Open Journal of Statistics, 2015, 5, 35-45
Published Online February 2015 in SciRes. http://www.scirp.org/journal/ojs
http://dx.doi.org/10.4236/ojs.2015.51005

Implicit Hypotheses Are Hidden Power Droppers in Family-Based Association Studies of Secondary Outcomes

Jean Gaschignard[1,2*], Quentin B. Vincent[1,2], Jean-Philippe Jaïs[1,2,3], Aurélie Cobat[1,2], Alexandre Alcaïs[1,2,4]

[1]Laboratoire de Génétique des Maladies Infectieuses, Institut National de la Santé et de la Recherche Médicale, Paris, France
[2]Université Paris Descartes, Sorbonne Paris Cité, Institut Imagine, Paris, France
[3]Biostatistique et Informatique Médicale, Hôpital Necker, Paris, Farnce
[4]URC, CIC, Necker and Cochin Hospitals, Paris, France
Email: *jean.gaschignard@inserm.fr, alexandre.alcais@inserm.fr

Received 8 January 2015; accepted 26 January 2015; published 30 January 2015

Abstract

Family-based tests of association between a genetic marker and a disease constitute a common design to dissect the genetic architecture of complex traits. The FBAT software is one of the most popular tools to perform such studies. However, researchers are also often interested in the genetic contribution to a more specific manifestation of the phenotype (e.g. severe vs. non-severe form) known as a secondary outcome. Here, what we demonstrate is the limited power of the classical formulation of the FBAT statistic to detect the effect of genetic variants that influence a secondary outcome, in particular when these variants also impact on the onset of the disease, the primary outcome. We prove that this loss of power is driven by an implicit hypothesis, and we propose a derivation of the original FBAT statistic, free from this implicit hypothesis. Finally, we demonstrate analytically that our new statistic is robust and more powerful than FBAT for the detection of association between a genetic variant and a secondary outcome.

Keywords

Family-Based Association Test, FBAT, Genetic Association Studies, Null Hypothesis, Secondary Outcome, Homogeneity Test

*Corresponding author.

How to cite this paper: Gaschignard, J., Vincent, Q.B., Jaïs, J.-P., Cobat, A. and Alcaïs, A. (2015) Implicit Hypotheses Are Hidden Power Droppers in Family-Based Association Studies of Secondary Outcomes. *Open Journal of Statistics*, **5**, 35-45.
http://dx.doi.org/10.4236/ojs.2015.51005

1. Introduction

The aim of genetic epidemiological studies is to identify the genetic factors influencing the development of common diseases. Genetic epidemiology combines classical epidemiological data (assessment of risk factors known to affect the expression of the phenotype studied) and genetic information (familial relationships, typing of genetic marker) and proposes a large range of tools to address the initial question, the use of one depending on the nature of your sample and the size of your wallet. Over the past ten years, however, our understanding of the pattern of genetic variation at the genome scale, coupled to an unprecedented decrease in the cost of measuring this variation, has put (genome-wide) association studies at the front. Although the vast majority of genetic association study designs are derived from usual case-control retrospective epidemiological studies (*i.e.* that compare the distribution of allelic/genotypic frequencies between a group of cases and a group of controls), one is quite specific to the field of genetic epidemiology and relies on the collection and analysis of families. Such family-based tests of association between a genetic item (allele, genotype...) and the disease under study offer interesting features as compared to case-control designs (Laird and Lange [1]; Chen and Abecasis [2]). They are robust against population stratification, allow the inference of both haplotype phase and missing genotypes (Chen and Abecasis [2]; Burdick *et al.* [3]), and can identify peculiar allelic segregation, for example, due to imprinting effect (Vincent *et al.* [4]).

The Transmission Desequilibrium Test (TDT) has emerged as the first popular family-based test of association (Spielman *et al.* [5]). It tests whether the transmission of a given allele from a heterozygote parent to an affected child is different from what is expected in the absence of any association between the genetic marker and the disease under study. The null hypothesis is written as $p = 0.5$ where p is the proportion of a given allele that has been transmitted to affected children by heterozygote parents. Whereas the TDT could only analyze binary traits in samples of pure trios (*i.e.* two parents and a single affected child), Laird *et al.* [6] proposed a more comprehensive approach designed to handle binary, quantitative or censored traits, multiple genetic models (e.g. additive, dominant or recessive) and more complex family structures (e.g. families with multiple children). This approach uses a natural measure of association between two variables, *i.e.* the covariance between phenotypes and genotypes, and relies on a score-test. It has been implemented in the popular *Family Based Association Test* software (FBAT, Laird *et al.* [6]; Rabinowitz and Laird [7]; Lange and Laird [8]). In this context of familial samples, FBAT has proved very efficient in identifying alleles associated with many phenotypes, whether binary or quantitative (e.g. Mira *et al.* [9]; Cobat *et al.* [10]).

