

ARTICLE

An omnibus test for family-based association studies with multiple SNPs and multiple phenotypes

Jessica Lasky-Su^{1,2,3}, Amy Murphy^{1,2,3}, Matthew B McQueen⁴, Scott Weiss^{1,2,3} and Christoph Lange^{1,3,5}

We propose an omnibus family-based association test (MFBAT) that can be applied to multiple markers and multiple phenotypes and that has only one degree of freedom. The proposed test statistic extends current FBAT methodology to incorporate multiple markers as well as multiple phenotypes. Using simulation studies, power estimates for the proposed methodology are compared with the standard methodologies. On the basis of these simulations, we find that MFBAT substantially outperforms other methods, including haplotypic approaches and doing multiple tests with single single-nucleotide polymorphisms (SNPs) and single phenotypes. The practical relevance of the approach is illustrated by an application to asthma in which SNP/phenotype combinations are identified and reach overall significance that would not have been identified using other approaches. This methodology is directly applicable to cases in which there are multiple SNPs, such as candidate gene studies, cases in which there are multiple phenotypes, such as expression data, and cases in which there are multiple phenotypes and genotypes, such as genome-wide association studies that incorporate expression profiles as phenotypes. This program is available in the PBAT analysis package.

European Journal of Human Genetics (2010) **18**, 720–725; doi:10.1038/ejhg.2009.221; published online 20 January 2010

Keywords: family-based association testing (FBAT); genome-wide association studies; FBAT-PC; multiple marker; multiple phenotypes; multiple testing

INTRODUCTION

To identify disease variants, genetic epidemiological methods perform association tests between single-nucleotide polymorphisms (SNPs) and phenotypes of interest. Most often, the number of SNPs range between hundreds in candidate gene studies to millions in genome-wide association studies. Not only can the genetic data be immense, but most often investigators analyze these SNPs with multiple phenotypes, which increases further the multiple testing problem. The number of phenotypes that are used in genetic association studies range from a few clinically relevant characteristics to tens of thousands that may originate from cellular expression measurements. With the use of multiple SNPs and phenotypes, the total number of statistical tests explodes and results in more and more stringent significance thresholds.

Typically when multiple SNPs and phenotypes are used in genetic analyses, the overarching hypothesis that researchers are interested in is whether a genetic variant has any relation to the specified disease of interest. The success of identifying these genetic variants with the marked increase in genetic data and the use of multiple phenotypes will depend to a large extent on the efficient statistical handling of these data. The immense amount of SNP and phenotype data must translate into increased statistical power to detect disease susceptibility loci.

A common approach to analyzing genetic data is to perform individual statistical tests for each SNP/phenotype combination and then adjust for the total number of tests using a multiple comparison correction method, such as Bonferroni,¹ or less conservative

corrections, such as the Holm² or Hochberg³ correction. Such approaches are limited because the number of statistical tests remains unchanged. Therefore, only association findings that can withstand multiple comparison corrections will achieve genome-wide significance in scenarios even with a moderate number of SNPs and phenotypes. The number of total association tests becomes large quickly and make it unlikely that true genetic variants will be identified.

Other methods, including global haplotype tests,⁴ regression methods,^{5,6} and multimarker tests,⁷ address multiple testing by combining multiple SNPs into a single test. Global haplotype tests enumerate the possible haplotypes and then evaluate the null hypothesis that no haplotype is associated with the disease. Regression-based approaches analyze associations between a trait and all linear combination of SNPs. Multimarker methods⁷ test multiple SNPs simultaneously by combining individual marker scores and pairwise correlations while avoiding the use of haplotype structure. Although all of these approaches dramatically reduce multiple testing over single SNP analyses, none of these methods have addressed the issue that multiple phenotypes are usually available and tested for association.

Statistical methods have also been developed to handle multiple phenotypes on a single SNP level. FBAT-PC⁸ is an approach implemented in the PBAT program⁹ that uses multiple phenotypes to construct an overall phenotype that amplifies the trait heritability at each SNP. The FBAT-PC methodology has been successful in identifying SNPs associated with complex diseases in genome-wide association studies, but has the distinct disadvantage that it requires all phenotypes to have similar distributional characteristics.

