Genetic Epidemiology 32: 255–263 (2008)

Increasing the Power of Identifying Gene × Gene Interactions in Genome-wide Association Studies

Charles Kooperberg* **and Michael LeBlanc**

Fred Hutchinson Cancer Research Center, Division of Public Health Sciences, Seattle, Washington

In this paper we investigate the power to identify gene × gene interactions in genome-wide association studies. In our analysis we focus on two-stage analyses: analyses in which we only test for interactions between single nucleotide polymorphisms that show some marginal effect. We give two algorithms to compute significance levels for such an analyses. One involves a Bonferoni correction on the number of interactions that are actually tested, and one is a resampling procedure similar to the one proposed by [Lin (2006) Am. J. Hum. Genet. 78:505–509]. We also give an algorithm to carry out approximate power calculations for studies that plan to use a two-stage analysis. We find that for most plausible interaction effects a two-stage analysis can dramatically increase the power to identify interactions compared to a single-stage analysis based on simulation studies using known genetic models and data from existing genome-wide association studies. *Genet. Epidemiol.* 32:255–263, 2008. © 2008 Wiley-Liss, Inc.

Key words: epistasis; two-stage analyses

Contract grant sponsor: NIH; Contract grant numbers: CA 74841; CA 53996; HL 74745; CA 125489.
*Correspondence to: Charles Kooperberg, Fred Hutchinson Cancer Research Center, Division of Public Health Sciences, 1100 Fairview Ave N/M3-A410, Seattle, WA 98109-1024. E-mail: clk@fhcrc.org
Received 17 September 2007; Revised 28 November 2007; Accepted 1 December 2007
Published online 16 January 2008 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/gepi.20300

INTRODUCTION

Genome-wide association studies (GWAs), which genotype hundreds of thousands of Single nucleotide polymorphisms (SNPs) on thousands of participants are, fueled by decreasing prices of genotyping, now carried out. Some initial successes have been reported [e.g. Tomlinson et al., 2007; Scott et al., 2007; Easton et al., 2007; WTCCC 2007]. Although the primary interest in GWAs is typically the identification of SNPs that are marginally associated with a disease, it is typically also of interest to identify SNPs that jointly have an epistatic (interaction) effect on the disease of interest. Such interactions may shed light on potential disease-associated pathways, and they may identify people who are at extreme high risks [e.g. Manolio and Collins, 2007].

It is easy to see that the potential number of interaction to be tested is enormous. When 500,000 SNPs are genotyped, there are $\binom{500,000}{2} \approx 10^{11}$ two SNP combinations and $\binom{500,000}{3} \approx 10^{16}$ three SNP combinations. This creates both computational and multiple comparisons problems: it is virtually impossible to evaluate each possible model, and, as a multiple comparisons correction needs to be made for many possible tests, there is limited power for the identification of any of the interactions that are tested.

In this paper we investigate strategies to improve the power in genome-wide association studies, while reducing the computational expense considerably. Our main tool is a two-stage analysis: we only investigate interactions between SNPs that show some (modest) marginal effect. Initially two-stage strategies were proposed as study designs to reduce the expense of a GWA, without a substantial reduction in the power to identify the (marginal) effect of SNPs on a disease [e.g. Lin, 2006; Skol et al., 2006, 2007; Thomas et al., 2004]. More recently, there have been proposals to use two-stage analysis strategies to possibly improve the power of identifying interaction effects in GWAs [Evans et al., 2006; Macgregor and Kahn, 2006]. A thorough discussion on how power and type 1 error are influenced by such a two-stage design is missing, however.

Marchini et al. [2005] and Evans et al. [2006] investigated whether a two-stage analysis was a viable approach to improve the power to identify SNPs that are marginally associated with a disease. They found, that, possibly because they carried out a multiple comparisons correction for all possible associations, their two-stage analysis did not improve the power for identifying SNPs that jointly have an epistatic effect in GWAs. One of the main goals of this paper is to better correct for multiple comparisons in two-stage analyses. Macgregor and Khan [2006] argued that one only needs to correct the number of interactions tested for. Their paper, however, gives neither a justification nor a simulation study validating this assertion. In this paper we

attempt to give a more solid foundation for two-stage analyses and provide extensive simulation studies to back up our results. We also provide an analytic algorithm to approximate power for the detection of interactions using a two-stage analysis.

