

Power of the 2-locus TDT for testing the interaction of two susceptibility genes.

Salma Kotti, Mathieu Bourgey, Françoise Clerget-Darpoux

► **To cite this version:**

Salma Kotti, Mathieu Bourgey, Françoise Clerget-Darpoux. Power of the 2-locus TDT for testing the interaction of two susceptibility genes.. BMC Proceedings, BioMed Central, 2007, 1 Suppl 1 (Suppl 1), pp.S65. inserm-00144404

HAL Id: inserm-00144404

<https://www.hal.inserm.fr/inserm-00144404>

Submitted on 13 Feb 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Power of the 2-locus TDT for testing the interaction of two susceptibility genes

Salma Kotti ^{1,2*}, Mathieu Bourgey ^{1,2}, Françoise Clerget-Darpoux ^{1,2}

¹ Univ Paris-Sud, UMR-S535, F-94817 Villejuif, France

² Inserm U535, F-94817 Villejuif, France

* Corresponding author

Salma Kotti
Inserm U535
Hôpital Paul brousse
BP1000
F-94817 Villejuif, France

Fax : + 33 1 45 59 53 31

kotti@vjf.inserm.fr

Email addresses :

Salma Kotti: kotti@vjf.inserm.fr

Mathieu Bourgey : bourgey@vjf.inserm.fr

Françoise Clerget-Darpoux: clerget@vjf.inserm.fr

Abstract

We recently proposed a new strategy: the 2-locus TDT for detecting two susceptibility genes through their interaction in trio families [1]. We apply our method to two candidate genes A and C on the GAW15 simulated rheumatoid arthritis data and study the power to conclude to an interactive effect of these genes.

This study was performed with full knowledge of the answers.

Introduction

Most multifactorial diseases result from complex interactions among multiple genes with modest effects and various environmental factors. Current strategies usually aim to detect the individual effect of each factor.

We have recently proposed a new strategy: the 2-locus TDT which tests for interaction between two candidate genes [1]. As the one locus TDT [2], it applies to trio families including patient and his two parents. Our method is used in the context of studying good candidate pathways. Indeed, we assume prior information that the disease under study is probably caused by particular genes recognized as key elements of the metabolic pathway. We apply the 2-locus TDT to problem 3 simulated GAW15 rheumatoid arthritis (RA) data and compute its power for detecting the interaction of two susceptibility genes.

Data

The GAW15 problem 3 consisted of 100 replicates of 1500 nuclear families with two affected sibs. From each replicate, we extract 1500 trio families: one affected (the index case) and his two parents.

We focus the analyses on two candidate genes in the RA susceptibility: locus A on chromosome 16 and locus C on chromosome 6. Locus A has an indirect effect on RA hazard. Its implication on the susceptibility to the disease depends on the locus HLA-DRB1 located on chromosome 6. Locus C is in strong LD with HLA-DRB1. Both genes are involved directly or indirectly in the RA outcome through the HLA-DRB1 gene and thus do not act independently. We apply the 2-locus TDT to the SNPs indicated as functional in the two candidate genes in each replicate of 1500 trios.

Notation

We denote A_1, A_2 and C_1, C_2 the SNP alleles of genes A and B respectively. For each patient, there are nine possible genotypes that we denote $A_i A_j C_k C_l$ where A_i, A_j and C_k, C_l are the alleles inherited from the two parents at the loci A and C, respectively. Subscripts i, j, k and l can be either 1 or 2. A_2 and C_2 are considered as the high risk alleles and $A_2 A_2 C_2 C_2$ as the reference genotype.

We denote the nine penetrances $P(\text{affected} | A_i A_j C_k C_l)$ by f_{ijkl} and by ε_{ijkl} the 8 penetrances relatively to the reference genotype $A_2 A_2 C_2 C_2$. Thus $\varepsilon_{2222} = 1$. Let M_G be the general model where ε_{ijkl} may take any value between 0 and 1. When the two genes A and B act independently, the lines and the columns of the relative penetrance matrix are proportional (multiplicative matrix) and may be written with 4 parameters: μ, ν, φ and π (see table 1). The corresponding restricted model is denoted M_R .

Method

The 2-locus TDT is achieved in two steps. In the first step, the parameters (μ , ν , φ and π) of the restricted model and ε_{ijkl} of the general model are estimated by their Maximum Likelihood Estimator (MLE). In the second one, the restricted model is compared to the general one by a maximum likelihood ratio test.