¹Channing Laboratory, Brigham and Women's Hospital, Boston, MA, USA; ²Department of Medicine, Harvard Medical School, Boston, MA, USA; ³Center for Genomic Medicine, Brigham and Women's Hospital, Boston, MA, USA; ⁴University of Colorado at Boulder, Boulder, CO, USA; ⁵Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA

*Correspondence: Dr J Lasky-Su, Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115, USA. Tel: +1 617 875 9992; Fax: +1 617 525 0958; E-mail: rejas@channing.harvard.edu

Received 2 December 2009; revised 30 October 2009; accepted 12 November 2009; published online 20 January 2010

Although many tests have been developed to aggregate either the genotype or the phenotype information, few methods have been developed that can reduce the multiple testing problem by examining marker and phenotype information simultaneously. In this manuscript, we propose a new one degree of freedom test, omnibus family-based association test (MFBAT), that incorporates both multiple markers and multiple phenotypes into a single test statistic. By reducing the number of tests across two dimensions, the total number of statistical tests is reduced to one and the multiple testing problem is addressed more efficiently than other methods. This methodology is also flexible in nature and can incorporate phenotypes that have varying distributional properties. Although this test has been developed with candidate gene studies, cellular expression phenotypes, and genome-wide association analyses in mind, the testing strategy is applicable to any family-based association analysis that uses multiple markers and multiple phenotypes. In addition, the MFBAT approach can be used in the context of the PBAT screening algorithm^{10,11} to further reduce multiple comparisons. Through simulation studies, we show that the MFBAT testing strategy outperforms standard methodology in terms of statistical power. We also provide a concrete example of the methodology using five phenotypes and five SNPs from an asthma clinical trial.

METHODS

Multi-marker/multi-phenotype: global expansion of FBAT

The general methodology is that the FBAT approach compares the difference between the observed marker score and the expected marker score that is computed based on Mendelian transmissions, conditional on the parental genotypes and the offspring's phenotype.¹² If the parental information is not available, the expected marker score is calculated conditional on the sufficient statistic for the pedigree that is then defined by the available genotypes on siblings and other relatives. As the genotype is the only random variable in the FBAT approach and its distribution is defined by Mendelian transmission, and not by any parameter estimates for the allele frequency, the FBAT approach is robust against confounding due to population admixture and stratification.

In this study we extend the FBAT methodology to generate an omnibus association test, MFBAT, that uses information on multiple markers and multiple phenotypes. For the sake of simplicity, we assume that there are trios, that is, probands and their parents, and that SNP data are analyzed. The proposed methodology extends easily to families with multiple siblings and missing parental data. If the parental data are missing/unavailable, the parental genotypes can be replaced below by the sufficient statistic^{5,12} and the construction of the weights for the overall test that we will describe below is directly extensible. For a more detailed discussion of this, we refer to the original paper.¹²

We assume that there are n parent-offspring trios, and that each offspring has m coded genotypes, $k=1,2,\dots,m$, and p traits of interest $l=1,2,\dots,p$. Let x_{ik} be the coded genotype of the k th marker for the i th proband that can be coded in an additive, dominant, or recessive manner. The variables p_{ik1} and p_{ik2} denote the parental genotypes for the parents of the i th proband at the k th marker, y_{il} denotes the l th trait of interest of the i th proband, and μ_l denotes the offset for that trait. Therefore, the FBAT statistic, $FBAT_{kb}$, tests for an association between the k th marker locus and l th trait and is given by the following:¹²

$$FBAT_{kl}^2 = \left\{ \sum_{i=1}^n (y_{il} - \mu_l) [x_{ik} - E(x_{ik}|p_{ik1}, p_{ik2})] \right\}^2 / \sum_{i=1}^n (y_{il} - \mu_l)^2 \text{Var}(x_{ik}|p_{ik1}, p_{ik2}) \sim X_1^2 \tag{1}$$

If all offspring are affected and an additive coding function is used for the genotype, the FBAT statistic and the original TDT statistic¹³ are equivalent.

Assuming that the FBAT statistic at each marker for each trait is $FBAT_{kb} \sim N(0,1)$, then an omnibus FBAT statistic that tests all phenotypes and genotypes of interest simultaneously can be constructed by the linear

combination of all FBAT statistics divided by the appropriate variance. The numerator for the MFBAT statistic is therefore given by:

$$\sum_{k=1}^m w_{kl} FBAT_{kl} \tag{2}$$

Following the work of O'Brien *et al*¹⁴ and Wei and Johnson,¹⁵ the optimal weights, w_{kl} , can be constructed as follows. Denote \mathbf{V} as the covariance matrix of $(FBAT_{1j}, \dots, FBAT_{mj})$ that is calculated empirically and the vector \mathbf{v} is defined by the expected FBAT statistic under the alternative hypothesis, that is, $E(FBAT|H_a) = [E(FBAT_{1j}|H_a), \dots, E(FBAT_{mj}|H_a)]$. Then the weights for the optimal linear combination of FBAT-statistics, that is, the linear combination with the highest statistical power under H_a is given by

$$\mathbf{W} = \mathbf{V}^{-1} \mathbf{v} \tag{3}$$

Specifically, \mathbf{W} is a vector of weights $[w_{1j}, \dots, w_{mj}]^t$ that are calculated from \mathbf{V} and \mathbf{v} . In the case in which the inverse variance matrix becomes unstable, we use the generalized inverse matrix.