Although developed to handle a large variety of tests according to the nature of both the traits and their genetic determinants, it is intrinsically designed to test primary outcomes (e.g. affected vs. unaffected) as the null hypothesis is based on the same underlying principles as the TDT (*i.e.* $p = 0.5$). However, in many cases researchers are interested in the genetic contribution to a more specific phenotype (e.g. severe vs. non-severe form), here denoted as a secondary outcome. Here, what we demonstrate is the limited power of the classical formulation of the FBAT statistic to detect the effect of genetic variants that influence a secondary outcome, in particular when these variants also impact on the onset of the disease, the primary outcome. We prove that this loss of power is driven by an implicit hypothesis and we propose a derivation of the original FBAT statistic, free from this implicit hypothesis. Finally, we demonstrate analytically that our new statistic is robust and more powerful than FBAT for the detection of association between a genetic variant and a secondary outcome.

2. Original FBAT Statistic

For sake of simplicity and without major loss of generality, we consider the analysis of a diallelic marker in a sample of trios with no missing parental data under an additive genetic model. Using the same notations as in the original FBAT paper (Laird *et al.* [6]),

$$\text{let } S = \sum_i T_i X_i$$

in which X_i represents the genotype at the locus being tested and T_i the phenotype of the child of family i. The expectation of X_i is calculated conditioned on the parental genotypes under the null hypothesis of no association.

$$\text{Let } E = E(S) = \sum_i E_i = \sum_i T_i E(X_i)$$

$$\text{Let } V = \text{Var}(S) = \sum_i T_i^2 \text{Var}(X_i)$$

$$\text{FBAT} = \frac{(S-E)^2}{V}$$

$$\text{FBAT} \underset{H_0}{\sim} \chi^2_{1df}$$

Under an additive model, X_i is the number of copy of the allele under study (0, 1 or 2). As the most common way to code the phenotype is $T=1$ for affected individuals and $T=0$ for unaffected ones. In a sample with no missing parental data, unaffected individuals do not contribute to the statistic; however, in the presence of missing parental data, such unaffected individuals will indirectly impact on the statistic as they can be used to infer missing parental genotypes under some conditions (Knapp [11]). *S* is generally written as:

$$S = \sum_{i \in \text{affected}} 1 \times X_i + \sum_{i \in \text{unaffected}} 0 \times X_i = \sum_{i \in \text{affected}} X_i.$$

The null hypothesis of no association between the phenotype and a given allele is the random transmission of this allele from heterozygote parents to (affected) children. By noting p the transmission probability of this allele, the null H_0 and alternate H_1 hypotheses can be written as:

$$H_0 : p = \frac{1}{2}$$

$$H_1 : p \neq \frac{1}{2}.$$

The tested allele will be considered "at risk" or "protective" for the disease, if $p > \frac{1}{2}$ or $p < \frac{1}{2}$, respectively[1].

3. FBAT Statistic to Test Secondary Outcomes

It is common practice to study a "primary" phenotype (e.g. disease yes/no) but as stated in the introduction, researchers are often interested in the genetic contribution to a "secondary" phenotype (e.g. severe vs. non-severe form of the disease). At first glance, FBAT could be used to test this hypothesis by computing the original statistic independently in the two modalities of the secondary outcome (e.g. severe and non-severe). Denoting D_1 and D_2 the two modalities of the secondary outcome, p_1 and p_2 the transmission probabilities of the tested allele to D_1 and D_2 children, respectively, we have:

$$S_1 = \sum_{i \in 1} T_i X_i, \quad S_1 = \sum_{i \in 1} X_i, \quad \text{FBAT}_1 = \frac{(S_1 - E_1)^2}{V_1}$$

$$H_0 : p_1 = \frac{1}{2}$$

$$H_1 : p_1 \neq \frac{1}{2}$$

$$S_2 = \sum_{i \in 2} T_i X_i, \quad S_2 = \sum_{i \in 2} X_i, \quad \text{FBAT}_2 = \frac{(S_2 - E_2)^2}{V_2}$$

$$H_0 : p_2 = \frac{1}{2}$$

$$H_1 : p_2 \neq \frac{1}{2}.$$

[1]More precisely, in the general case, the null hypothesis of FBAT is "no association OR no linkage" and therefore the alternate hypothesis is "association AND linkage". H_0 can be written as a composite hypothesis: "no association AND no linkage" ∪ "no association AND linkage ∪ "association AND no linkage". In the particular case of a sample limited to trios, there is no linkage information, and the hypotheses are: H_0 = association, H_1 = no association.

However, because of the bivariate nature of the phenotype under study (*i.e.* disease AND severe form or disease AND non-severe form), rejection of the null hypothesis cannot distinguish between alleles associated with the disease *per se* (*i.e.* independently of its severity) or alleles specifically associated with the severity of the disease. FBAT offers no immediate solution to study such secondary outcomes, *i.e.* to distinguish between alleles impacting the primary (e.g. disease *per se*) or the secondary (e.g. severe vs. non-severe) outcome. Below we propose two new tests denoted as FBAT_{het} and $\text{FBAT}_{\text{het free}}$ that can be used to directly assess the association between a marker allele and a secondary outcome.