The goal in this paper is to develop methods to identify interacting SNPs in (genome-wide) association studies. Several authors [Chapman and Clayton, 2007; Chatterjee et al., 2006; Marchini et al., 2005] recently have proposed methods that make use of interactions in developing powerful tests to determine whether SNPs are marginally associated with a disease outcome, which is a slightly different objective from ours. For example, in the situation where there is in fact no interaction, we would hope that these procedures would still identify SNPs that are associated with the disease, whereas our procedure would not be expected to identify such combinations. On the other hand, when there is an interaction, the bar is higher for the procedure that we proposing, as we want to identify the SNPs in combination.

METHODS

AN ELEMENTARY INDEPENDENCE RESULT

Let Y_i, $i = 1, \ldots, n$ be independent identically distributed (iid) random variables, and let x_{1i} and x_{2i}, $i = 1, \ldots, n$ be predictor variables. Consider the three linear regression models

$$Y_i = \gamma_{10} + \gamma_{11} x_{1i} + \varepsilon_i^1 \tag{1}$$

$$Y_i = \gamma_{20} + \gamma_{21} x_{2i} + \varepsilon_i^2 \tag{2}$$

and

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \varepsilon_i. \tag{3}$$

Let $\widehat{\gamma}_{10}$, $\widehat{\gamma}_{11}$, $\widehat{\gamma}_{20}$, $\widehat{\gamma}_{21}$, $\widehat{\beta}_0$, $\widehat{\beta}_1$, $\widehat{\beta}_2$, and $\widehat{\beta}_3$ be the ordinary least squares estimates of the parameters in (1), (2), and (3). Then $\widehat{\beta}_3 \perp \widehat{\gamma}_{11}$ and $\widehat{\beta}_3 \perp \widehat{\gamma}_{21}$.

Proof: Note that ordinary least square estimates are linear in the response. Let $\widehat{\gamma}_{j1} = \sum_i a_{ji} Y_i$, and $\widehat{\beta}_1 = \sum_i b_i Y_i$. Expressions for the a_i and b_i can be found in any elementary linear regression text. Because the Y_i are iid, $\mathrm{cov}(\widehat{\gamma}_{j1}, \widehat{\beta}_1) = \sum_i a_{ji} b_i \mathrm{var}(Y_i) = \mathrm{var}(Y_1) \sum_i a_{ji} b_i$. Algebra yields that for models (1) and (3) $\sum_i a_{ji} b_i = 0$; thus, $\mathrm{cov}(\widehat{\gamma}_{j1}, \widehat{\beta}_3) = 0$.

The implication of this result is that for a cohort study with a continuous outcome we can carry out a hypothesis test for interactions in a two-stage approach:

- test only interactions between those predictors (SNPs) that are marginally significant at a prespecified level α_1; and
- control the global (family-wise) type 1 error by controlling for the number of interactions that are actually being tested for (for example, in a Bonferoni fashion), rather than the one that could have been tested for.

CASE-CONTROL SAMPLING

Most genome-wide association studies employ a case-control study design. Unlike for cohort studies with a continuous outcome the samples are not iid, and it is likely that the study will be analyzed using logistic, rather than linear regression. In particular, we would now estimate using the models

$$\mathrm{logit}(P(Y_i = 1 | x_{1i}, x_{2i})) = \gamma_{10} + \gamma_{11} x_{1i} \tag{4}$$

$$\mathrm{logit}(P(Y_i = 1 | x_{1i}, x_{2i})) = \gamma_{20} + \gamma_{21} x_{2i} \tag{5}$$

and

$$\mathrm{logit}(P(Y_i = 1 | x_{1i}, x_{2i})) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i}. \tag{6}$$

We believe that in general, in models (4)–(6) the maximum likelihood estimate $\widehat{\beta}_3$ is not independent of the estimates $\widehat{\gamma}_{11}$ and $\widehat{\gamma}_{21}$. However, in simulation studies, including those reported below, we have found that this dependence is sufficiently small that an inference using a Bonferoni approach on the number of tested interactions is still valid. Alternatively, inference can be carried out using a permutation approach described below.