Although in general, it is difficult to estimate $E(FBAT|H_a)$ before the computation of the test statistic, family-based studies allow for this estimation in a way that is statistically independent of the subsequent FBAT statistics that are computed using the conditional mean model. This method is described in detail elsewhere.¹⁶ We provide a brief summary in this study. The standard quantitative genetic model for the phenotypic mean for the k th genetic marker is given by

$$E(Y_{ikl}) = \mu_l + a_l x_{ik} \tag{4}$$

In this equation, μ_l is the overall mean for the l th phenotype and a_l is the additive genetic effect of the l th phenotype. In most cases, a_l can only be calculated using the actual genotypic data that are used in the statistical analysis itself. Family-based studies offer a unique situation in which we can generate an estimate for a_l from the data without biasing the subsequent test statistic. Because we compute the FBAT statistic using the offspring marker scores in the informative families, information from the noninformative families (ie, both parents are homozygous) can be used in the calculation of the expected value of the FBAT under the alternative hypothesis. Estimation on the sole basis of noninformative families is problematic for several reasons discussed elsewhere.¹⁶ To permit the use of both informative and noninformative families for a_l without biasing the resulting test statistic, we replace the marker score x_{ik} in Equation Eq. (4) by its expected value conditional on the parental genotypes of the i th proband at the l th phenotype and the k th genetic marker, $E(X_{ik}|p_{ik1}, p_{ik2})$. This equation is also called the conditional mean model.

$$E(Y_{ikl}|p_{ik1}, p_{ik2}) = \mu_l + E(X_{ik}|p_{ik1}, p_{ik2}) \tag{5}$$

It is important to note that these two are identical when the family is noninformative, because then the observed marker score x_{ik} and the expected marker $E(X_{ik}|p_{ik1}, p_{ik2})$ are identical, that is, $x_{ik} = E(X_{ik}|p_{ik1}, p_{ik2})$. As the test statistic is based on the use of offspring genotypes conditional on parental genotypes, the use of $E(X_{ik}|p_{ik1}, p_{ik2})$ to estimate a_l does not bias subsequent testing, even in the informative families. We can therefore use $E(X_{ik}|p_{ik1}, p_{ik2})$ in place of x_{ik} to generate an estimate of the FBAT statistic under the alternative hypothesis without biasing the subsequent statistical test. The conditional mean model has been specified similarly by Vansteem *et al*¹⁰ for continuous traits, Jiang *et al*¹⁷ for time-to-onset, and Murphy *et al*¹⁸ for affection status. Therefore, the most powerful combination of FBATs can be constructed as follows:

$$MFBAT = \sum_{k=1}^m w_{kl} FBAT_{kl} / \sqrt{w^t V w} \sim N(0, 1) \tag{6}$$

Alternatively, an omnibus test statistic could be constructed based on the multivariate score test, the FBAT-GEE,¹⁹ but the degrees of freedom for this test increases rapidly with increasing markers and phenotypes, that is, $df=mp$, making this test less optimal than MFBAT with one df proposed in this study. Similar to the approach proposed by Schaid *et al*,²⁰ this approach also only uses one degree of freedom, which makes the resulting test statistic more powerful when multiple markers and/or phenotypes are added. There are three fundamental differences between the approach proposed by Schaid *et al* and the

methodology described in this study: (1) we incorporate not only multiple markers into the test statistic, but also multiple phenotypes that the Schaid *et al* methodology does not do; (2) our approach is based on family data whereas the Schaid *et al* approach is based on case-control data and therefore does not incorporate quantitative phenotypes; and (3) the weightings used in our analysis are calculated using the expected offspring genotypes conditional on the parental genotypes, and such a calculation cannot be performed with case-control data.

RESULTS

The simulation is designed around the asthma study discussed in the data analysis section of this paper. The markers of interest comprise a five-SNP haplotype modeled after five SNPs in the *ARG1* gene. We generated the parental haplotypes by drawing from a uniform distribution, in which the probability that any parent has a given haplotype is the haplotypic frequency as measured in the Childhood Asthma Management Program (CAMP) population.²¹ The haplotypes of the probands are obtained by simulating Mendelian transmissions of the parental haplotypes, assuming complete linkage disequilibrium in each haplotype. For the computation of the MFBAT statistics, the genotypes of probands and their parents are assumed to be known.