3.1. The FBAT_{het} Test

A first straightforward idea is to perform a homogeneity test of the allelic transmission rate between the two subgroups D_1 and D_2.

$$\text{Let } \text{FBAT}_{\text{het}}\left(D_1, D_2\right) = \text{homogeneity}\left(S_1, S_2\right)$$

$$= \frac{\left(S_1 - E_1\right)^2}{V_1} + \frac{\left(S_2 - E_2\right)^2}{V_2} - \frac{\left(S_1 - E_1 + S_2 - E_2\right)^2}{V_1 + V_2}$$

$$\text{FBAT}_{\text{het}} = \frac{\left(\frac{S_1 - E_1}{V_1} - \frac{S_2 - E_2}{V_2}\right)^2}{\frac{1}{V_1} + \frac{1}{V_2}} = \frac{\left(\frac{S_1 - E_1}{V_1} - \frac{S_2 - E_2}{V_2}\right)^2}{\frac{1}{V_1^2}V_1 + \frac{1}{V_2^2}V_2}$$

FBAT_{het} = FBAT with the phenotypes coded as $T = \frac{1}{V_1}$ for individuals D_1 and $T = -\frac{1}{V_2}$ for individuals D_2.

Indeed,

$$S\left(T_1 = \frac{1}{V_1}, T_2 = -\frac{1}{V_2}\right) = \sum_{i\in 1}\frac{1}{V_1}X_i - \sum_{i\in 2}\frac{1}{V_2}X_i = \frac{1}{V_1}S_1 - \frac{1}{V_2}S_2$$

$$E = \frac{1}{V_1}E_1 - \frac{1}{V_2}E_2 \quad \text{and} \quad V = \frac{1}{V_1^2}V_1 + \frac{1}{V_2^2}V_2$$

$$\text{and} \quad \text{FBAT}\left(T_1 = \frac{1}{V_1}, T_2 = -\frac{1}{V_2}\right) = \frac{\left(\frac{S_1 - E_1}{V_1} - \frac{S_2 - E_2}{V_2}\right)^2}{\frac{1}{V_1} + \frac{1}{V_2}} = \text{FBAT}_{\text{het}}.$$

The two hypotheses can then be written as:

$$H_0 : p_1 = p_2 = \frac{1}{2}$$

$$H_1 : p_1 \neq \frac{1}{2} \cup p_2 \neq \frac{1}{2}.$$

Note that under an additive genetic model and in a sample of trios with no missing parental data, coding $T_1 = \frac{1}{V_1}$ and $T_2 = -\frac{1}{V_2}$ is equivalent to coding $T_1 = \frac{1}{n_1}$ and $T_2 = -\frac{1}{n_2}$, where n_1 and n_2 are the number of heterozygote parents of children with phenotype D_1 and D_2 (see Appendix A)[2].

[2]FBAT_{het} can be implemented in FBAT by using the offset option "-o" while coding $T_1 = 1$ and $T_2 = 0$: the software then calculates, for each allele, an offset μ used to transform the phenotypic values in $T_1 = 1 - \mu$ and $T_2 = -\mu$ that minimizes the variance of the statistics. We show in Appendix B that using the offset option is equivalent to coding $T_1 = \frac{1}{V_1}$ and $T_2 = -\frac{1}{V_2}$, thus testing for secondary outcome. Here, one should not code unaffected individuals as 0 but as missing to avoid that the controls interfere in the calculation of the statistics. FBAT software can be downloaded from: http://www.biostat.harvard.edu/fbat/fbat.htm.

3.2. The FBAT$_{\text{het free}}$ Test

A somewhat hidden/under evaluated constraint of FBAT$_{\text{het}}$ is that the null hypothesis forces the transmission probabilities in both groups to be 0.5. Although valid and likely efficient in quite a number of practical situations, this can dramatically impact the power of the test in the study of a secondary outcome. A simple example being that carrying one copy of the allele is sufficient to develop the disease *per se* but that carrying two alleles will be associated with developing a severe form of the disease.

We propose a new statistic denoted as FBAT$_{\text{het free}}$ that relaxes this 0.5 constraint. Consider a diallelic locus (A and a) and denote n_{A1} (n_{A2}) the number of transmissions of allele A from Aa heterozygote parents to their children with phenotype D_1 (D_2). Then $\frac{n_{A1}+n_{A2}}{n_1+n_2}=\frac{n_A}{N}$ is the mean number of transmission of allele A from Aa heterozygote parents to affected children (whether D_1 or D_2).

