

# Genomic screening and replication using the same data set in family-based association testing

Kristel Van Steen<sup>1</sup>, Matthew B McQueen<sup>2</sup>, Alan Herbert<sup>3</sup>, Benjamin Raby<sup>4</sup>, Helen Lyon<sup>4,5</sup>, Dawn L DeMeo<sup>4</sup>, Amy Murphy<sup>1</sup>, Jessica Su<sup>2</sup>, Soma Datta<sup>4</sup>, Carsten Rosenow<sup>6</sup>, Michael Christman<sup>3</sup>, Edwin K Silverman<sup>4</sup>, Nan M Laird<sup>1</sup>, Scott T Weiss<sup>4</sup> & Christoph Lange<sup>1,4</sup>

The Human Genome Project and its spin-offs are making it increasingly feasible to determine the genetic basis of complex traits using genome-wide association studies. The statistical challenge of analyzing such studies stems from the severe multiple-comparison problem resulting from the analysis of thousands of SNPs. Our methodology for genome-wide family-based association studies, using single SNPs or haplotypes, can identify associations that achieve genome-wide significance. In relation to developing guidelines for our screening tools, we determined lower bounds for the estimated power to detect the gene underlying the disease-susceptibility locus, which hold regardless of the linkage disequilibrium structure present in the data. We also assessed the power of our approach in the presence of multiple disease-susceptibility loci. Our screening tools accommodate genomic control and use the concept of haplotype-tagging SNPs. Our methods use the entire sample and do not require separate screening and validation samples to establish genome-wide significance, as population-based designs do.

In humans, SNPs are the most common type of genetic variation; eight million SNPs have already been documented and deposited in the dbSNP database<sup>1,2</sup>. Their dense distribution across the genome and their low mutation rate make them ideal markers for large-scale genome-wide association studies for common complex diseases<sup>3</sup>.

The success of genome-wide association studies will depend on whether the increase in numbers of SNPs can be translated into increased statistical power or whether the positive effects of the increased number of SNPs will be diluted by the multiple-comparison problem. When thousands of SNPs are tested for association, the *P* values need to be adjusted for the number of tests computed to control type I error rates, which include the family-wise error rate and the false discovery rate (FDR). Multiple-testing procedures such as those proposed by Bonferroni<sup>4</sup> and Hochberg<sup>5</sup> adjust *P* values to control the family-wise error rate. They often generate unrealistically small significance levels for the individual tests, in part because the dependence between test statistics is ignored. Alternative multiple-testing approaches control the FDR<sup>6,7</sup>. Most procedures become more conservative as more tests are done.

Ideally, SNP-reduction techniques are applied first, so that the number of association tests is diminished and the correction for multiple testing is less severe. To avoid biasing test results, the data used in the reduction process should differ from the data used for testing. For family data, it is possible to create two sources of information<sup>8,9</sup> using one sample. The basic idea is to estimate

the genetic effect using a regression model that is statistically independent of the family-based analysis, using data from all families. The genetic effect estimate for each SNP is used to screen and select SNPs for association testing. The association testing on a much smaller set of SNPs uses family-based tests (FBATs), which are robust to population admixture.

Here we report new strategies for genomic screening. We derived lower bounds for the estimated power of the screening method to detect a gene carrying a disease-susceptibility locus (DSL), regardless of the linkage disequilibrium (LD) structure. We show that population stratification has a minimal effect on power and illustrate the potential of the approach for genome-wide association studies using the software package PBAT<sup>10,11</sup>.

## RESULTS

### Simulation studies: power

We assessed the power of the testing strategies using simulations. We used 291 SNPs in candidate genes from 651 trios in the Childhood Asthma Management Program (CAMP) Genetics Ancillary Study<sup>12</sup>, who were affected with mild to moderate asthma. We chose the interleukin gene *IL10* on chromosome 1 as the DSL<sup>13</sup>. We selected each of the six typed SNPs in *IL10* individually as the DSL and, for each offspring, simulated a trait value  $Y_{ij}$  from the normal distribution with unit variance  $Y_{ij} \approx N(aX_{ij}, 1)$ , where *a* denotes the genetic effect size and  $X_{ij}$  denotes the observed marker score of the selected SNP in

Departments of <sup>1</sup>Biostatistics and <sup>2</sup>Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA. <sup>3</sup>Department of Genetics and Genomics, Boston University School of Medicine, Boston, Massachusetts 02115, USA. <sup>4</sup>Channing Laboratory, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>5</sup>Division of Genetics, Children's Hospital, Boston, Massachusetts 02115, USA. <sup>6</sup>Genomics Collaboration Genotyping, Affymetrix, Inc., Santa Clara, California 95051, USA. Correspondence should be addressed to K.V.S. (kvanstee@hsph.harvard.edu).

Published online 5 June 2005; doi:10.1038/ng1582

**Table 1** Estimated power for gene or SNP selection

Method	<i>h</i>	Causal mutation in CAMP <i>IL10</i>					
		SNP1	SNP2	SNP3	SNP4	SNP5	SNP6
Method I (top 1)	0.05	0.726 (0.368)	0.745 (0.741)	0.673 (0.271)	0.730 (0.548)	0.481 (0.479)	0.695 (0.590)
	0.07	0.871 (0.454)	0.873 (0.872)	0.833 (0.351)	0.865 (0.680)	0.561 (0.561)	0.812 (0.746)
	0.10	0.936 (0.514)	0.943 (0.943)	0.922 (0.403)	0.943 (0.773)	0.687 (0.687)	0.933 (0.874)
Method II (top 1)	0.05	0.842 (0.646)	0.884 (0.864)	0.891 (0.683)	0.843 (0.447)	0.813 (0.813)	0.898 (0.598)
	0.07	0.967 (0.798)	0.976 (0.964)	0.978 (0.800)	0.955 (0.512)	0.967 (0.967)	0.947 (0.620)
	0.10	0.998 (0.881)	0.998 (0.994)	0.999 (0.851)	0.999 (0.559)	0.996 (0.996)	1.000 (0.710)
Method III	0.05	0.074 (0.059)	0.408 (0.408)	0.203 (0.155)	0.076 (0.063)	0.265 (0.265)	0.235 (0.229)
	0.07	0.172 (0.129)	0.649 (0.647)	0.364 (0.294)	0.167 (0.154)	0.389 (0.389)	0.443 (0.437)
	0.10	0.309 (0.257)	0.882 (0.882)	0.622 (0.556)	0.332 (0.315)	0.584 (0.582)	0.735 (0.724)
Method	<i>h</i>	Causal mutation in Affymetrix block					
		SNP1	SNP2	SNP3	SNP4		
Method I (top 1)	0.05	0.587 (0.268)	0.690 (0.264)	0.455 (0.091)	0.527 (0.054)		
	0.07	0.771 (0.400)	0.841 (0.333)	0.783 (0.116)	0.794 (0.066)		
	0.10	0.950 (0.511)	0.964 (0.379)	0.958 (0.125)	0.967 (0.069)		
Method II (top 1)	0.05	0.406 (0.152)	0.460 (0.092)	0.318 (0.046)	0.365 (0.122)		
	0.07	0.686 (0.293)	0.739 (0.116)	0.688 (0.130)	0.720 (0.241)		
	0.10	0.957 (0.345)	0.950 (0.179)	0.958 (0.167)	0.937 (0.373)		
Method III	0.05	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)		
	0.07	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.008 (0.000)		
	0.10	0.032 (0.032)	0.040 (0.032)	0.057 (0.057)	0.049 (0.041)		
Method IV	0.05	0.024 (0.008)	0.008 (0.008)	0.008 (0.008)	0.000 (0.000)		
	0.07	0.041 (0.041)	0.008 (0.008)	0.033 (0.033)	0.024 (0.016)		
	0.10	0.153 (0.153)	0.113 (0.105)	0.098 (0.098)	0.146 (0.138)		