The estimation of these parameters requires specifying the allele frequencies at each locus. This was achieved by the Affected Family Based Controls (AFBAC) method using the parental alleles untransmitted to the patients from which the family was ascertained [3]. Here, the families were not ascertained through an affected patient but through a patient with an affected sib. Our parameter estimation also needs to take into account this specific mode of ascertainment and consequently was achieved by using the Marker Association Segregation Chi-square (MASC) program [4].

We denote by $\max_G L_G$ and $\max_R L_R$ the likelihood values for the estimated parameters of the general and restricted models, respectively. The likelihood ratio test is calculated as follows:

$$\text{LRT} = -2 \ln \frac{\max_R L_R}{\max_G L_G}$$

This ratio follows, under the null hypothesis, a chi-square distribution with four degrees of freedom (df) corresponding to the difference in the number of parameters between the two models.

For each replicate of 1500 trios, the LRT was computed and the significance of the test was assessed in order to evaluate the power of our test in the present situation.

Results

Because patients with genotypes $A_1A_1C_1C_1$, $A_1A_2C_1C_1$ and $A_2A_2C_1C_1$ are very rare in all replicates, we fixed the relative penetrances of these genotypes at 10^{-3} . Consequently, under the general model, we estimate only 5 parameters and, under the restricted model, only 3 parameters ν , φ and π . Thus, in the present situation, we have 2 degrees of freedom for the LRT. Figure 1 and 2 show the distributions of the LRT and p-values in 98 replicates (the program failed in 2 replicates), respectively. There is a good percentage of replicates (44/98) for which the test was significant at 1% level. The power of 2-locus TDT test (at a 5% level) is 66%.

We illustrate the variation of the LRT and p-values on two contrasted replicates: the 9 and 27. In the replicate 27, there is a strong evidence for interaction: $\text{LRT} = 14.35$ and $p = 7 \times 10^{-4}$. In contrast, we cannot conclude for interaction in the replicate 9 since the $\text{LRT} = 2.49$ and $p = 0.29$.

Tables 2 and 3 show the parameter estimates under the general model for replicates 9 and 27, respectively. The parameter estimates under the restricted model are $\nu = 0.135$, $\varphi = 0.688$ and $\pi = 0.886$ for replicate 9 and $\nu = 0.168$, $\varphi = 0.682$ and $\pi = 0.909$ for replicate 27. It is clear that the departure to the multiplicativity in the relative penetrance matrix of the general model is stronger for the replicate 27 than for the replicate 9. In particular, for replicate 27, the relative penetrance for the genotype $A_1A_2C_2C_2$ is half of the one expected under the restricted model.

Discussion

In this paper, we studied the performance of the 2-locus TDT for detecting the interaction of the genes A and C in the susceptibility to RA simulated data and show that interaction may be detected with a power of 66% at a 5% level.

The 2-locus TDT searches for a departure to the multiplicativity in the two-locus penetrance matrix. This departure cannot be observed without an effect of the two genes. Therefore, the 2-locus TDT may also be used for detecting the effect of two candidate genes through their interaction.

More generally, the 2-locus TDT may be particularly interesting in the situation of weak effect of one or both genes and strong interaction. In such a situation, traditional method will fail because they will stop in the absence of noticeable main effect. In fact, they ignore the possibility that the interactive effect of multilocus functional genetic variants may play a larger role than the single-locus effect in the determinism of one trait [1].

Our test applies to candidate genes already recognized as key elements of a metabolic pathway thought to be involved in the disease pathogenesis. There are difficulties in good candidate choice and the use of tagSNPs as markers for most genetic variations in the human genome has become a cornerstone of genetic research. There are, however, reasons to think that the expectations they have raised are somewhat inflated and may lead to unduly neglect other strategies. We believe in particular that identification of biologically relevant pathways followed by candidate gene strategies may be quite effective at identifying new variants and their interaction in disease susceptibility [5].

Acknowledgements

This work was supported by a grant from the medical research foundation FRM (Fondation pour la Recherche Médicale)

References

1. Kotti S, Bickeboller H, Clerget-Darpoux F: **Strategy for detecting susceptibility genes with weak or no marginal effect.** *Hum Hered* 2007, **63**(2):85-92.
2. 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**(3):506-516.
3. Thomson G: **Mapping disease genes: family-based association studies.** *Am J Hum Genet* 1995, **57**(2):487-498.
4. Clerget-Darpoux F, Babron MC, Prum B, Lathrop GM, Deschamps I, Hors J: **A new method to test genetic models in HLA associated diseases: the MASC method.** *Ann Hum Genet* 1988, **52**(Pt 3):247-258.
5. Bourgain C, Genin E, Cox N, Clerget-Darpoux F: **Are genome-wide association studies all that we need to dissect the genetic component of complex human diseases?** *Eur J Hum Genet* 2007, **15**(3):260-263.