Whereas in the above-mentioned FBAT and FBAT$_{\text{het}}$ tests the expected transmission of the allele of interest under the null hypothesis of no association is 0.5, in FBAT$_{\text{het free}}$ it is $\frac{n_A}{N}$. We can calculate S, E and V for FBAT, FBAT$_{\text{het}}$ and FBAT$_{\text{het free}}$. The contribution to $S-E$ of each transmission of an allele A from any Aa parent is 1/2 in FBAT and FBAT$_{\text{het}}$, and $\frac{n_A}{N}$ in FBAT$_{\text{het free}}$. Similarly, its contribution to V is 1/4 in FBAT and FBAT$_{\text{het}}$, and $\left(1-\frac{n_A}{N}\right)\frac{n_A}{N}$ in FBAT$_{\text{het free}}$ (**Figure 1**). Note that for all three statistics, the expectancy and variance of a trio including two heterozygote parents are twice those of a trio with only one heterozygote parent. Symmetrically, Aa heterozygote parents transmitting allele a each contributes for 1/2 and $\left(1-\frac{n_A}{N}\right)$ to $S-E$, and for 1/4 and $\left(1-\frac{n_A}{N}\right)\frac{n_A}{N}$ to V in FBAT or FBAT$_{\text{het}}$ and FBAT$_{\text{het free}}$, respectively. Then with $T_1=\frac{1}{n_1}$ and $T_2=-\frac{1}{n_2}$, we have:

$$\text{FBAT}_{\text{het free}}=\frac{\left(\frac{n_{A1}}{n_1}\left(1-\frac{n_A}{N}\right)+\frac{n_{a1}}{n_1}\left(0-\frac{n_A}{N}\right)-\frac{n_{A2}}{n_2}\left(1-\frac{n_A}{N}\right)-\frac{n_{a2}}{n_2}\left(0-\frac{n_A}{N}\right)\right)^2}{\frac{n_1}{n_1^2}\frac{n_An_a}{N^2}+\frac{n_2}{n_2^2}\frac{n_An_a}{N^2}}$$

$$=\frac{N}{n_1n_2n_An_a}\left(n_{A1}n_{a2}-n_{a1}n_{A2}\right)^2.$$

It is shown in Appendix C that FBAT$_{\text{het free}}$ is a Pearson's chi-squared test. In summary, the hypotheses of the FBAT$_{\text{het free}}$ test can be written as:

$$H_0: p_1 = p_2$$

$$H_1: p_1 \neq p_2.$$

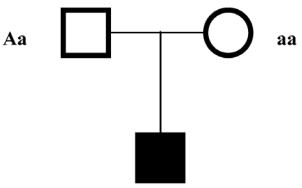
As opposed to FBAT and FBAT$_{\text{het}}$, the implicit/hidden 0.5 constraint has disappeared.

3.3. Comparison of FBAT$_{\text{het}}$ and FBAT$_{\text{het free}}$

To illustrate the magnitude of the differential power of FBAT$_{\text{het}}$ and FBAT$_{\text{het free}}$, we could have gone for large simulation studies. However, we show analytically in Appendix D that:

$$\text{FBAT}_{\text{het}}=\rho\text{FBAT}_{\text{hetfree}} \quad \text{with} \quad \rho=\frac{4n_A\left(N-n_A\right)}{N^2},\ \rho\in[0,1].$$

Trio with 1 heterozygote parent



Child genotype		*aa*	*Aa*	*AA*
	X_i	0	1	2
FBAT &FBAT$_{het}$	$p(X_i)$	$\frac{1}{2}$	$\frac{1}{2}$	0
	$X_i - E(X)$	$-\frac{1}{2}$	$\frac{1}{2}$	
FBAT$_{het\ free}$	$p(X_i)$	$1-\frac{n_A}{N}$	$\frac{n_A}{N}$	0
	$X_i - E(X)$	$-\frac{n_A}{N}$	$1-\frac{n_A}{N}$	

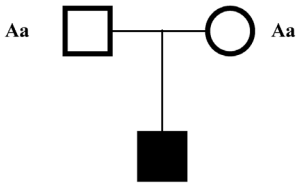
For FBAT and FBAT$_{het}$ $E(X)=\frac{1}{2}$

$\mathrm{Var}(X)=\frac{1}{4}$

For FBAT$_{het\ free}$, $E(X) = \frac{n_A}{N}$

$\mathrm{Var}(X)=\left(1-\frac{n_A}{N}\right)\frac{n_A}{N}$

Trio with 2 heterozygote parents



Child genotype		*aa*	*Aa*	*AA*
	X_i	0	1	2
FBAT &FBAT$_{het}$	$p(X_i)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
	$X_i - E(X)$	−1	0	1
FBAT$_{het\ free}$	$p(X_i)$	$\left(1-\frac{n_A}{N}\right)^2$	$2\left(1-\frac{n_A}{N}\right)\frac{n_A}{N}$	$\left(\frac{n_A}{N}\right)^2$
	$X_i - E(X)$	$-\frac{2n_A}{N}$	$1-\frac{2n_A}{N}$	$2\left(1-\frac{n_A}{N}\right)$

$E(X)=1$

$\mathrm{Var}(X)=\frac{1}{2}$

$E(X)=\frac{2n_A}{N}$

$\mathrm{Var}(X)=2\left(1-\frac{n_A}{N}\right)\frac{n_A}{N}$

Figure 1. Contribution of a trio to FBAT, FBAT$_{het}$ and FBAT$_{het\ free}$ according to the number of heterozygote parents. In a trio with one (left panel) and two (right panel) heterozygote parents, the expected genotypes *aa*, *Aa* and *AA* of the child will vary according to the statistics used. In FBAT and FBAT$_{het}$, the transmission probability of an allele A from an heterozygote parent is $\frac{1}{2}$, whereas it is $\frac{n_A}{N}$ for FBAT$_{het\ free}$ (with N denoting the total number of alleles transmitted from heterozygote parents in the whole sample, n_A the number of alleles A transmitted, and $\frac{n_A}{N}$ the mean transmission of allele *A*).