Estimated power levels to detect *IL10* using data from 291 SNPs in the CAMP study or a selected LD block of four SNPs from 10,000 SNPs in Affymetrix data, based on simulations with 10,000 replicates. Heritability (*h*) values are in the range 0.05–0.10. The nominal significance level is set to 5%. Method I uses conditional power calculations. Method II uses the overall Wald test for genetic effects. Method III uses the FDR controlling approach<sup>7</sup> to calculate power levels using adjusted *P* values. Method IV uses the FDR corrective approach<sup>6</sup>. Values in parentheses are estimates of power to detect the simulated causal mutation.

*IL10* for the *j*th offspring in family *i*. We specified the genetic effect size in terms of the heritability *h*:  $h = \text{Var}(aX_{ij})/\text{Var}(Y_{ij})$ , where *h* is the proportion of phenotypic variance explained by the marker<sup>14</sup>. We selected the most promising SNP from the entire pool, in terms of highest power (method I) or smallest *P* value for the Wald test (method II), and tested it for association using the FBAT statistic<sup>15</sup>.

We estimated the power level as the proportion of trials successfully identifying the DSL in *IL10* (the SNP in *IL10* is selected by the screening technique and the corresponding FBAT statistic is significant at  $\alpha = 5\%$ ). In general, screening 291 SNPs using method I gives power levels of  $\geq 60\%$ , and power levels of  $> 80\%$  with heritability values  $\geq 0.07$  (Table 1). We obtained similar results with method II (Table 1), which outperformed method I. Our methods did not perform well for SNP5 in *IL10*, which is a rare SNP (minor allele frequency of 0.055) and is not correlated with other SNPs in *IL10*.

To assess the power of our screening technique with thousands of markers, we carried out a simulation study using genetic data from an Affymetrix GeneChip Mapping 10K Array. Blood samples were collected from 467 subjects in 167 families, and SNPs were typed using the Affymetrix 10,000 SNP mapping array<sup>16</sup>. Data from a genome linkage screen were previously reported<sup>17</sup>. We selected an LD block of four SNPs as the region carrying the DSL in our simulations. We selected each SNP as the alleged DSL and estimated power levels. We observed a loss of power when we increased the number of SNPs to 10,000 (Table 1). Power levels were acceptable for larger heritability values ( $h > 0.05$ ). Method I was more powerful than method II for heritability values  $\geq 0.05$ . The simulations assume

one DSL, with the functional variant being an observed SNP. Only the top trait-marker combination is retained for further FBAT testing. More trait-marker pairs can be pushed forward to the second level of the screening method, but there is a trade-off between the number of such pairs and the loss in power owing to controlling type I error for FBAT results on those pairs.

To simulate a more realistic situation where the causal SNP is not observed, we chose each SNP as the alleged DSL, simulated trait values as described above, removed the DSL from the pool of SNPs and determined the most promising SNP from the pool using screening method I or II. We estimated power levels as the proportion of trials successfully identifying the *IL10* gene in CAMP or in the Affymetrix LD block of size 4. The identification is successful if one of the SNPs in the gene or block is selected by the screening technique and found significant by the FBAT statistic at a 5% level (Table 2).

We achieved adequate power to detect the gene carrying the DSL for larger heritability values using only the top screening selection. Low heritability values ( $\leq 0.05$ ) may result in poor power, particularly for low allele frequencies and large numbers of SNPs. Even though the SNP pool does not include the DSL, high correlations between SNPs ensure that the gene containing the mutation is identified by the screening technique (SNP3 and SNP6;  $r^2 = 0.91$ ). When we selected SNP5 in *IL10* as the unobserved DSL, neither screening method (I and II) had power to identify *IL10*; SNP5 is poorly correlated with the other five SNPs in *IL10* (Table 2). The high correlations among the four Affymetrix SNPs in the LD block resulted in similar power levels (Tables 1 and 2).

**Table 2 Power to detect indirect association**

Method	<i>h</i>	Causal mutation in CAMP <i>IL10</i>					
		SNP1	SNP2	SNP3	SNP4	SNP5	SNP6
Method I (top 1)	0.05	0.694	0.167	0.653	0.664	0.005	0.557
	0.07	0.838	0.324	0.809	0.817	0.007	0.679
	0.10	0.923	0.569	0.913	0.918	0.012	0.882
Method II (top 1)	0.05	0.750	0.258	0.849	0.788	0.004	0.837
	0.07	0.918	0.456	0.955	0.934	0.006	0.909
	0.10	0.989	0.709	0.998	0.991	0.011	0.995
Method III	0.05	0.059	0.017	0.161	0.041	0.000	0.109
	0.07	0.133	0.039	0.309	0.085	0.006	0.229
	0.10	0.253	0.069	0.558	0.186	0.002	0.460

Method	<i>h</i>	Causal mutation in Affymetrix block			
		SNP1	SNP2	SNP3	SNP4
Method I (top 1)	0.05	0.544	0.678	0.455	0.527
	0.07	0.757	0.833	0.790	0.808
	0.10	0.935	0.957	0.958	0.974
Method II (top 1)	0.05	0.370	0.460	0.318	0.351
	0.07	0.650	0.739	0.696	0.731
	0.10	0.935	0.950	0.958	0.944
Method III	0.05	0.000	0.000	0.000	0.000
	0.07	0.000	0.000	0.000	0.008
	0.10	0.024	0.016	0.057	0.049
Method IV	0.05	0.016	0.008	0.000	0.000
	0.07	0.041	0.008	0.033	0.024
	0.10	0.065	0.105	0.082	0.130

Estimated power levels (%), using simulations with 10,000 replicates based on the data from CAMP SNPs ( $m = 291$ ) or Affymetrix SNPs ( $m = 10,000$ ), when the causal SNP is removed from the SNP set tested. Heritability ( $h$ ) values are in the range 0.05–0.10. Either the probability that *IL10* is selected by one of the six available SNPs by first-level screening and is significant in terms of the FBAT statistic at the 5% level or the probability that one of four SNPs in a fixed block is selected by first-level screening and is significant in terms of the FBAT statistic at the 5% level is shown, using method I (based on conditional power), method II (based on the overall Wald test for genetic effects) or controlling FDR in method III (ref. 7) or method IV (ref. 6). The  $r^2$  measures of LD for CAMP SNP pairs ( $i, j$ ) are as follows: (1,2) is 0.11; (1,3) is 0.32; (1,4) is 0.94; (1,5) is 0.02; (1,6) is 0.31; (2,3) is 0.33; (2,4) is 0.11; (2,5) is 0.02; (2,6) is 0.33; (3,4) is 0.31; (3,5) is 0.07; (3,6) is 0.91; (4,5) is 0.02; (4,6) is 0.30; and (5,6) is 0.07. The  $r^2$  measure of LD for Affymetrix SNP pairs ( $i, j$ ) are as follows: (1,2) is 0.92; (1,3) is 0.92; (1,4) is 0.88; (2,3) is 0.96; (2,4) is 0.92; and (3,4) is 0.92.