We simulate 1000 trios with five phenotypes and five SNPs and then evaluate the power of the proposed testing strategy to other existing testing strategies. Using the haplotypes that were generated from these five SNPs in the CAMP population, the haplotypes with frequencies of 0.1, 0.2, and 0.3 are each selected to be the disease susceptibility loci, and the genotypic distribution under the alternative hypothesis is generated using $E(X_i|p_{i1}, p_{i2})$ for the marker score as described by Lange and Laird.¹⁶ The haplotypes for the remaining SNPs are simulated under the null hypothesis, assuming Hardy-Weinberg equilibrium and complete linkage disequilibrium. Five phenotypes are generated, in which one phenotype is associated with one haplotype and the four remaining phenotypes are only associated with the haplotype by their correlation with the associated phenotype. Three different phenotypic correlations were used in the simulations: (1) a low phenotypic correlation, in which the phenotypic correlation between all phenotypes is 0.2; (2) a moderate phenotypic correlation, in which the correlation between each phenotype is 0.38, which reflects the average phenotypic correlation matrix of the five asthma phenotype measurements in the CAMP clinical trial; and (3) a high phenotypic correlation, in which the phenotypic correlation between all phenotypes is 0.8. The strength of the additive effect relative to the phenotypic variance is measured by the heritability, h^2 . Specifically, when the phenotypes are generated for the simulation, they are generated using a regression equation that results in the specified heritability measurement between the phenotype and genotype. The phenotypic vector Y_i for each offspring is a random sample from a multivariate normal distribution, that is, $Y_i \sim N([a_1x_{i1}, \dots, a_5x_{i5}], V)$, in which a_l is the additive effect for the l th phenotype, x_i is the individual genotype, and V is the (5×5) variance matrix. We measure the strength of the additive genetic effect on a phenotypic trait by the heritability h^2 ,²² which is the proportion of phenotypic variation explained by the genetic variation – that is, $h^2 = \text{Var}(a_l x_i) / \text{Var}(Y_{ij})$.²² This expression for h^2 can be solved for a_l ,¹⁶ which is a measure of the genetic effect size.

We evaluate the power of the proposed testing strategy with other analysis approaches, including testing five phenotypes at five SNPs separately using the FBAT statistic (denoted as a single-SNP/single-phenotype test or SSSP), testing five phenotypes separately with a five-SNP haplotype using a family-based haplotype test (denoted as hap), testing five phenotypes at each SNP separately using the FBAT-PC

methodology (denoted as FBAT-PC), using MFBAT on each SNP but combining all five phenotypes (denoted as MFBAT-1S), testing all five SNPs with each phenotype individually using the MFBAT approach (denoted as MFBAT-1P), and testing all five SNPs and five phenotypes simultaneously using the MFBAT statistic (denoted as MFBAT). All multiple comparison corrections were made using the Bonferroni procedure with overall $\alpha=0.05$. The R^2 measures the degree of linkage disequilibrium between two SNPs, in which an R^2 of 0 indicates that there is no linkage disequilibrium between two SNPs and an R^2 of 1 indicates perfect linkage disequilibrium between two SNPs. When simulating the SNPs with an R^2 of 0.7, we used the actual structure of the SNPs in the *ARG1* gene that is used in the data application. These SNPs have the following haplotypic distributions:

Haplotype AATAT=0.534
Haplotype TGATT=0.303
Haplotype TATAT=0.108
Haplotype TGATC=0.028
Haplotype AGATT=0.019
Haplotype TGAAT=0.008

We assumed that the disease haplotype had frequencies of 0.10 (haplotype TATAT), 0.20 (we changed haplotype TGATT to have this frequency), and 0.30 (haplotype TGATT). An additive mode of inheritance was assumed in the simulations. The reported power estimates are based on 1000 replicates. The results of the simulations are shown in Table 1. Similarly, when simulating the haplotypes with an R^2 of 0.2, we used the following haplotypes:

Haplotype AACAA=0.367
Haplotype TACAA=0.129
Haplotype TGTA=0.100
Haplotype AGCGT=0.012
Haplotype AGTAA=0.066
Haplotype AGCAA=0.039
Haplotype TGTAT=0.038
Haplotype AATAA=0.029
Haplotype AACGT=0.026
Haplotype TGCGT=0.092

Because we simulate the data using a five-SNP haplotype, presumably the statistical tests using all five SNPs would have the most power. In the simulations, we use all five SNPs together (MFBAT, MFBAT-1P, haplotype) and scenarios in which we only use one SNP at a time (MFBAT-1S, FBAT-PC). Because the simulations presented in the paper analyze the power using all five SNPs individually and together, we believe that we have captured the power that MFBAT approach has to detect the DSL in the two extreme situations, one in which what is being tested is exactly the DSL (ie, when all five markers are used simultaneously) and one in which we are testing only a part of the true DSL. The results of the simulations are shown in Table 1.