The distribution of ρ according to $\frac{n_A}{N}$ is shown in **Figure 2**. As an example, consider a sample of 300 trios with an affected child (150 D_1 and 150 D_2), all with one herterozygote parent. Consider the mean transmission of allele *A* is 0.7 in D_1 and 0.8 in D_2. Then $\frac{n_A}{N}=0.75$, $\rho=0.75$, $\mathrm{FBAT}_{het}=3$ and $\mathrm{FBAT}_{het\ free}=4$, $p(\mathrm{FBAT}_{het})=0.083$ and $p(\mathrm{FBAT}_{het\ free})=0.046$.

When there is an equivalent number of transmissions of alleles A and a from Aa heterozygote parents to their children, $n_A=n_a=\frac{N}{2}$ and $\rho=1$. In practice, this is observed when the mean transmission of allele

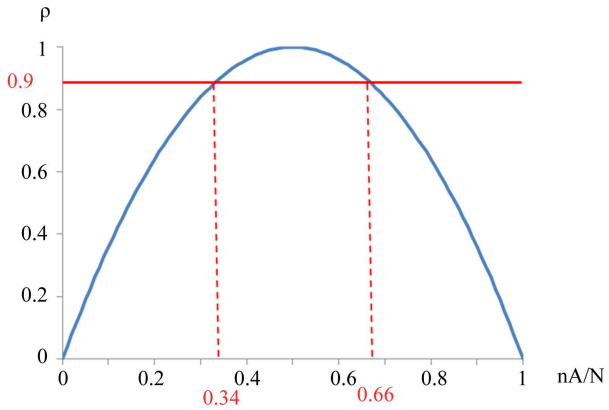


Figure 2. Distribution of ρ according to $\frac{n_A}{N}$. $\rho = \frac{4n_A(N-n_A)}{N^2}$ is the link function between $FBAT_{het}$ and $FBAT_{het\ free}$. When the mean transmission of allele A among affected cases is close to 0.5, ρ is also close from 1. When $\frac{n_A}{N} \in [0.34;0.66]$, $\rho > 0.9$.

A among all affected individuals $(D_1 + D_2)$ is 0.5. In that particular case, $FBAT_{het} = FBAT_{het\ free}$. In all other cases, $\rho < 1$ and $FBAT_{het} < FBAT_{het\ free}$ as shown in **Figure 3**.

4. Discussion

Family-based association studies have gained popularity to dissect the genetic architecture of complex traits and FBAT is likely the most popular tool to perform such studies. We have shown that at first glance it can be conveniently used to test for secondary outcomes, e.g. genetic heterogeneity between severe and non-severe forms of a disease. As an example, in a sample of trios, one can weight each "sub-phenotype" (severe and non-severe) by the inverse of the variance of each statistic. We called this test $FBAT_{het}$, for which the null and alternative hypotheses are $H_0 : p_1 = p_2 = \frac{1}{2}$ and $H_1 : p_1 \neq \frac{1}{2}$ or $p_2 \neq \frac{1}{2}$, respectively.

However, in the previous test, the transmission probabilities under the null hypothesis are fixed to 0.5 in both groups. This may not be optimal in the context of secondary outcomes when the transmission of the tested allele has already been found to significantly differ from 0.5 with respect to the primary outcome. We show that it is possible to relax this constraint by modifying the expectation in the $FBAT_{het}$ statistic so that the test is defined as $H_0 : p_1 = p_2$ and $H_1 : p_1 \neq p_2$, which are the classical hypotheses in the vast majority of homogeneity tests. This new test, $FBAT_{het\ free}$, is proven to be equivalent to a classical test for homogeneity. $FBAT_{het\ free}$ is the most powerful test when the mean transmission to affected children ($D_1 + D_2$, primary outcome) is not 0.5. Stated differently, each time an allele is found associated with the disease *per se*, $FBAT_{het\ free}$ will be the most powerful to detect heterogeneity between the transmission rates of this allele across the modalities of the secondary outcome.

For sake of simplicity, we have derived our main statistic $FBAT_{het\ free}$ in the context of the analysis of a diallelic marker under an additive genetic model in a sample of trios with no missing parental data. However, generalization to other genetic models and more complex family structures should be possible by using, for a given marker, the estimated mean transmission of the allele under study among affected individuals, in preference to the actual 0.5 that prevents testing $p_1 = p_2$. By doing so, one will be able to take advantage of all the features of FBAT ranging from the analysis of all kinds of phenotypes to the simultaneous testing of several alleles either in a classic multivariate way or taking into account the phase through haplotypic analysis.