We also compared PBAT screening tools with procedures to control FDR at a 5% level on the basis of reports by Benjamini and Yekutieli<sup>7</sup> (method III) and Benjamini and Hochberg<sup>6</sup> (method IV). To account for general dependencies, such as those arising from LD patterns between SNPs in candidate genes, we included only results for method III (ref. 7) for the CAMP data in **Tables 1** and **2**. The power to detect the gene carrying the DSL, or the SNP itself, was always higher for PBAT screening methods. For the Affymetrix data and large heritability values ( $\geq 0.07$ ), our screening methodology outperformed method III, with up to 30 times greater power (**Table 1**). We reached similar conclusions regarding method IV (ref. 6). Differences in power performance between selections based on FDR controlling and selections based on our screening techniques were even more apparent when the disease-causing mutation was not being tested (**Table 2**). Although method IV is expected to be less conservative than method III, both failed to detect a significant association in the Affymetrix data. Because our results (**Tables 1** and **2**) refer to probabilities of both selecting a gene or SNP (using methods I or II) and obtaining a significant FBAT statistic at the 5% level, power estimates can be smaller than the nominal significance level.

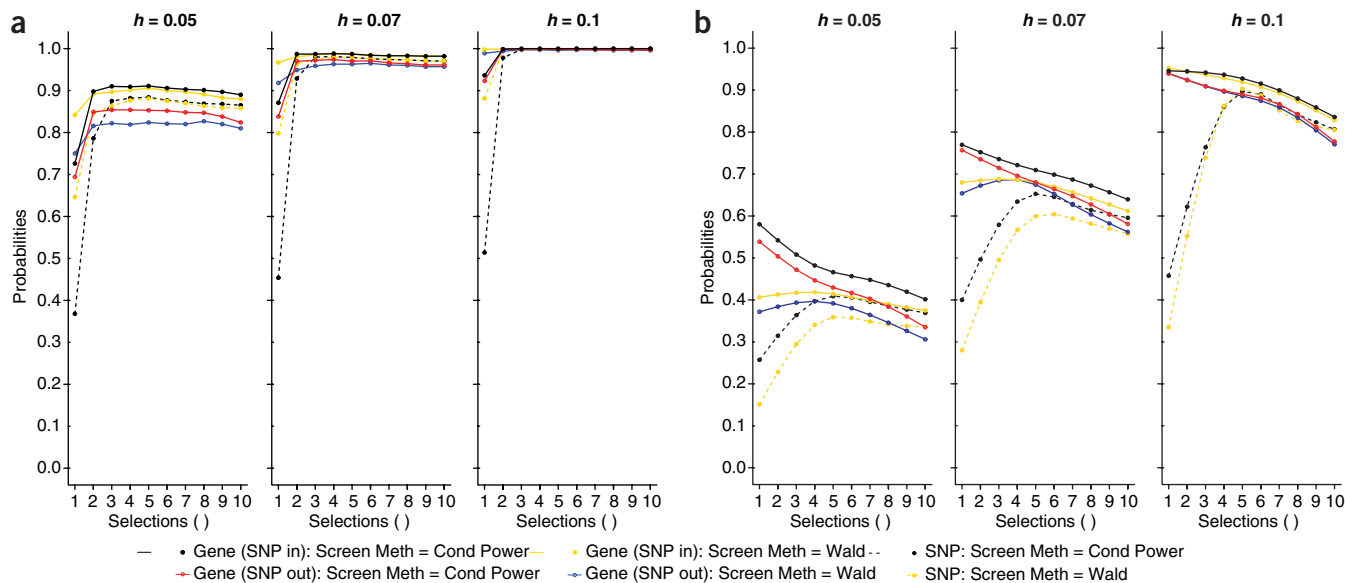
### Guidelines for the number of top selections

To address the issue of selecting a single trait-marker combination versus multiple combinations, we repeated the simulations, defining the power to identify the DSL as the probability that it is observed in

the top  $K$  ranking SNPs of the chosen screening method and results in a significant FBAT statistic at the 5% level, Bonferroni-corrected for  $K$  comparisons made. We defined the power to find the gene carrying the DSL in a similar fashion. The simulation results using genotype data from the CAMP study and the Affymetrix 10,000 SNP platform are illustrated in **Figure 1**.

With respect to SNP selection, it is beneficial to include more than one top selection with either screening method (**Fig. 1**). At some point, these benefits are offset by the correction for multiple testing. This became apparent when we used relatively few SNPs for screening (**Fig. 1a**). For large SNP pools, selecting the top five combinations gave an acceptable balance between detecting associations and reducing power (**Fig. 1b**).

Our results for gene selection using large data sets (Affymetrix data) favor top trait-marker selection. The power to detect the gene carrying the DSL increases with increasing heritability and decreases with increasing number of top selections. We observed a steeper decrease in power when we removed the DSL from the data set. For the smaller CAMP sample (**Fig. 1a**), retaining a single trait-marker combination did well for screening on the basis of Wald tests (method II) but was less powerful for screening on the basis of power (method I). We observed similar trends when we removed the DSL from the total set of SNPs being screened (**Fig. 1**), but power was generally lower. Overall, screening on the basis of power and retaining the five most promising trait-marker combinations seemed to yield excellent results.



**Figure 1** Plots of power versus the number of top trait-marker combination retained after first-level screening under different screening scenarios. **(a)** 291 SNPs from CAMP. **(b)** The 10K Affymetrix setting. Method I is based on conditional power sorting (high to low), and method II is based on ranking the  $P$  values (low to high) obtained by the Wald test for genetic effects. Solid lines indicate probabilities of identifying the gene carrying the disease mutation; dashed lines indicate probabilities of identifying the causal variant. Filled symbols indicate that the disease mutation SNP1 is included in the SNP search set.

Our simulations reconfirm the belief that LD patterns are important in terms of power to detect SNPs or genes. High correlations exist between SNP1 and the other three SNPs in the gene selected for the Affymetrix-based simulations. If the DSL was removed from the search SNP pool, the gene or region of interest could still be picked up by a single SNP that is in high correlation with the removed SNP. In the CAMP data sets, retaining more than one marker-trait combination boosted the power to detect the gene. We could increase the signal by adding more trait-marker pairs for further testing, but adding too many combinations could decrease the signal.

### Type I error

By simulating the quantitative trait from a normal distribution with mean 0 and variance 1, we assessed the significance levels attained by our two-level testing strategy. For each SNP-trait combination, we either estimated the power or computed the Wald test and then selected the most promising SNP to be tested for association with the FBAT statistic. We estimated the empirical significance levels as the proportion of SNPs that are selected by the screening technique and then found significant by the FBAT statistic at the 5% level.