SCORE STATISTICS BASED "PERMUTATION" TESTS

A common approach to controlling the type 1 error in situations where the exact distribution of the test statistic is not known is a permutation test. However, permutation tests for interactions are generally not possible, as permutations do not just remove the interaction effect, but they also remove the main effect [Pesarin, 2001]. Inference on an interaction should be conditional on the main effect, but in fact a straightforward permutation test tests for a combination of main effects and interactions. For linear models, for a given permutation, it is possible to permute the residuals or the fitted interaction component (similar to the parametric bootstrap [Efron and Tibshirani, 1993]). For logistic regression, the usual regression model employed for case-control studies, such an approach does not work.

Lin [2006] proposes an approach to obtaining test statistics in two-stage genome-wide association studies that can be adapted to the testing for interaction effects. In our situation the null-hypothesis to be tested is that there is no interaction, although there may be main effects. First, consider the logistic regression model

$$\begin{aligned} &\mathrm{logit}(P(Y_i = 1 | X_{ji}, X_{ki})) \\ &\quad = \beta_0 + \beta_1 X_{ji} + \beta_2 X_{ki} + \beta_3 X_{ji} X_{ki}. \end{aligned} \tag{7}$$

Set $Z_{ijk} = X_{ij}X_{ik}$. Then the efficient score for β_3 is

$$U_{jk} = \sum_{i=1}^{n} U_{ijk} = \sum_{i=1}^{n} (Y_i - p_{ijk})(X_{ij}X_{ik} - \mu_{ijk}),$$

where p_{ijk} is the fitted probability for the ith subject from the logistic regression model that includes X_{ij} and X_{ik}, but not Z_{ijk}, and μ_{ijk} is the fitted value for the ith subject from the linear regression model $Z_{ijk} = \rho_0 + \rho_1 X_{ij} + \rho_2 X_{ik}$. Under the null-hypothesis of no interaction effect, U_{jk} has approximately a normal distribution with mean 0 and variance $V_{jk} = \sum_{i=1}^{n} U_{ijk}^2$. Set $T_{jk} = U_{jk}^2/V_{jk}$. Let $G_1, \ldots, G_n$ be independent normal random variables. Set $\widetilde{U}_{jk} = \sum_i U_{ijk}G_i$. Then the $\widetilde{U}_{jk}$ have mean 0 and approximately the same covariance matrix as the U_{jk}. Set $\widetilde{T}_{jk} = \widetilde{U}_{jk}^2/V_{jk}$. Thus, a strategy to get a sample from the T_{jk} under the null-hypothesis of no association is to generate repeated samples of $G_1, \ldots, G_n$ and compute the $\widetilde{T}_{jk}$. Note that this approach of controlling the type 1 error does not assume independence of the stage one and two tests, as the Bonferoni approach does, but rather the "permutations" for stage two are carried out conditional on the results of stage one, as the permutations are only carried out for the selected SNPs (j and k) and the p_{ijk} depend on the marginal models.

A STRATEGY TO IDENTIFY INTERACTIONS IN GWAS

Our strategy to test for the presence of an interaction is as follows:

1. Fix the first-stage significance level α_1.
2. Test every SNP marginally at level α_1 using a logistic regression model. Say that n^* SNPs were significant at level α_1.
3. Test the interactions between all SNPs that pass Step 2, using the logistic model (7).
4. *Bonferoni*: Correct the P-value from Step 3 using a Bonferoni correction for $\binom{n^*}{2}$ tests; or
5. *Score*: Compute T_{jk} for each of the tests from Step 3. Let $T^* = \max_{1 \le j,k \le n^*} T_{jk}$. Also generate L independent sets of $\widetilde{T}_{jk}^l$, $l = 1, \ldots, L$, and set $\widetilde{T}^{*l} = \max_{1 \le j,k \le n^*} \widetilde{T}_{jk}^l$. Compare T^* to the $\widetilde{T}^{*l}$.

Generalizations of this procedure that use the false discovery rate, or that can identify more than one interactions are straightforward.

APPROXIMATIONS FOR POWER CALCULATIONS

When the data are generated from model (7) the logistic regression estimate of β_3 has approximately a normal distribution with mean β_3 and variance

$$\sum_{l_1 l_2 l_3 \in \{0,1\}} \frac{1}{n_{l_1 l_2 l_3}},$$

where $n_{l_1 l_2 l_3}$ is the number of observations for which $Y_i = l_1$, $X_{ij} = l_2$, and $X_{ik} = l_3$. This can be used to compute a useful approximation to the power of identifying an interaction, under the assumption that all SNPs and all interactions between SNPs are independent. Details are given in Appendix A.