Figures

Figure 1 - Distribution of the LRT values

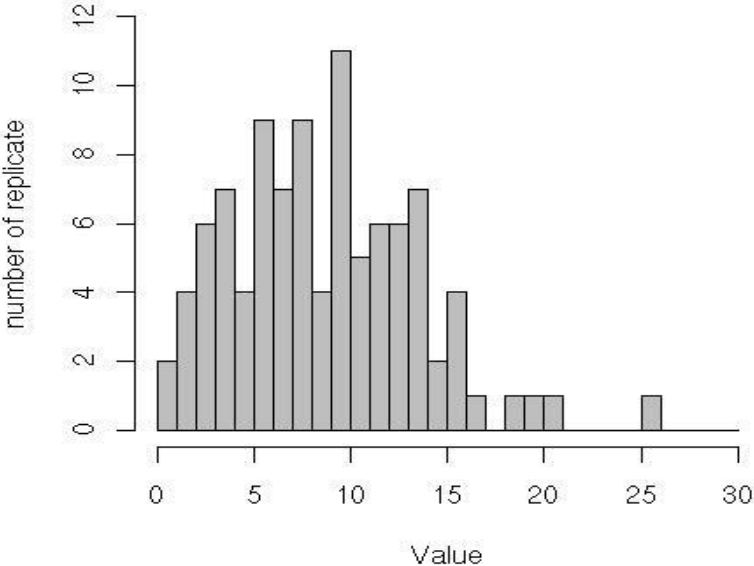
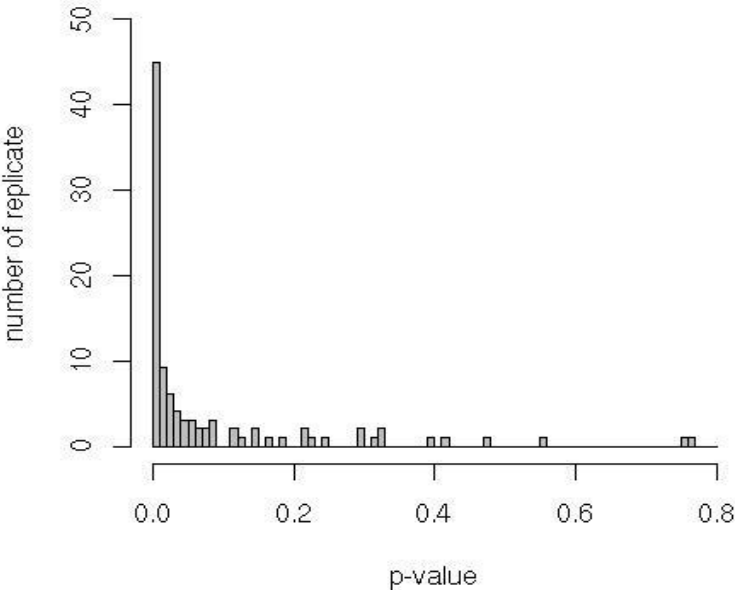


Figure 2 - Distribution of the LRT p-values



Tables

Table 1 - Penetrance estimation under the restricted model

		Gene A		
		A ₁ A ₁	A ₁ A ₂	A ₂ A ₂
Gene C	C ₁ C ₁	$\varphi \mu$	$\pi \mu$	μ
	C ₁ C ₂	$\varphi \nu$	$\pi \nu$	ν
	C ₂ C ₂	φ	π	1

Table 2 - Replicate 9: Matrix of relative penetrances for the general model

Parameter estimates (in blue)

		Gene A		
		A ₁ A ₁	A ₁ A ₂	A ₂ A ₂
Gene C	C ₁ C ₁	0.001	0.001	0.001
	C ₁ C ₂	0.105	0.109	0.113
	C ₂ C ₂	0.668	0.879	1

Table 3 - Replicate 27: Matrix of relative penetrances for the general model

Parameter estimates (in blue)

		Gene A		
		A ₁ A ₁	A ₁ A ₂	A ₂ A ₂
Gene C	C ₁ C ₁	0.001	0.001	0.001
	C ₁ C ₂	0.137	0.133	0.083
	C ₂ C ₂	0.623	0.868	1