For 100,000 replicates, the estimated significance levels for the 291 CAMP SNPs were 4.92% and 4.87% for methods I and II, respectively. Both screening methods maintained the specified significance level. Under the null hypothesis of no association, the screening techniques and the FBAT statistic are statistically independent<sup>8,9</sup>. The significance level of the FBAT statistic is therefore expected to be the significance level of the overall procedure. We obtained similar results using the 10K Affymetrix SNP data. For  $\alpha = 5\%$ , the estimated significance levels were 4.86% and 5.14% for methods I and II, respectively.

### Analytical power considerations

We analytically derived the power of PBAT screening techniques for a single quantitative trait and one DSL. The null hypothesis assumes no

association and no linkage between any SNP and the DSL. Throughout, we assumed that  $n$  trios were genotyped. The quantitative trait is described by an additive model  $Y_i = \mu_i + a_k X_{ik}$ , where  $Y_i$  denotes the quantitative trait for the  $i$ th proband,  $a_k$  denotes the genetic effect size for the  $k$ th SNP and  $X_{ik}$  denotes the marker score for the  $k$ th SNP in the  $i$ th offspring. The number of genotyped SNPs is  $m$ . We take  $\mu_i$  to be 0 and  $\text{Var}(Y_i)$  to be 1. We assume that the first SNP is the DSL ( $a_1 > 0$  and  $a_k = 0$ , where  $k = 2, \dots, m$ ).

The power of the screening technique is defined as  $\text{power}^{\text{PBAT}} = P(\text{true SNP is selected and FBAT statistic is significant}) = P(\text{FBAT statistic is significant} \mid \text{true SNP is selected}) \times P(\text{true SNP is selected})$ . Although the probability  $P(\text{FBAT statistic is significant} \mid \text{true SNP is selected})$  can be computed directly<sup>18,19</sup>, a lower bound for the probability  $P(\text{true SNP is selected})$  is given by the equation

$$\begin{aligned} P(\text{true SNP is selected}) &= P(\hat{a}_1 > \hat{a}_k, k = 2, \dots, m) \\ &\geq [P_{N(-a_1, \frac{2}{n \cdot p_{\min} (3p_{\min} + 1)}})(Z < 0)]^{(m-1)} \\ &= \pi_{h, p_{\min}, n, m}, \end{aligned}$$

where  $\hat{a}_k$  is the conditional mean model based least-squares estimator for  $a_k$ ,

$$\hat{a}_k = \frac{\sum_i E(X_{ik} \mid P_{ik1}, P_{ik2}) Y_i}{\sum_i E(X_{ik} \mid P_{ik1}, P_{ik2})^2}$$

and  $p_{\min} = \min_{1 < k \leq m} (p_k)$ , where  $p_k$  is the target allele frequency of the  $k$ th SNP (**Supplementary Note** online).

Because we ignored LD structures between SNPs, the lower bound for the overall power is probably too conservative. But given promising probabilities  $P(\text{FBAT statistic is significant} \mid \text{true SNP is selected})$ , high values for  $\pi_{h, p_{\min}, n, m}$  are indicative of adequate power levels.

Values for  $\pi_{h, p_{\min}, n, m}$  under various settings are shown in **Table 3**. The factor  $\pi_{h, p_{\min}, n, m}$  is modestly influenced by the minimum of all allele frequencies,  $p_{\min}$ . The influence of the number of SNPs on

**Table 3 Analytical power approximations**

<i>n</i>	$p_{\min}$	<i>h</i>	Power FBAT	<i>m</i>						
				250	500	1,000	10,000	100,000	500,000	
700	0.001	0.05	0.050	0.7454	0.5550	0.3077	0.0000	0.0000	0.0000	0.0000
		0.07	0.050	0.9662	0.9334	0.8711	0.2512	0.0000	0.0000	0.0000
		0.10	0.050	0.9988	0.9975	0.9950	0.9515	0.6083	0.0833	0.0833
	0.010	0.05	0.874	0.7830	0.6124	0.3747	0.0001	0.0000	0.0000	0.0000
		0.07	0.925	0.9737	0.9480	0.8986	0.3431	0.0000	0.0000	0.0000
		0.10	0.962	0.9991	0.9983	0.9966	0.9663	0.7094	0.1797	0.1797
	0.100	0.05	0.981	0.9676	0.9361	0.8761	0.2660	0.0000	0.0000	0.0000
		0.07	0.997	0.9984	0.9968	0.9936	0.9376	0.5248	0.0398	0.0398
		0.10	1.000	1.0000	1.0000	0.9999	0.9994	0.9942	0.9714	0.9714
1,500	0.001	0.05	0.270	0.9989	0.9979	0.9957	0.9583	0.6529	0.1186	0.1186
		0.07	0.275	1.0000	1.0000	0.9999	0.9995	0.9949	0.9749	0.9749
		0.10	0.279	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.010	0.05	0.997	0.9993	0.9985	0.9971	0.9711	0.7461	0.2312	0.2312
		0.07	0.999	1.0000	1.0000	1.0000	0.9997	0.9970	0.9851	0.9851
		0.100	1.000	1.0000	1.0000	1.0000	0.9995	0.9953	0.9770	0.9770
2,000	0.001	0.05	0.455	1.0000	0.9999	0.9999	0.9986	0.9864	0.9337	0.9337
		0.07	0.466	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
		0.10	0.481	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	0.010	0.05	1.000	1.0000	1.0000	0.9999	0.9992	0.9917	0.9591	0.9591
		0.07	1.000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999
		0.100	1.000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9998

Estimated lower bounds  $\pi_{h,p_{\min},n,m}$  for PBAT genome-wide association screening technique and a selection of different sample sizes *n*, minimum allele frequency  $p_{\min}$  (over all available SNPs in the data), heritability values *h* and number of SNPs *m*. Simulations are based on 10,000 replicates. Power FBAT values are power levels for a single SNP analysis, obtained by unconditional power calculations by approximation<sup>19</sup>. Data for heritability settings (0.05, 0.07 or 0.10) where both the estimated power for FBAT and the  $\pi_{h,p_{\min},n,m}$  equal 1 are omitted.

$\pi_{h,p_{\min},n,m}$  is almost diminished by the level of heritability *h* and sample size *n*. But the power levels for single SNP analysis (Table 3) are dictated by allele frequencies and the minimum of all allele frequencies,  $p_{\min}$ .

Assuming 1,500 trios, power levels for heritability values in the range of 0.07–0.20 exceed 70% with five million SNPs. Power may drop below 50% when more than 13.6 million SNPs are analyzed at once, but only for moderately low heritability values ( $\sim 0.07$ ). For heritability values  $> 0.10$ , power is maintained, even with three billion SNPs, in which case power still exceed 85%. For low heritability values ( $\leq 0.05$ ) and a low disease allele frequency ( $< 0.001$ ), power levels drop below 50% when more than 162,500 SNPs are genotyped.

Our screening technique uses the conditional mean model in which the estimated effects might be distorted by population stratification. Genomic control<sup>20</sup> does not affect PBAT screening method II. Power is reduced with increasing degrees of population stratification using method I but remains acceptable (Supplementary Note and Supplementary Fig. 1 online).

