test to compare haplotype frequencies in transmitted and non-transmitted group was used to detect association. When the information of parents is not available in the nuclear family, classical TDT is no longer suitable. However, as you can see in the application, missing parental genotypes are allowed in our method. In addition, when there are siblings available, the information can be used to reduce the uncertainty of phase, and the likelihood can be given similarly.

Under different simulated conditions where heredity mode, linkage disequilibrium coefficient are specified, 5000 and 1000 replicated samples were generated to evaluate the distribution of test statistics and statistical power respectively. Our method is more powerful than TRANSMIT.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Spielman, R.S., McGinnis, R.E. and Ewens, W.J. (1993) Transmission Test for Linkage Disequilibrium: The Insulin Gene Region and Insulin-Dependent Diabetes Mellitus. *American Journal of Human Genetics*, **52**, 506-516.

[2] Spielman, R.S. and Ewens, W.J. (1996) The TDT and Other Family-Based Tests for Linkage Disequilibrium and Association. *American Journal of Human Genetics*, **59**, 983-989.

[3] Clayton, D.G. (1999) Transmission/Disequilibrium Tests for Extended Marker Haplotypes. *American Journal of Human Genetics*, **65**, 1161-1169. https://doi.org/10.1086/302566

[4] Stephens, M., Smith, N.J. and Donnelly, P. (2001) A New Statistical Method for Haplotype Reconstruction from Population Data. *American Journal of Human Genetics*, **68**, 978-989. https://doi.org/10.1086/319501

[5] Niu, T., Qin, Z.S., Xu, X. and Liu, J.S. (2002) Bayesian Haplotype Inference for Multiple Linked Single-Nucleotide Polymorphisms. *American Journal of Human Genetics*, **70**, 157-169. https://doi.org/10.1086/338446

[6] Zhao, H., Zhang, S., Merikangas, K.R., Trixler, M., Wildenauer, D.B., Sun, F., Kidd, K.K. (2000) Transmission/Disequilibrium Test Using Multiple Tightly Linked Markers. *American Journal of Human Genetics*, **67**, 936-946. https://doi.org/10.1086/303073

[7] Clayton, D.G. (1999) A Generalization of the Transmission/Disequilibrium Test for Uncertain-Haplotype Transmission. *American Journal of Human Genetics*, **65**, 1170-1177. https://doi.org/10.1086/302577

[8]   Abdulkadir, M., Londono, D., Gordon, D., *et al.* (2017) Investigation of previously Implicated Genetic Variants in Chronic Tic Disorders: A Transmission Disequilibrium Test Approach. *European Archives of Psychiatry and Clinical Neuroscience*, **268**, 1-16.

[9]   Wang, M., Ji, Z., Wang, S., *et al.* (2017) Mechanisms to Protect the Privacy of Families When Using the Transmission Disequilibrium Test in Genome-Wide Association Studies. *Bioinformatics*, **33**,

[10]   Li, Y.J., Du, Y., Li, C.X., *et al.* (2004) Family-Based Association Study Showing That Immunoglobulin a Nephropathy Is Associated with the Polymorphisms 2093C and 2180T in the 3′ Untranslated Region of Megsin Gene. *Journal of the American Society of Nephrology*, **15**, 1739-1743.
https://doi.org/10.1097/01.ASN.0000130858.00688.46

[11]   Morris, A.P., Whittaker, J.C. and Curnow, R.N. (1997) A Likelihood Ratio Test for Detecting Patterns of Disease-Marker Association. *Annals of Human Genetics*, **61**, 335-350. https://doi.org/10.1017/S0003480097006349

[12]   Bickeböller, H. and Clerget-Darpoux, F. (1995) Statistical Properties of the Allelic and Genotypic Transmission/Disequilibrium Test for Multiallelic Markers. *Genetic Epidemiology*, **12**, 865-870. https://doi.org/10.1002/gepi.1370120656

[13]   Sham, P.C. and Curtis, D. (1995) An Extended Transmission/Disequilibrium Test (TDT) for Multi-Allele Marker Loci. *Annals of Human Genetics*, **59**, 323-336.
https://doi.org/10.1111/j.1469-1809.1995.tb00751.x

Appendix (E Step in EM Algorithm)

Give the current estimate $\left(F^{(t)}, F^{*(t)}\right)$, the distribution of missing data

$$P_M^{(t)}\left\{(i_1, j_1, k_1, l_1), \cdots, (i_N, j_N, k_N, l_N)\right\}$$

$$= \prod_{r=1}^{N} P\left(g_{rf} = i_r / j_r, g_{rm} = k_r / l_r, g_{rc} = i_r / k_r \mid A, g_{rf}, g_{rm}, g_{rc}, F^{(t)}, F^{*(t)}\right)$$

$$= \prod_{r=1}^{N} w_r^{(t)}(i_r, j_r, k_r, l_r), \qquad i_1, j_1, k_1, l_1, \cdots, i_N, j_N, k_N, l_N = 1, 2, \cdots, H,$$

where

$$w_r^{(t)}(i, j, k, l) = \begin{cases} \dfrac{F_i^{*(t)} F_j^{(t)} F_k^{*(t)} F_l^{(t)}}{\displaystyle\sum_{(i', j', k', l') \in \tilde{G}_r} F_{i'}^{*(t)} F_{j'}^{(t)} F_{k'}^{*(t)} F_{l'}^{(t)}}, & (i, j, k, l) \in \tilde{G}_r, \\[6pt] 0, & (i, j, k, l) \notin \tilde{G}_r. \end{cases}$$

And then the expected complete-data log-likelihood

$$Q\left(F, F^* \mid F^{(t)}, F^{*(t)}\right) = E\left[\ln L_c\left(F, F^*\right) \mid F^{(t)}, F^{*(t)}\right]$$

$$= \sum_{i_1, j_1, k_1, l_1 = 1}^{H} \cdots \sum_{i_N, j_N, k_N, l_N = 1}^{H} \left[\sum_{r=1}^{N} \ln\left(F_{i_r}^* F_{j_r} F_{k_r}^* F_{l_r}\right)\right] P_M^{(t)}\left\{(i_1, j_1, k_1, l_1), \cdots, (i_N, j_N, k_N, l_N)\right\}$$

$$= \sum_{i_1, j_1, k_1, l_1 = 1}^{H} \cdots \sum_{i_N, j_N, k_N, l_N = 1}^{H} \left\{\left[\sum_{r=1}^{N} \ln\left(F_{i_r}^* F_{j_r} F_{k_r}^* F_{l_r}\right)\right] \prod_{n=1}^{N} w_n^{(t)}(i_n, j_n, k_n, l_n)\right\}$$

$$= \sum_{r=1}^{N} \left\{\left[\sum_{i,j,k,l=1}^{M} w_r(i, j, k, l) \ln\left(F_i^* F_j F_k^* F_l\right)\right] \prod_{n=1, n \neq r}^{N} \sum_{i,j,k,l=1}^{M} w_n^{(t)}(i, j, k, l)\right\}$$

$$= \sum_{r=1}^{N} \sum_{i,j,k,l=1}^{M} w_r(i, j, k, l) \ln\left(F_i^* F_j F_k^* F_l\right)$$