RESULTS

SIMULATION SETUP

We generate 10,000 binary SNPs $X_1, \ldots, X_{10,000}$ as a first-order Markov chain such that $\mathrm{cor}(X_i, X_{i+1}) = \rho$. We generate a response Y according to

$$\mathrm{logit}(P(Y = 1|X)) = \beta_0 + \beta_1 X_{2,500} + \beta_2 X_{7,500} + \beta_3 X_{2,500} X_{7,500}.$$

The minor allele frequency $P(X_i = 1) = p$ is taken constant for all SNPs i. Note that a minor allele frequency p for this binary SNP model corresponds to a minor allele frequency of $q = 1 - \sqrt{1-p}$ for a dominant genetic model with bi-allelic SNPs.

We generate data until we have c cases ($Y = 1$) and controls ($Y = 0$). For computational reasons we take $c = 500$ and relatively large effect sizes. For the power analysis approximations below we consider larger sample sizes and smaller effect sizes. In all our simulations we took $\beta_0 = -2$, generating outcomes for a moderately rare outcome, and $\beta_1 = \beta_2$, identical main effects for both SNPs.

In our analysis we consider selection on the marginal effects using $\alpha_1 = 0.002$, 0.005, 0.01, 0.02, and 1.0. We compute the power when significant SNP × SNP interactions are identified using the modified approach of Lin [2006] described above with 1,000 permutations, using a Bonferoni correction for the number of interactions that are actually tested, and the analytic approximation. Note that for the analytic approximation the correlation between SNPs is ignored. For $\alpha_1 = 1$ we do not compute the power using the approach of Lin [2006]; as computing 1,000 permutations for $\binom{10000}{2} \approx 5 \times 10^7$ interactions for a single simulation is not well feasible. We control the overall type 1 error at a global (family-wise) level $\alpha = 0.05$ using the Bonferoni method.

SIMULATION RESULTS

In Figure 1 we show the results for $\beta_3 = 0$, $p = 0.3$ ($q = 0.173$), $\rho = 0.7$, $\beta_2 = \beta_1$ various values of α_1 as a function of the main effect $\beta_1 = \beta_2$. In this situation there are no interactions, and we would want any procedure to yield the designed type 1 error. On the basis of 1,000 simulations (so that the standard error

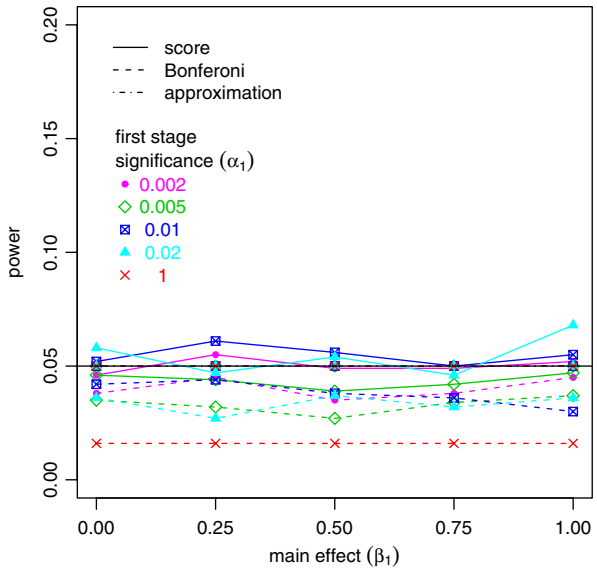

Fig. 1. Type 1 error for identifying an interaction when there is in fact no interaction effect, but there is a dominant main effect for two single nucleotide polymorphisms each with minor allele frequency 0.173 for various levels of two-stage testing ($\alpha_1 < 1$) and one-stage testing ($\alpha_1 = 1$). [Color figure can be viewed in the online edition which is available at www.interscience.wiley.com]