In all simulation scenarios, MFBAT or MFBAT-1S had the highest power estimates of the approaches that were used, and in many cases these approaches substantially outperformed both the haplotype and SSSP approaches. As MFBAT-1S and FBAT-PC are methodologically very similar, it is not surprising that these two approaches have comparable power estimate in all of the tested scenarios. When the R^2 between the SNPs is 0.20, MFBAT is consistently the highest powered of the various testing strategies regardless of the haplotype

Table 1 Power simulations of the MFBAT methodology compared with the standard approaches using 1000 trios

Heritability	Average R^2	Phenotypic correlation	Haplotype frequency	FBAT-PC	MFBAT-1P	MFBAT-1S	MFBAT	Haplotype	SSSP		
0.02	0.70	0.2	0.1	0.02	0.02	0.02	0.07	0.03	0.003		
			0.2	0.11	0.06	0.11	0.11	0.02	0.02		
			0.3	0.11	0.06	0.11	0.14	0.02	0.02		
		0.4	0.1	0.02	0.02	0.02	0.09	0.03	0.003		
			0.2	0.14	0.06	0.15	0.14	0.02	0.02		
			0.3	0.15	0.06	0.16	0.14	0.02	0.02		
		0.8	0.1	0.03	0.02	0.03	0.10	0.03	0.003		
			0.2	0.59	0.06	0.59	0.16	0.02	0.02		
			0.3	0.59	0.06	0.59	0.17	0.02	0.02		
	0.2	0.2	0.1	0.02	0.03	0.02	0.10	0.03	0.004		
			0.2	0.05	0.04	0.05	0.14	0.06	0.01		
			0.3	0.05	0.04	0.05	0.14	0.05	0.01		
		0.4	0.1	0.03	0.03	0.04	0.14	0.03	0.006		
			0.2	0.05	0.03	0.05	0.17	0.06	0.007		
			0.3	0.08	0.04	0.08	0.18	0.05	0.01		
		0.8	0.1	0.13	0.02	0.12	0.44	0.03	0.005		
			0.2	0.24	0.03	0.24	0.64	0.07	0.008		
			0.3	0.31	0.04	0.31	0.64	0.05	0.01		
		0.05	0.70	0.2	0.1	0.06	0.08	0.06	0.25	0.18	0.009
					0.2	0.50	0.15	0.51	0.29	0.17	0.09
					0.3	0.50	0.15	0.51	0.30	0.17	0.09
0.4	0.1			0.08	0.09	0.08	0.22	0.18	0.01		
	0.2			0.60	0.16	0.60	0.30	0.17	0.09		
	0.3			0.78	0.20	0.78	0.36	0.17	0.14		
0.8	0.1			0.29	0.08	0.29	0.33	0.18	0.01		
	0.2			0.81	0.15	0.82	0.34	0.17	0.09		
	0.3			0.80	0.20	0.80	0.38	0.17	0.14		

Estimated power levels to detect the causal haplotype using 1000 trios and five phenotypes and five SNPs, assuming complete linkage disequilibrium among markers and an additive mode of inheritance. The significance level is set to 5%. These simulations incorporate the Bonferroni correction when multiple testing has occurred.

frequency or the phenotypic correlation. In this scenario, MFBAT has between two and three times the power of the haplotype approach and over 10 times the power of SSSP. In fact, when the phenotypic correlation is 0.8 and R^2 is 0.20, MFBAT is nine times as powerful as the haplotypic approach and over 60 times as powerful as SSSP. When the R^2 between the SNPs is 0.70 and the phenotypic correlation is moderate (0.4) or low (0.20) with a minor allele frequency >0.2 , then the power of FBAT-PC, MFBAT-1S, and MFBAT do not deviate notably from each other. The power of any of these three approaches is between 2 and 10 times higher than the overall haplotype test and SSSP. When R^2 is high (0.7) and the phenotypic correlation is high (0.8), then MFBAT-1S and FBAT-PC substantially outperform the other approaches, with three times the power of MFBAT and approximately 30 times the power of the haplotypic approach or SSSP. In virtually all of the simulations, SSSP has the lowest power, which is not surprising because there are many comparisons that need to be adjusted for and the correlation between tests is not accounted for in any way. When the heritability is increased from 1 to 5%, the overall power increases, but the relative relationships among the statistical tests remains constant. In addition, simulations using an R^2 of 0.5 show results in between 0.2 and 0.7 (data not shown).

We generated simulations under the null hypothesis of no genetic effect and calculated the empirical type I error rates for the MFBAT under three scenarios: (1) single phenotype and five SNPs, (2) five phenotypes and single SNP, and (3) five phenotypes and five SNPs. Simulations of 10 000 replicates were performed at minor allele

frequencies of 10 and 30% and the α was set at 5.0%. These results were compared with FBAT-PC, in which accurate type I error rates has been established many times. In all cases, the type I error rate does not vary by more than 0.2% above or below 5.0%. This small deviation suggests that the different MFBAT scenarios do not vary from the selected alpha level in any meaningful or systematic way. Therefore, we conclude that the type I error is being maintained appropriately.