### Multiple DSLs

So far we have assumed a single DSL in the sample. Using the CAMP and Affymetrix genotypic data as a platform, we selected five regions in each data set and selected one DSL in every region. We generated traits from a normal distribution, including as many as five genetic contributions. For each replicate, we generated heritability values for each locus from a uniform distribution with a mean *h* value of either 0.03 (for all loci considered) or 0.05 (assuming that when multiple loci elevate disease risk, their singular effects are small).

The screening results using methods I and II are reported in Table 4. Although we explored the effects on power of selecting the top *K* SNPs, using various values for *K*, we show results only for selecting the top five or ten SNPs. Overall, selecting the ten most

promising combinations for subsequent FBAT testing works best for multiple DSLs. Comparing methods I and II in the presence of multiple DSLs, screening on the basis of power outperformed screening method II in both data sets. For the CAMP data, excellent power was achieved to detect as many as three DSLs with method I, even for heritability values as low as 0.03. For method II, this observation held for detecting as many as two DSLs. Using the Affymetrix data (Table 4), screening on the basis of power in the presence of multiple DSLs gave excellent power to detect two of the loci. Given four DSLs, the power to detect two of them was 0.754 and to detect three of them was 0.246 (Table 4).

The FDR methods (methods III and IV) had modest power to detect one locus (Table 4). There was no power to detect multiple loci using the 10K Affymetrix platform. In general, the power to detect any DSL with methods III and IV was much lower than with method I. This effect was amplified when the number of SNPs was increased. For  $h = 0.03$  and using the CAMP data set, the power estimates with method I and the top ten most promising combinations were  $\sim 20$  times higher than those for method III. Similar conclusions held for method IV (Table 4).

### CAMP data analysis

Asthma is a complex genetic disorder with increasing prevalence and substantial heritability<sup>21</sup>. To date, five genes involved in asthma have been identified by positional cloning. More than 200 positive genetic association studies of asthma and its phenotypes have been reported<sup>22</sup>.

We applied our screening technique to a data set including parent-child trios in the CAMP Genetics Ancillary Study, which included 1,041 asthmatic children randomly assigned to one of three treatment groups (CAMP 1999). Appropriate informed consent from all participating subjects was obtained. Blood samples from probands and parents were collected, and 291 SNPs in 701 children from 651

**Table 4 Power to identify genes given multiple DSLs**

CAMP														
Method I	DSLs	Identified genes ( $h = 0.03$ )					DSLs	Identified genes ( $h = 0.05$ )						
		1	2	3	4	5		1	2	3	4	5		
Method I	2	0.732	0.229	–	–	–	2	0.764	0.250	–	–	–		
		0.971	0.668					0.981	0.682					
	3	0.927	0.553	0.128	–	–	3	0.939	0.642	0.199	–	–		
		0.997	0.913	0.460				1.000	0.979	0.700				
	4	0.890	0.487	0.121	0.005	–	4	0.948	0.676	0.266	0.042	–		
		0.994	0.842	0.394	0.039			0.998	0.976	0.742	0.278			
	5	0.844	0.417	0.075	0.003	0.000	5	0.925	0.618	0.230	0.032	0.000		
		0.979	0.727	0.282	0.047	0.003		0.995	0.926	0.672	0.243	0.033		
	Method II	DSLs	Identified genes ( $h = 0.03$ )					DSLs	Identified genes ( $h = 0.05$ )					
			1	2	3	4	5		1	2	3	4	5	
	Method II	2	0.545	0.086	–	–	–	2	0.514	0.085	–	–	–	
			0.804	0.274					0.766	0.249				
		3	0.740	0.289	0.031	–	–	3	0.720	0.270	0.039	–	–	
			0.919	0.562	0.116				0.899	0.551	0.128			
		4	0.790	0.336	0.049	0.003	–	4	0.825	0.414	0.111	0.009	–	
			0.961	0.675	0.227	0.024			0.963	0.751	0.363	0.082		
		5	0.753	0.278	0.038	0.001	0.000	5	0.804	0.383	0.089	0.011	0.001	
			0.946	0.562	0.178	0.016	0.001		0.959	0.706	0.321	0.071	0.004	
		Method III	DSLs	Identified genes ( $h = 0.03$ )					DSLs	Identified genes ( $h = 0.05$ )				
				1	2	3	4	5		1	2	3	4	5
Method III		2	0.153	0.014	–	–	–	2	0.471	0.098	–	–	–	
		3	0.358	0.098	0.005	–	–	3	0.804	0.458	0.116	–	–	
		4	0.416	0.142	0.035	0.002	–	4	0.855	0.615	0.298	0.058	–	
		5	0.439	0.150	0.025	0.000	0.000	5	0.861	0.519	0.221	0.027	0.002	
Affymetrix														
Method I		DSLs	Identified genes ( $h = 0.03$ )					DSLs	Identified genes ( $h = 0.05$ )					
			1	2	3	4	5		1	2	3	4	5	
Method I		2	0.646	0.000	–	–	–	2	0.969	0.000	–	–	–	
			0.665	0.000					0.984	0.000				
	3	0.776	0.079	0.000	–	–	3	0.984	0.390	0.000	–	–		
		0.823	0.063	0.000				0.996	0.287	0.000				
	4	0.846	0.247	0.010	0.000	–	4	0.972	0.643	0.116	0.003	–		
		0.914	0.255	0.025	0.000			0.997	0.754	0.246	0.015			
	5	0.730	0.205	0.005	0.000	0.000	5	0.947	0.534	0.051	0.000	0.000		
		0.822	0.222	0.025	0.000	0.000		0.987	0.696	0.185	0.005	0.000		
	Method II	DSLs	Identified genes ( $h = 0.03$ )					DSLs	Identified genes ( $h = 0.05$ )					
			1	2	3	4	5		1	2	3	4	5	
	Method II	2	0.098	0.000	–	–	–	2	0.323	0.000	–	–	–	
			0.236	0.000					0.504	0.000				
		3	0.118	0.000	0.000	–	–	3	0.260	0.008	0.000	–	–	
			0.307	0.004	0.000				0.472	0.024	0.000			
		4	0.106	0.003	0.000	0.000	–	4	0.268	0.005	0.000	0.000	–	
			0.424	0.008	0.000	0.000			0.711	0.099	0.003	0.000		
		5	0.072	0.000	0.000	0.000	0.000	5	0.182	0.005	0.000	0.000	0.000	
			0.345	0.008	0.000	0.000	0.000		0.648	0.089	0.000	0.000	0.000	

Table 4 continued on following page

Table 4 Continued

Affymetrix													
		Identified genes ( $h = 0.03$ )							Identified genes ( $h = 0.05$ )				
Methods III and IV	DSLs	1	2	3	4	5	DSLs	1	2	3	4	5	
	2	0.059	0.000	–	–	–	2	0.413	0.000	–	–	–	
		0.201	0.000					0.587	0.000				
	3	0.138	0.000	0.000	–	–	3	0.630	0.004	0.000	–	–	
		0.303	0.000	0.000				0.799	0.004	0.000			
	4	0.258	0.000	0.000	0.000	–	4	0.770	0.018	0.000	0.000	–	
		0.485	0.010	0.000	0.000			0.909	0.071	0.000	0.000		
	5	0.368	0.000	0.000	0.000	0.000	5	0.833	0.003	0.000	0.000	0.000	
		0.563	0.008	0.000	0.000	0.000		0.937	0.033	0.003	0.000	0.000	

Estimated power levels using simulations (10,000 replicates) to detect a given number of genes among multiple DSLs based on the top five (upper row) or top ten (lower row) SNPs in the CAMP ( $m = 291$ ) or the Affymetrix ( $m = 10,000$ ) genetic data sets. The average heritability ( $h$ ) of a DSL in simulating trait values is either 0.03 or 0.05 for all loci considered. The nominal significance level is set to 5%.

pedigrees were genotyped. As a quantitative phenotype, we used the log of LNPC20 scores (a measure of airway responsiveness) measured repeatedly over time.