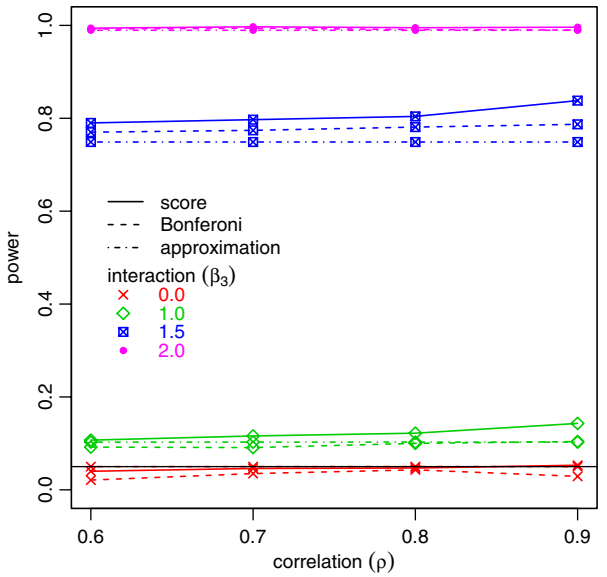

Fig. 2. Power for identifying an interaction when there is no main effect, for various interaction effects between two single nucleotide polymorphisms each with minor allele frequency 0.173 for two-stage testing at $\alpha_1 = 0.05$ for various correlations between the SNPs. [Color figure can be viewed in the online edition which is available at www.interscience.wiley.com]

of the power estimates is about 0.007), it appears that the type 1 error is well controlled by both the Bonferoni and the Lin approach, and that the type 1 error is not inflated by using a two-stage testing procedure for interactions. The power approximation yields exactly a type 1 error of 0.05 in this situation. The Bonferoni correction on all interactions ($\alpha_1 = 1$) appears somewhat conservative, and the Lin approach is well within the range what would be expected based on the standard error. We repeated these simulations for various other parameter settings and consistently obtained the same results.

In the remaining simulations we include interactions, and thus prefer approaches in which the power to identify interactions is large. In particular, in Figure 2 we show the results for $\beta_1 = \beta_2 = 0$, $\beta_3 = 0$, 1, 1.5, and 2, $\alpha_1 = 0.005$, and $p = 0.3$ ($q = 0.173$) as a function of the correlation ρ between SNPs. We note that all three methods to compute the power provide similar results with the power computed using the Lin [2006] approach to suggest slightly larger power. We also note that the power is slightly larger when the correlation is 0.9 than when the correlation is smaller but that the difference is very small. Because of this small effect of the correlation we show further results only for $\rho = 0.7$.

In Figures 3 and 4 we show the results for $\beta_1 = \beta_2 = 0$, $\rho = 0.7$, $p = 0.3$ ($q = 0.173$; Fig. 3), and $p = 0.4$ ($q = 0.235$; Fig. 4) for a variety of choices of α_1 as a function of the interaction effect β_3. We note that

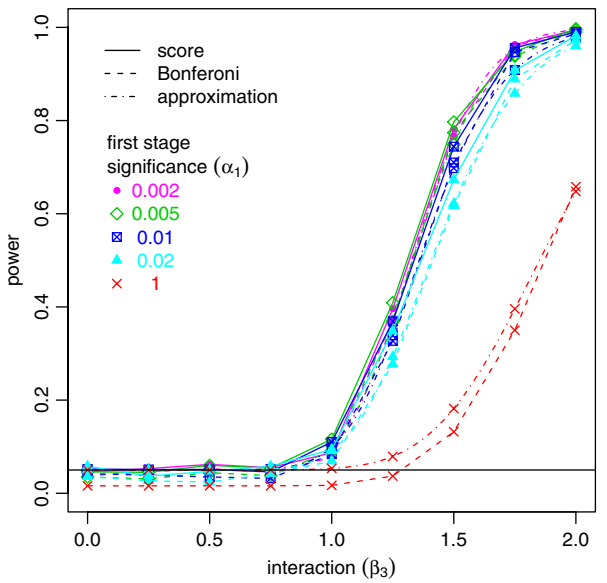

Fig. 3. Power for identifying an interaction when there is no main effect, for various interaction effects between two single nucleotide polymorphisms each with minor allele frequency 0.173 for various levels of two-stage testing ($\alpha_1 < 1$) and one-stage testing ($\alpha_1 = 1$). [Color figure can be viewed in the online edition which is available at www.interscience.wiley.com]

the power to identify an interaction is substantially larger when we first filter on main effects, even though there is in fact no main effect when $\beta_1 = \beta_2 = 0$. The power gain of filtering in the first stage over a global Bonferoni correction can be as