Data analysis: an asthma study (CAMP)

We applied MFBAT to a collection of parent/child trios in the CAMP Genetics Ancillary Study. The CAMP study randomized asthmatic children to three different asthma treatments.²¹ Blood samples for DNA were collected from 696 complete parent/child trios from 640 nuclear families in the CAMP Ancillary Genetics Study. Genotyping was performed on five polymorphic loci located in the *ARG1* gene.²³ Previously, an association with the *ARG1* gene and bronchodilator response was identified and replicated in three independent populations.²³ Therefore, we used this gene along with five phenotypes, including bronchodilator response and four measures of pulmonary function. The related phenotypes were pre- and post-bronchodilator measures of forced expiratory volume in one second (FEV_1) and forced vital capacity (FVC). FEV_1 is the amount of air that can be forcibly blown out of the lung in 1 s. FVC is the total amount of air that can be forcibly blown out of the lung after full inspiration. Therefore, the following phenotypes were used in the analysis:

Table 2 Analysis of ARG1 and five asthmatic phenotypes

Test statistic	SNP	Phenotype	P-value	Significance level
MFBAT	All SNPs	All phenotypes	9.53×10^{-6}	0.05
MFBAT-1P	All SNPs	Bronchodilator response	0.95	0.01
	All SNPs	Post-FVC	0.244	
	All SNPs	Post-FEV1	0.718	
	All SNPs	Pre-FVC	0.54	
	All SNPs	Pre-FEV1	0.623	
MFBAT-1S	RS2749935	All phenotypes	0.03	0.01
	RS2781659	All phenotypes	0.0059	
	RS2781663	All phenotypes	0.0036	
	RS2781665	All phenotypes	0.022	
	RS3756780	All phenotypes	0.284	
FBAT-PC	RS2749935	All phenotypes	0.036	0.01
	RS2781659	All phenotypes	0.007	
	RS2781663	All phenotypes	0.005	
	RS2781665	All phenotypes	0.017	
	RS3756780	All phenotypes	0.444	
Haplotype	All SNPs	Bronchodilator response	0.404	0.01
	All SNPs	Post-FVC	0.021	
	All SNPs	Post-FEV1	0.117	
	All SNPs	Pre-FVC	0.02	
	All SNPs	Pre-FEV1	0.391	
SSSP	RS2749935	Bronchodilator response	0.714	0.002
	RS2781659	Bronchodilator response	0.895	
	RS2781663	Bronchodilator response	0.599	
	RS2781665	Bronchodilator response	0.905	
	RS3756780	Bronchodilator response	0.541	
	RS2749935	Post-FEV1	0.76	
	RS2781659	Post-FEV1	0.695	
	RS2781663	Post-FEV1	0.868	
	RS2781665	Post-FEV1	0.765	
	RS3756780	Post-FEV1	0.932	
	RS2749935	Post-FVC	0.639	
	RS2781659	Post-FVC	0.296	
	RS2781663	Post-FVC	0.547	
	RS2781665	Post-FVC	0.405	
	RS3756780	Post-FVC	0.588	
	RS2749935	Pre-FEV1	0.76	
	RS2781659	Pre-FEV1	0.588	
	RS2781663	Pre-FEV1	0.64	
	RS2781665	Pre-FEV1	0.709	
	RS3756780	Pre-FEV1	0.738	
	RS2749935	Pre-FVC	0.813	
	RS2781659	Pre-FVC	0.04	
	RS2781663	Pre-FVC	0.089	
	RS2781665	Pre-FVC	0.072	
	RS3756780	Pre-FVC	0.458	

This table provides the association results for the five SNPs in *ARG1* and five pulmonary function phenotypes. This example is used for comparison.

measurements of (1) bronchodilator response, (2) pre-bronchodilator response to FEV₁, (3) post-bronchodilator response to FEV₁, (4) pre-bronchodilator response to FVC, and (4) post-bronchodilator response to FVC.

FEV₁ is known to be an important asthma phenotype that also depends on the age, height, weight, and sex of the individual. It is standard practice to adjust the FEV₁ measurements for these covariates before analyzing them. Therefore, using the conditional mean

model approach,¹⁶ we regressed the FEV₁ and FVC measurements on the recorded values for age, height, weight, and sex and used the residuals in MFBAT.

Table 2 lists the results from the various analyses that were used in the power simulations sections. We list the statistical test, the SNPs analyzed, the phenotypes used, the observed *P*-values, and the significance level. The column labeled significance level refers to the significance level (ie, alpha) that is necessary when accounting for the multiple testing. When there is one statistical test, then the *P*-value must be <0.05 to be significant whereas when there are five statistical tests, then the *P*-value must be <0.01 to be significant.