We first selected the baseline values at randomization and then adjusted them for age, age of onset, weight, height and gender. We restricted the analysis to individuals of European ancestry. We screened 291 SNPs using methods I and II, with a significance level of 5%. We used recessive genetic models (ref. 23 and R. Lazarus *et al.*, unpublished data). To ensure the asymptotic validity of the FBAT statistic, we did not calculate it when fewer than 20 families were informative.

Table 5 shows the selected trait-marker combinations, from a first-level screening by selecting the five highest power levels. After we

reduced the total number of SNP-trait combinations to five, the *IL10* SNP  $-627A \rightarrow C$  reaches significance with a  $P$  value of 0.0058 (compared with  $0.05/5 = 0.01$  using Bonferroni correction for five comparisons). Screening results after selecting the five smallest  $P$  values for the Wald test are also shown in Table 5. This screening method did not identify significant associations. Not accounting for multiple testing, only four FBAT  $P$  values in the entire data set were  $<0.01$ . The smallest FBAT  $P$  value was 0.0022 (power = 0.0003; WALD  $P$  value = 0.6944).

We next analyzed the development of LNPC20-measurements, using the FBAT-PC statistic<sup>24</sup> (Table 5). LNPC20 measurements at five time points were available. The FBAT-PC approach aggregates and amplifies the heritability values of each measurement and is a powerful

Table 5 Association of SNPs with LNPC20 measure of airway responsiveness

Gene	SNP	rs number	Allele	Inf. families	Power	Wald $P$ value	FBAT $P$ value	$h$
Univariate FBAT								
<i>IKBKAP</i>	L1023L	rs11791783	C	220	0.2831	$1.5381 \times 10^{-2}$	0.1921	0.0306
<i>IL10</i>	$-627A \rightarrow C$	rs1800872	A	67	0.2138	$2.4775 \times 10^{-2}$	0.0058	0.0656
<i>IL10</i>	$-854T \rightarrow C$	rs1800871	T	78	0.2019	$8.1205 \times 10^{-2}$	0.0355	0.0596
<i>Tbx21</i>	$-16115A \rightarrow G$	rs1808192	T	142	0.1381	$6.3913 \times 10^{-2}$	0.7327	0.0264
<i>VDR</i>	I353I	rs731236	G	179	0.1364	$1.4245 \times 10^{-2}$	0.4800	0.0364
<i>VDR</i>	I352I	rs731236	G	179	0.1364	$1.4245 \times 10^{-2}$	0.4800	0.0364
<i>IKBKAP</i>	L1023L	rs11791783	C	220	0.2831	$1.5381 \times 10^{-2}$	0.1921	0.0306
<i>IKBKAP</i>	I816L	rs10759326	A	233	0.1087	$1.6590 \times 10^{-2}$	0.0172	0.0347
<i>ADRB2</i>	I164T	rs1800888	C	22	0.0147	$2.1569 \times 10^{-2}$	0.9755	0.0217
<i>IL10</i>	$-1117A \rightarrow G$	rs1800896	A	194	0.1120	$2.1745 \times 10^{-2}$	0.4073	0.0426
Multivariate FBAT								
<i>VDR</i>	$34059A \rightarrow C$	rs7975232	A	221	0.7814	$2.0505 \times 10^{-3}$	0.9122	
<i>TLR4</i>	$-6143A \rightarrow G$	rs1927914	A	276	0.7407	$2.9748 \times 10^{-4}$	0.0086	
<i>CRHBP</i>	$-8093C \rightarrow T$	rs1700676	T	135	0.6970	$5.0940 \times 10^{-2}$	0.4878	
<i>VDR</i>	I352I	rs731236	G	180	0.6922	$6.4160 \times 10^{-3}$	0.5728	
<i>ADRB2</i>	$45702C \rightarrow T$	rs1036173	T	242	0.6035	$1.0130 \times 10^{-2}$	0.5434	
<i>PPARG</i>	$30132C \rightarrow G$	rs709150	G	184	0.0999	$2.7953 \times 10^{-8}$	0.9890	
<i>LOX</i>	$-2241T \rightarrow G$	rs840466	T	58	0.2989	$3.9514 \times 10^{-8}$	0.3382	
<i>CRHR1</i>	$49823C \rightarrow A$	rs242949	C	218	0.0663	$7.6767 \times 10^{-8}$	0.7501	
<i>IL12B</i>	$11776C \rightarrow A$	rs1368439	A	187	0.0478	$8.8965 \times 10^{-8}$	0.5671	
<i>IL13</i>	R130Q	rs20541	G	218	0.0330	$1.6390 \times 10^{-7}$	0.1063	

Data analysis results from screening a moderate number of SNPs using the CAMP data. The reported FBAT  $P$  values are not corrected for multiple testing. Method I is based on conditional power calculations; method II is based on the Wald test for genetic effects. The last column shows the estimated proportions of phenotypic variance explained by the analyzed SNP ( $h$ ).

statistic for any type of multivariate or longitudinal data. Screening on the basis of power highlighted a SNP in the gene *TLR4* (−6143A → G; allele A,  $P = 0.0086$ ) with a significant FBAT  $P$  value after Bonferroni correction for five comparisons. Screening method II did not identify significant associations. Not accounting for multiple testing, only five FBAT  $P$  values in the entire data set were  $<0.01$ . The smallest FBAT  $P$  value was 0.0025 (power = 0.0092; WALD  $P$  value = 0.3171).

The number of comparisons using the screening technique is smaller than the number of possible tests (440), which would lead to a Bonferroni correction of  $0.05/440 = 0.0001$ . We computed the adjusted FBAT  $P$  values using the procedures of Bonferroni<sup>4</sup>, Holm<sup>25</sup>, Hochberg<sup>5</sup>, Benjamini and Hochberg<sup>6</sup>, Sidak<sup>26</sup> and Benjamini and Yekutieli<sup>7</sup>. None of these methods identified significant associations.

Observed discrepancies between the population-based  $P$  values (from Wald tests) and the family-based  $P$  values (from the FBAT statistic) are due to the fact that the Wald statistic uses between-family information in all families whereas the FBAT statistic uses only within-family information from informative families.