Acknowledgements

We thank Laurent Abel, Jean-Laurent Casanova and all members of the Epidemiological Group for their support

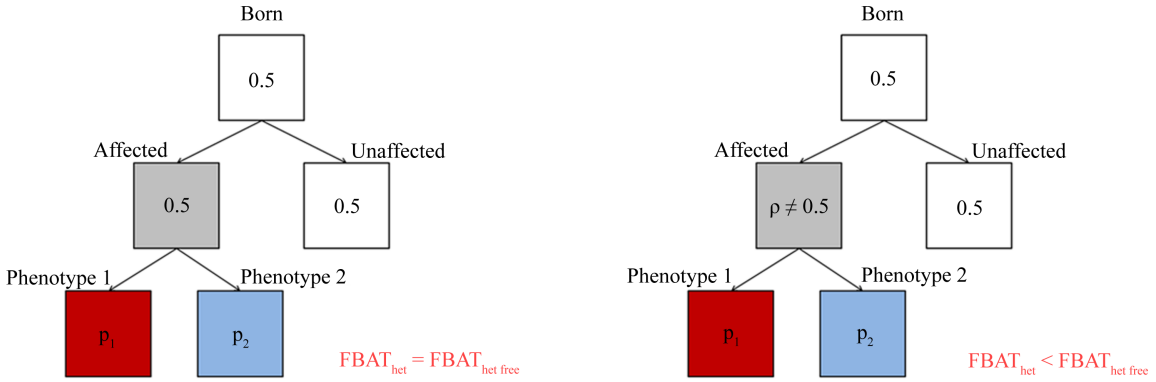


Figure 3. Power of $FBAT_{het}$ vs. $FBAT_{het\ free}$ according to the mean transmission rate of the tested allele among the affected children.

and constructive criticism. JG is funded by the Fondation pour la Recherche Médicale, and QV by the Institut Imagine. This work was supported by the Programme Blanc de l'Agence National de la Recherche.

References

[1] Laird, N.M. and Lange, C. (2006) Family-Based Designs in the Age of Large-Scale Gene-Association Studies. *Nature Reviews Genetics*, **7**, 385-394. http://dx.doi.org/10.1038/nrg1839

[2] Chen, W.M. and Abecasis, G.R. (2007) Family-Based Association Tests for Genomewide Association Scans. *American Journal of Human Genetics*, **81**, 913-926. http://dx.doi.org/10.1086/521580

[3] Burdick, J.T., Chen, W.M., Abecasis, G.R. and Cheung, V.G. (2006) In Silico Methods for Inferring Genotypes in Pedigrees. *Nature Genetics*, **38**, 1002-1004. http://dx.doi.org/10.1038/ng1863

[4] Vincent, Q., Alcais, A., Alter, A., Schurr, E. and Abel, L. (2006) Quantifying Genomic Imprinting in the Presence of Linkage. *Biometrics*, **62**, 1071-1080. http://dx.doi.org/10.1111/j.1541-0420.2006.00610.x

[5] 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 (IDDM). *American Journal of Human Genetics*, **52**, 506-516.

[6] Laird, N.M., Horvath, S. and Xu, X. (2000) Implementing a Unified Approach to Family-Based Tests of Association. *Genetic Epidemiology*, **19**, S36-S42. http://dx.doi.org/10.1002/1098-2272(2000)19:1+<::AID-GEPI6>3.0.CO;2-M

[7] Rabinowitz, D. and Laird, N. (2000) A Unified Approach to Adjusting Association Tests for Population Admixture with Arbitrary Pedigree Structure and Arbitrary Missing Marker Information. *Human Heredity*, **50**, 211-223. http://dx.doi.org/10.1159/000022918

[8] Lange, C. and Laird, N.M. (2002) Power Calculations for a General Class of Family-Based Association Tests: Dichotomous Traits. *American Journal of Human Genetics*, **71**, 575-584. http://dx.doi.org/10.1086/342406

[9] Mira, M.T., Alcais, A., Van Thuc, N., Moraes, M.O., Di Flumeri, C., Hong Thai, V., Chi Phuong, M., Thu Huong, N., Ngoc Ba, N., Xuan Khoa, P., *et al*. (2004) Susceptibility to Leprosy Is Associated with PARK2 and PACRG. *Nature*, **427**, 636-640. http://dx.doi.org/10.1038/nature02326

[10] Cobat, A., Gallant, C.J., Simkin, L., Black, G.F., Stanley, K., Hughes, J., Doherty, T.M., Hanekom, W.A., Eley, B., Jais, J.P., *et al*. (2009) Two Loci Control Tuberculin Skin Test Reactivity in an Area Hyperendemic for Tuberculosis. *Journal of Experimental Medicine*, **206**, 2583-2591. http://dx.doi.org/10.1084/jem.20090892