All five SNPs in the *ARG1* gene were also included in the analysis. The data were analyzed in six ways: (1) using MFBAT with all phenotypes and SNPs, (2) using MFBAT with all phenotypes and each SNP individually, (3) using MFBAT with all SNPs and each phenotype individually, (4) using the FBAT-PC association test for each SNP individually, (5) using a haplotype test for each phenotype individually, and finally (6) analyzing each SNP/phenotype combination individually. The results of these findings are summarized in Table 2. The minor allele frequency of the SNPs in this analysis ranged between 0.46 and 0.03, whereas the average phenotypic correlation was 0.38 and the average *R*² was in between 0.2 and 0.7. In this example, significant associations are identified with FBAT-PC, MFBAT-1S, and MFBAT; however, the association findings for FBAT-PC and MFBAT-1S are not nearly as striking at low *P*-value observed using MFBAT (*P*-value= 9.53×10^{-6}).

From the data analysis we see that the MFBAT analysis, using all phenotypes and all SNPs simultaneously, has the strongest genetic association with a *P*-value of 9.5×10^{-6} . When we analyzed the association *P*-values for individual SNPs and phenotypes, there is no specific phenotype/SNP combination that seems to be driving the effect. When analyzing the FBAT-PC and MFBAT-1S, it seems that rs2781659 and rs2781663 are driving the association. It also seems as if the pre- and post-FVC measures have the strongest association. When examined individually, however, the association is not as strong as when all SNPs and phenotypes are examined together. This is likely because the combinations of phenotypes are more informative than any single phenotype that is evaluated individually.

From the simulations described in Table 1, MFBAT would have the greatest power to detect an association, followed by MFBAT-1S and FBAT-PC. The haplotype and SSSP analyses would have the lowest power to detect an effect. In this analysis, MFBAT resulted in an overall *P*-value on the order of 10^{-6} , which clearly shows a very strong association with the asthma-related phenotypes. Two SNPs were significantly associated with the phenotypes for both MFBAT-1S and FBAT-PC. As was observed in the simulations, these two test statistics have similar results and their association *P*-values track together closely. Both the haplotype analysis and SSSP had no significant associations after adjusting for multiple comparisons. This analysis illustrates that the MFBAT approach, which uses all of the information simultaneously, stands out as a possible association *P*-value that one would carry forward for further analysis.

DISCUSSION

In this manuscript we propose a new testing methodology in which a test statistic, MFBAT, is generated by using information across multiple SNPs and phenotypes. This approach was developed to help reduce the overall number of statistical tests, a problem that has prohibited researchers from establishing genome-wide significance in many studies, most notably in genome-wide association analyses. The simulation studies suggest that MFBAT or MFBAT-1S have the

most power when compared with standard methods, particularly when the correlation among SNPs is low and the phenotypic correlation is high. One advantage of the MFBAT approach is that it can easily incorporate multiple phenotypes with varying characteristics. With a group of affected parent–offspring trios, the clinical traits of interest are often both quantitative and qualitative in nature. Therefore, having a statistical test that can incorporate any combination of qualitative and quantitative traits offers a distinct advantage over other statistical tests that can only incorporate either quantitative or qualitative measures. This is a great advantage over FBAT-PC, which has similar power estimates to MFBAT-1S, but is restricted in using quantitative phenotypes with distributions similar to each other. To date, we know of no other statistical test for family data that can incorporate this type of phenotypic data in such an efficient manner.

It is important to point out that when using the MFBAT test, we will be identifying a genetic association among a group of genetic markers and phenotypes. That is, we are not testing any single phenotype or single SNP directly. Once a significant association is found with multiple SNPs and phenotypes, the approach by which to proceed may be divided between two differing philosophies. The first philosophy would seek to identify the single-SNP/phenotype combination that is largely responsible for the effect. In this case, one may deem it imperative to determine what specific SNP/phenotype combination is driving the effect. There are several ways to determine this. First, one can analyze the weightings that are applied for each phenotype/genotype combination in generating the overall test statistic. Second, once a significant association is determined, individual tests can be performed to determine what phenotypes and genotypes are driving the effect, similar to what is performed for analysis of variance tests in classical statistics. The second philosophy regarding a significant association while using multiple SNPs and phenotypes argues something completely different. This perspective would suggest that the optimal phenotype is not any single one of the phenotypes that were used, but that a combination of the selected phenotypes best describes the clinical condition that is being analyzed. Similarly, this philosophy would suggest that it is not likely to be any single SNP that is driving the association, but that several of SNPs are in linkage disequilibrium with an unknown function genetic variant responsible for the observed effect. Therefore, if one's beliefs fall into this philosophy, then the next step would be to perform subsequent statistical and genetic testing on the entire region in which the association effect was identified.