## DISCUSSION

With advances in genotyping technology, genome-wide association using thousands of markers will be standard in the near future. In such studies, larger numbers of association tests are computed and correction of statistical significance for multiple testing is warranted. In this paper, we address the multiple-testing problem, the most important statistical hurdle in genome-wide family-based association studies.

There are two types of association tests: population-based and family-based. In the first design, subjects are cross-classified by genotype and data are analyzed by  $\chi^2$  tests or logistic regression models. In family-based designs, allele-transmission rates are used. Compared with case-control studies, family data elevate genotyping costs by recruiting parents or additional siblings but help to resolve haplotype phase. Family-based designs are not susceptible to confounding due to population substructure. Family-based tests assess both association and linkage, whereas population-based data allow testing of association only.

Our studies show that genome-wide association studies have the best power when family data with at least 1,500 probands are available, when the minor allele frequency for SNPs in the analysis is at least 0.01 and when heritability is moderate ( $>0.03$ ). Under these conditions, genome-wide association studies with as many as one million SNPs can be successful. The information gained from the increased number of SNPs are not diluted by the multiple-comparison problem. The number of SNPs has only a modest effect on the overall power of our screening techniques, which are robust against effects of population stratification and admixture.

To identify one DSL, it is best to screen on the basis of power (method I) and to consider the five most promising combinations for FBAT testing. To identify multiple loci of small effects, a substantial gain in power can be achieved when considering the ten most promising combinations in stage I for further FBAT testing. The better performance of method I for large numbers of SNPs may alter views on current strategies<sup>27</sup> for genome-wide association studies in case-control designs.

Because of the methodological issues involved in analyzing thousands of SNPs, investigators have searched for methods to reduce the amount of data, using LD patterns, with a minimum loss of information<sup>28</sup>. Leaving aside whether or not SNPs should be selected on the basis of haplotype blocks<sup>29</sup>, the end product is a smaller set of tagging SNPs, often with low LD between them. Using simulations, we showed that our proposed screening techniques do not require *a priori*

identification of causal variants to identify disease-associated regions; their success relies on the assumption that untyped DSLs are correlated with one or more typed SNPs. From this perspective, adding more SNPs is beneficial, because it increases potential LD with the DSL(s). In general, adding more SNPs comes at the cost of power loss when corrections for multiple testing need to be applied. Our screening methods are only moderately affected by adding 'noncausal' SNPs.

Type I error rates can also be controlled by permutation-based methods<sup>30,31</sup>. Such permutation procedures are numerically challenging for large data sets with thousands of SNPs. As the field is moving towards genotyping arrays with more than 500,000 SNPs, small  $P$  values are expected, and the number of permutations needs to be large to guarantee acceptable accuracy levels. In PBAT, a genomic screen of 2,000 trios genotyped on 300,000 SNPs takes 1 day on a single processor. Alternatively, the transmitted and untransmitted status of alleles from parents to offspring can be randomized. This approach was adopted by Lin *et al.*<sup>32</sup> in the context of genome-wide association studies using the classical TDT test<sup>33</sup> but excludes LD structure between SNPs.

Our screening tools increase detection power, provide an analytical method to use genotype data in full and can be used with multivariate quantitative traits, time-to-onset traits, covariate adjustment, multiple alleles, haplotypes, extended pedigrees and missing parents. Compared with FDR methods, our screening tools are particularly suited to detect multiple trait-influencing loci in data sets with thousands of SNPs, even for low heritability values. Unlike FDR methods, our screening technique does not necessarily identify the SNP with the smallest FBAT  $P$  value but combines  $P$  value information with an estimate for genetic effect. Furthermore, our screening technique does not require a screening and replication sample; both screening and replication steps can be accomplished in a single data set.

## METHODS

**New tool for genome-wide association screening.** Our tools for genomic association screening are implemented in the PBAT software<sup>11</sup> and use the unified approach to FBAT<sup>34,35</sup>. FBAT builds on the original TDT method<sup>33</sup> and has been generalized to accommodate quantitative phenotypes, missing parental information, use of different genetic models, etc. The FBAT statistic is based on a linear combination of offspring genotypes and traits:  $FBAT = (S - E[S])/\sqrt{V}$ ,  $S = \sum_{ij} T_{ij} * X_{ij}$ , where  $V = \text{Var}(S)$  and  $T_{ij}$  represents the coded phenotype (*i.e.*, the phenotype adjusted for any covariates) of the  $j$ th offspring in family  $i$ .  $X_{ij}$  denotes the offspring's coded genotype at the locus being tested. It depends on the genetic model under consideration. The FBAT statistic has an approximate standard normal distribution; the null hypothesis being tested is that there is no linkage and no association. Extension of this equation to multiple traits is straightforward<sup>34</sup>, in which case the distribution of the multivariate FBAT statistic can be approximated by a  $\chi^2$  distribution (with the degrees of freedom equal to the rank of  $V$ ). For studies with quantitative traits that are measured repeatedly, generalized principal component analysis can be used to derive an overall phenotype that maximizes the proportion of phenotypic variance explained by the marker. The newly defined trait is used in a univariate FBAT statistic, called FBAT-PC<sup>24</sup>.

The screening strategy consists of two steps. The first step is a data reduction technique to select the most promising trait-marker combinations, which involves repeating four components of an algorithm. First, a plausible linear regression model that functionally relates the phenotype of interest to genotypic information is specified. The coding of the marker genotypes is a reflection of the underlying disease model. In linking traits to coded genotype, different selections of covariates with the test locus can be considered as well. Here we use linear regression model to link an offspring's phenotype  $Y_{ij}$  to its genotype  $X_{ij}$  and a covariate vector  $Z_{ij}$ :  $E(Y_{ij}) = aX_{ij} + bZ_{ij}$ . Second, the observed offspring genotypes are replaced by their conditional means (*i.e.*, the between-family component). When the parental genotypes are observed,

the conditional mean is computed based on the genotypes (i.e.,  $X_{ij}$  is replaced by  $E[X_{ij} | \text{parental genotypes}]$ ). When parental genotypes are incomplete, the conditional mean is computed based on the sufficient statistic<sup>35</sup> (i.e.,  $X_{ij}$  is replaced by  $E[X_{ij} | \text{sufficient statistic}]$ ). Hence, instead of the linear regression model given above, we use the conditional mean model  $E(Y_{ij}) = aE[X_{ij} | \text{parental genotypes}] + bZ_{ij}$ . For double homozygous parents,  $X_{ij} = E[X_{ij} | \text{parental genotypes}]$ . Third, the genetic effect size  $a$  in the conditional mean model is estimated.

Fourth, one of the two methods is used to evaluate each combination of trait, marker, covariates and genetic model. For method I, the conditional power of the FBAT statistic is computed given the observed data<sup>8</sup>. The power of the test statistic is computed, conditional on the offspring's phenotypes and the parental genotypes (or the sufficient statistics when parental genotypes are missing). The conditional power depends on the estimated genetic effect from the conditional mean model. For method II, the Wald test statistic for the genetic effect in the conditional mean model<sup>9</sup> is calculated:

$$\frac{\hat{a}^2}{\text{Var}(\hat{a})} = \chi^2(1).$$

Trait-marker combinations can be retained in a variety of ways, according to different criteria. For example, the criterion to retain combinations may be a specific cutoff value for the conditional power of the FBAT statistic (e.g., 80%), or it may be based on a preset number of smallest p-values for the Wald test (e.g., 5 smallest).