[11] Knapp, M. (1999) The Transmission/Disequilibrium Test and Parental-Genotype Reconstruction: The Reconstruction-Combined Transmission/Disequilibrium Test. *American Journal of Human Genetics*, **64**, 861-870. http://dx.doi.org/10.1086/302285

Appendix A. Proof That Coding $T_1 = \frac{1}{V_1}$ and $T_2 = -\frac{1}{V_2}$ Is Equivalent to $T_1 = \frac{1}{n_1}$ and $T_2 = -\frac{1}{n_2}$ under an Additive Genetic Model

Let N_1 and N_2 be the number of trios with phenotype D_1 and D_2, and N_{id} (N_{is}) the number of trios with double (d) or single (s) heterozygote parent (s). Let n_i be the number of heterozygote parents. Then

$$n_i = 2N_{id} + N_{is}.$$

Let V_s and V_d be the unitary variance for trios with 1 or 2 heterozygote parents.

For FBAT and FBAT$_{\text{het}}$, $V_s = \frac{1}{4}$ and $V_d = \frac{1}{2} = 2V_s$. Then

$$V_1 = \sum_{\substack{\text{pheno 1} \\ \&_{1 \text{ parent}}}} \text{Var}(S_j) + \sum_{\substack{\text{pheno 1} \\ \&_{2 \text{ parents}}}} \text{Var}(S_j) = N_{1s}V_s + N_{1d}V_d = \frac{N_{1s}}{4} + \frac{N_{1d}}{2} = \frac{n_1}{4}$$

$$\text{and } V_2 = \sum_{\substack{\text{pheno 2} \\ \&_{1 \text{ parent}}}} \text{Var}(S_j) + \sum_{\substack{\text{pheno 2} \\ \&_{2 \text{ parents}}}} \text{Var}(S_j) = \frac{n_2}{4}.$$

Given that $\text{FBAT}(T_1 = x, T_2 = y) = \text{FBAT}(T_1 = kx, T_2 = ky)$, coding $T_1 = \frac{1}{V_1}$ and $T_2 = -\frac{1}{V_2}$ is equivalent to $T_1 = \frac{1}{n_1}$ and $T_2 = -\frac{1}{n_2}$ for FBAT and FBAT$_{\text{het}}$.

For FBAT$_{\text{het free}}$, $V_s = \left(1 - \frac{n_A}{N}\right)\frac{n_A}{N}$ and $V_d = 2\left(1 - \frac{n_A}{N}\right)\frac{n_A}{N} = 2V_s$. Then

$$V_1 = \sum_{\substack{\text{pheno 1} \\ \&_{1 \text{ parent}}}} \text{Var}(S_j) + \sum_{\substack{\text{pheno 1} \\ \&_{2 \text{ parents}}}} \text{Var}(S_j) = (N_{1s} + 2N_{1d})\left(1 - \frac{n_A}{N}\right)\frac{n_A}{N} = n_1\left(1 - \frac{n_A}{N}\right)\frac{n_A}{N}$$

$$\text{and } V_2 = \sum_{\substack{\text{pheno 2} \\ \&_{1 \text{ parent}}}} \text{Var}(S_j) + \sum_{\substack{\text{pheno 2} \\ \&_{2 \text{ parents}}}} \text{Var}(S_j) = n_2\left(1 - \frac{n_A}{N}\right)\frac{n_A}{N}.$$

Then coding $T_1 = \frac{1}{V_1}$ and $T_2 = -\frac{1}{V_2}$ is also equivalent to $T_1 = \frac{1}{n_1}$ and $T_2 = -\frac{1}{n_2}$ for FBAT$_{\text{het free}}$.

Appendix B. Proof That $\mu = \frac{n_1}{n_1 + n_2}$ Is the Offset That Minimizes the Variance under an Additive Genetic Model

Let μ be the offset.

$$T_1 = 1 - \mu \quad \text{and} \quad T_2 = -\mu$$

With the same notations as in Appendix A,

$$\begin{aligned}\text{Var} &= \sum_{\substack{\text{pheno 1} \\ \&_{1 \text{ parent}}}} \text{Var}(S_j) + \sum_{\substack{\text{pheno 1} \\ \&_{2 \text{ parents}}}} \text{Var}(S_j) + \sum_{\substack{\text{pheno 2} \\ \&_{1 \text{ parent}}}} \text{Var}(S_j) + \sum_{\substack{\text{pheno 2} \\ \&_{2 \text{ parents}}}} \text{Var}(S_j) \\ &= N_{1s}T_1^2V_s + N_{1d}T_1^2V_d + N_{2s}T_2^2V_s + N_{2d}T_2^2V_d \\ &= (1-\mu)^2(N_{1s}V_s + N_{1d}V_d) + (-\mu)^2(N_{2s}V_s + N_{2d}V_d).\end{aligned}$$

For FBAT, $V_s = \frac{1}{4}$, $V_d = \frac{1}{2} = 2V_s$ and

$$\text{Var} = \frac{1}{4}\left((1-\mu)^2 n_1 + \mu^2 n_2\right)$$

and $\min_{\mu}(\text{Var}) = \min_{\mu}\left((1-\mu)^2 n_1 + \mu^2 n_2\right)$ is obtained for $\mu = \frac{n_1}{n_1+n_2}$.