In conclusion, we propose a new statistical test, MFBAT, that can be used to test multiple phenotypes and genetic markers simultaneously, thereby reducing the multiple testing problem. This test has good statistical power, is easy to use, flexible in nature, and is incorporated into the PBAT program.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank all subjects for their ongoing participation in this study. We acknowledge the CAMP investigators and research team, supported by NHLBI, for collection of CAMP Genetic Ancillary Study data. All work on data collected

from the CAMP Genetic Ancillary Study was conducted at the Channing Laboratory of the Brigham and Women's Hospital under appropriate CAMP policies and human subject's protections. The CAMP Genetics Ancillary Study is supported by U01 HL075419, U01 HL65899, P01 HL083069, R01 HL086601, and T32 HL07427 from the National Heart, Lung and Blood Institute and the National Institute of Health. We would also like to acknowledge the support from the following grants: R01MH081862, P01 HL083069, U01 HL065899.

- 1 Bonferroni C: *Teoria statistica delle classi e calcolo delle probability* Volime in Onore di Ricardo della Volta. Universita di Firenze, 1937.
- 2 Holm S: A simple sequentially rejective multiple test procedure. *Scand J Statist* 1979; **6**: 65–70.
- 3 Hochberg Y: A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 1988; **75**: 800–802.
- 4 Horvath S, Xu X, Lake SL, Silverman EK, Weiss ST, Laird NM: Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. *Genet Epidemiol* 2004; **26**: 61–69.
- 5 Clayton D, Chapman J, Cooper J: Use of unphased multilocus genotype data in indirect association studies. *Genet Epidemiol* 2004; **27**: 415–428.
- 6 Chapman JM, Cooper JD, Todd JA, Clayton DG: Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. *Hum Hered* 2003; **56**: 18–31.
- 7 Rakovski CS, Xu X, Lazarus R, Blacker D, Laird NM: A new multimarker test for family-based association studies. *Genet Epidemiol* 2007; **31**: 9–17.
- 8 Lange C, van Steen K, Andrew T *et al*: A family-based association test for repeatedly measured quantitative traits adjusting for unknown environmental and/or polygenic effects. *Stat Appl Genet Mol Biol* 2004; **3**: 1–27.
- 9 Lange C, DeMeo D, Silverman EK, Weiss ST, Laird NM: PBAT: tools for family-based association studies. *Am J Hum Genet* 2004; **74**: 367–369.
- 10 Van Steen K, McQueen MB, Herbert A *et al*: Genomic screening and replication using the same data set in family-based association testing. *Nat Genet* 2005; **37**: 683–691.
- 11 Van Steen K, Lange C: PBAT: a comprehensive software package for genome-wide association analysis of complex family-based studies. *Hum Genomics* 2005; **2**: 67–69.
- 12 Rabinowitz D, Laird N: A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Hum Hered* 2000; **50**: 211–223.
- 13 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.
- 14 O'Brien PC: Procedures for comparing samples with multiple endpoints. *Biometrics* 1984; **40**: 1079–1087.
- 15 Wei L, Johnson W: Combining dependent tests with incomplete measurements. *Biometrika* 1985; **72**: 359–364.
- 16 Lange C, Laird NM: On a general class of conditional tests for family-based association studies in genetics: the asymptotic distribution, the conditional power, and optimality considerations. *Genet Epidemiol* 2002; **23**: 165–180.
- 17 Jiang H, Harrington D, Raby BA *et al*: Family-based association test for time-to-onset data with time-dependent differences between the hazard functions. *Genet Epidemiol* 2006; **30**: 124–132.
- 18 Murphy A, Weiss ST, Lange C: Screening and replication using the same data set: testing strategies for family-based studies in which all probands are affected. *PLoS Genet* 2008; **4**: e1000197.
- 19 Lange C, Silverman EK, Xu X, Weiss ST, Laird NM: A multivariate family-based association test using generalized estimating equations: FBAT-GEE. *Biostatistics* 2003; **4**: 195–206.
- 20 Schaid DJ, Sinnwell JP, Thibodeau SN: Robust multipoint identical-by-descent mapping for affected relative pairs. *Am J Hum Genet* 2005; **76**: 128–138.
- 21 Long-term effects of budesonide or nedocromil in children with asthma: The Childhood Asthma Management Program Research Group. *N Engl J Med* 2000; **343**: 1054–1063.
- 22 Falconer DS, Mackay TFC: *Introduction to Quantitative Genetics*. New York: Longman, 1997.
- 23 Litonjua AA, Lasky-Su J, Schneider K *et al*: ARG1 is a novel bronchodilator response gene: screening and replication in four asthma cohorts. *Am J Respir Crit Care Med* 2008; **178**: 688–694.