The second step in the screening process involves applying the FBAT statistic to the selected combinations of phenotypes and markers. Although population admixture and stratification may bias the estimate of  $a$  and will thus affect the power of the proposed testing strategy, the second step of the screening technique avoids confounding due to model mis-specification as well as admixture or population stratification: The final decision on potential marker associations is based on the FBAT test statistic, which guards against these confounding factors. The null hypothesis being tested is that there is no association and no linkage between any SNP and a DSL.

From a theoretical point of view, screening on the basis of conditional power calculations is preferred, because conditional power calculations are a natural yardstick using both the genetic effect size estimates and the number of informative families in the FBAT statistic, whereas screening on the basis of overall Wald test does not account for the available number of informative families.

**URL.** PBAT software is available for the Windows XP, Linux and UNIX operating systems<sup>11</sup> at <http://www.biostat.harvard.edu/~clange/default.htm>.

*Note: Supplementary information is available on the Nature Genetics website.*

#### ACKNOWLEDGMENTS

We thank the families with CAMP for their participation in the CAMP Genetics Ancillary Study, supported by the US National Heart, Lung and Blood Institute; the CAMP investigators and research team, supported by the US National Heart, Lung and Blood Institute, for collection of CAMP Genetic Ancillary Study data; N.A. Beattie and J.T. Follweiler for editorial help; and D. Cutler for comments and suggestions. Additional support for this research came from the US National Heart Lung and Blood Institute. K.V.S. was supported by a grant from the US National Institutes of Health. All work undertaken in the CAMP Genetics Ancillary Study was done at the Channing Laboratory at the Brigham and Women's Hospital under appropriate CAMP policies and human subject protections. M.B.M. was supported by the National Research Service Award, Training Program in Psychiatric Epidemiology and Biostatistics.

#### COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 29 November 2004; accepted 25 April 2005

Published online at <http://www.nature.com/naturegenetics/>

1. Sherry, S.T., Ward, M. & Sirotkin, K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res.* **9**, 677–679 (1999).

2. International SNP Map Working Group. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
3. Marnellos, G. High-throughput SNP analysis for genetic association studies. *Curr. Opin. Drug Discov. Devel.* **6**, 317–321 (2003).
4. Bonferroni, C.E. Teoria statistica delle classi e calcolo delle probabilita. In *Volume in Onore di Riccardo dalla Volta, Universita di Firenze*, 1–62 (1937).
5. Hochberg, Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **75**, 800–802 (1988).
6. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
7. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
8. Lange, C., DeMeo, D.L., Silverman, E., Weiss, S. & Laird, N.M. Using the noninformative families in family-based association tests: a powerful new testing strategy. *Am. J. Hum. Genet.* **73**, 801–811 (2003).
9. Lange, C. *et al.* A new powerful non-parametric two-stage approach for testing multiple phenotypes in family-based association studies. *Hum. Hered.* **56**, 10–17 (2003).
10. Lange, C., DeMeo, D.L., Silverman, E.K., Weiss, S.T. & Laird, N.M. PBAT: Tools for family-based association studies. *Am. J. Hum. Genet.* **74**, 367–369 (2004).
11. Van Steen, K. & Lange, C. PBAT: a comprehensive software package for genome-wide association analysis of complex family-based studies. *Hum. Genomics* **2**, 70–74 (2005).
12. Childhood Asthma Management Program Research Group. The childhood asthma management program (CAMP): design, rationale, and methods. *Control. Clin. Trials* **20**, 91–120 (1999).
13. Lyon, H. *et al.* IL10 gene polymorphisms are associated with asthma phenotypes in children. *Genet. Epidemiol.* **26**, 155–165 (2004).
14. Lange, C. & Laird, N.M. 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.* **23**, 165–180 (2002).
15. Laird, N., Horvath, S. & Xu, X. Implementing a unified approach to family based tests of association. *Genet. Epidemiol.* **19** Suppl 1, S36–S42 (2000).
16. Kennedy, G.C. *et al.* Large-scale genotyping of complex DNA. *Nat. Biotechnol.* **21**, 1233–1237 (2003).
17. Schaid, D.J. *et al.* Comparison of microsatellites versus single nucleotide polymorphisms by a genome linkage screen for prostate cancer susceptibility loci. *Am. J. Hum. Genet.* **75**, 948–965 (2004).
18. Lange, C. & Laird, N.M. Power calculations for a general class of family-based association tests: Dichotomous traits. *Am. J. Hum. Genet.* **71**, 575–584 (2002).
19. Lange, C., DeMeo, D.L. & Laird, N.M. Power calculations for a general class of family-based association tests: Quantitative traits. *Am. J. Hum. Genet.* **71**, 1330–1341 (2002).
20. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
21. Duffy, D.L., Martin, N.G., Battistutta, D., Hopper, J.L. & Mathews, J.D. Genetics of asthma and hay fever in Australian twins. *Am. Rev. Respir. Dis.* **142**, 1351–1358 (1990).
22. Weiss, S.T. & Raby, B.A. Asthma genetics 2003. *Hum. Mol. Genet.* **13**, R83–R89 (2004).
23. Randolph, A.G. *et al.* The IL12B gene is associated with asthma. *Am. J. Hum. Genet.* **75**, 709–715 (2004).
24. Lange, C. *et al.* A family-based association test for repeatedly measured quantitative traits adjusting for unknown environmental and/or polygenic effects. *Statistical Applications in Genetics and Molecular Biology [online]* **3**, Article 17 (2004).
25. Holm, S. A simple sequentially rejective multiple testing procedure. *Scand. J. Stat.* **6**, 65–70 (1979).
26. Sidak, Z. On probabilities of rectangles in multivariate Student distributions: their dependence on correlations. *Ann. Math. Statist.* **42**, 169–175 (1971).
27. Satagopan, J.M., Venkatraman, E.S. & Begg, C.B. Two-stage designs for gene-disease association studies with sample size constraints. *Biometrics* **60**, 589–597 (2004).
28. International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
29. Zhai, W., Todd, M.J. & Nielsen, R. Is haplotype block identification useful for association mapping studies? *Genet. Epidemiol.* **27**, 80–83 (2004).
30. Churchill, G.A. & Doerge, R.W. Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971 (1994).
31. Churchill, G.A. & Doerge, R.W. Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**, 285–294 (1996).
32. Lin, S., Chakravarti, A. & Cutler, D.J. Exhaustive allelic transmission disequilibrium tests as a new approach to genome-wide association studies. *Nat. Genet.* **36**, 1181–1188 (2004).
33. Spielman, R.S., McGinnis, R.E. & Ewens, W.J. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52**, 506–516 (1993).
34. Lange, C., Silverman, E., Weiss, S.T., Xu, X. & Laird, N.M. A multivariate family-based test using generalized estimating equations: FBAT-GEE. *Biostatistics* **4**, 195–206 (2003).
35. Rabinowitz, D. & Laird, N.M. A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Hum. Hered.* **50**, 211–223 (2000).