For $\text{FBAT}_{\text{het free}}$, $V_s = \left(1-\frac{n_A}{N}\right)\frac{n_A}{N}$, $V_d = 2\left(1-\frac{n_A}{N}\right)\frac{n_A}{N} = 2V_s$ and

$$\text{Var} = \left(1-\frac{n_A}{N}\right)\frac{n_A}{N}\left((1-\mu)^2 n_1 + \mu^2 n_2\right)$$

and $\min_{\mu}(\text{Var}) = \min_{\mu}\left((1-\mu)^2 n_1 + \mu^2 n_2\right)$ is also obtained for $\mu = \frac{n_1}{n_1+n_2}$.

Appendix C. Proof That $\text{FBAT}_{\text{het free}}$ Is a Pearson's χ^2

With the notations of the manuscript, let us write the table of contingency of the transmission of alleles *A* and *a* in two phenotypic groups.

Transmission	*A*	*a*	Total
D_1	n_{A1}	n_{a1}	n_1
D_2	n_{A2}	n_{a2}	n_2
	n_A	n_a	N

Pearson's $\chi^2(n_{A1}, n_{a1}, n_{a2}, n_{A2})$

$$= \frac{\left(n_{A1} - \frac{n_A n_1}{N}\right)^2}{\frac{n_A n_1}{N}} + \frac{\left(n_{a1} - \frac{n_a n_1}{N}\right)^2}{\frac{n_a n_1}{N}} + \frac{\left(n_{A2} - \frac{n_A n_2}{N}\right)^2}{\frac{n_A n_2}{N}} + \frac{\left(n_{a2} - \frac{n_a n_2}{N}\right)^2}{\frac{n_a n_2}{N}}$$

$$= \frac{\left(n_{A1}N - n_A n_1\right)^2}{N n_A n_1} + \frac{\left(n_{a1}N - n_a n_1\right)^2}{N n_a n_1} + \frac{\left(n_{A2}N - n_A n_2\right)^2}{N n_A n_2} + \frac{\left(n_{a2}N - n_a n_2\right)^2}{N n_a n_2}$$

$$= \left(\frac{1}{N n_A n_1} + \frac{1}{N n_a n_1} + \frac{1}{N n_A n_2} + \frac{1}{N n_a n_2}\right)\left(n_{A1}n_{a2} - n_{a1}n_{A2}\right)^2$$

$$= \frac{N}{n_1 n_2 n_A n_a}\left(n_{A1}n_{a2} - n_{a1}n_{A2}\right)^2$$

$$= \text{FBAT}_{\text{het free}}.$$

Appendix D. Proof That $\text{FBAT}_{\text{free}} = \rho\text{FBAT}_{\text{het free}}$

With the notations used in the main text, for FBAT_{het},

$$S - E = \frac{1}{n_1}\left(n_{A1}\left(1-\frac{1}{2}\right) + n_{a1}\left(0-\frac{1}{2}\right)\right) - \frac{1}{n_2}\left(n_{A2}\times\left(1-\frac{1}{2}\right) + n_{a1}\left(0-\frac{1}{2}\right)\right)$$

and $V = \frac{1}{4}\left(\frac{1}{n_1}\right)^2\left(n_{A1}+n_{a1}\right) + \frac{1}{4}\left(\frac{1}{n_2}\right)^2\left(n_{A2}+n_{a2}\right) = \left(\frac{1}{n_1}\right)^2\frac{n_1}{4} + \left(\frac{1}{n_2}\right)^2\frac{n_2}{4}.$

$$\text{Then FBAT}_{\text{het}} = \frac{(S-E)^2}{V} = \frac{\left(\frac{n_{A1}}{n_1}\left(1-\frac{1}{2}\right)+\frac{n_{a1}}{n_1}\left(0-\frac{1}{2}\right)-\frac{n_{A2}}{n_2}\left(1-\frac{1}{2}\right)-\frac{n_{a2}}{n_2}\left(0-\frac{1}{2}\right)\right)^2}{\frac{1}{4}\frac{n_1}{n_1^2}+\frac{1}{4}\frac{n_2}{n_2^2}}$$

$$= \frac{4\left(n_{A1}n_{a2}-n_{a1}n_{A2}\right)^2}{Nn_1n_2} = \frac{4n_An_a}{N^2}\frac{N\left(n_{A1}n_{a2}-n_{a1}n_{A2}\right)^2}{n_An_an_1n_2} = \rho\text{FBAT}_{\text{het free}},$$

with $\rho = \frac{4n_A(N-n_A)}{N^2}$.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.