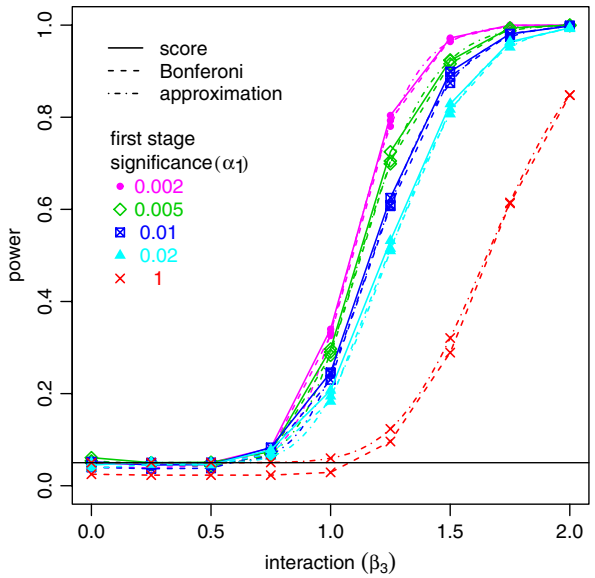

Fig. 4. Power for identifying an interaction when there is no main effect, for various interaction effects between two single nucleotide polymorphisms each with minor allele frequency 0.235 for various levels of two-stage testing ($\alpha_1 < 1$) and one-stage testing ($\alpha_1 = 1$). [Color figure can be viewed in the online edition which is available at www.interscience.wiley.com]

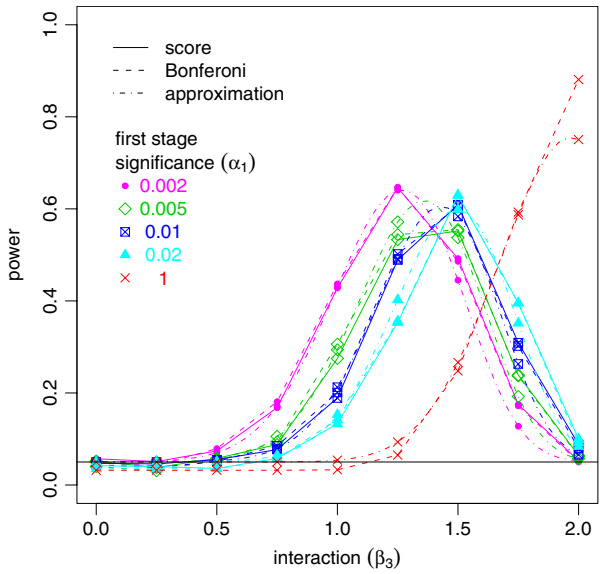

Fig. 5. Power for identifying an interaction when the main effect goes in the opposite direction of the interaction effect, for various interaction effects between two single nucleotide polymorphisms each with minor allele frequency 0.235 for various levels of two-stage testing ($\alpha_1 < 1$) and one-stage testing ($\alpha_1 = 1$). [Color figure can be viewed in the online edition which is available at www.interscience.wiley.com]

much as 40%. These power gains are universal and were observed in many other situations. In Figure 4, where the minor allele frequency was larger it appears that most power is gained when taking the first stage α_1 small. The difference in power between the various choices of α_1 is small compared to the difference between two-stage selection and overall testing with $\alpha_1 = 1$.

The only situation that we identified where a two-stage selection does not help is when the main effect goes in the opposite direction of the interaction effect. In Figure 5 we show the results for $\beta_1 = \beta_2 = -1$, $\rho = 0.7$, and $p = 0.4$ ($q = 0.235$). For some values of β_3, testing all interactions now yields more power than two-stage testing. This corresponds to the situation where the magnitude of the interaction leads to approximately canceling out the marginal association of the two genes to the outcome. Therefore, the non-monotonic shape of the power curves can be explained by low power to detect the interaction at the second stage for β_3 small, increasing as the interaction effect increases, but decreasing as $\beta_3 \to 2$, as the chance that the relevant genes are selected at the first stage is dramatically reduced as the marginal association is weakened.

Clearly some more unusual interaction patterns similar to those in Marchini et al. [2005] may also yield more power for overall testing. The main question that a researcher will have to ask before testing is what she/he thinks to be more likely: if an epistatic effect where one SNP enhances the effect of another gene is more likely than a more complicated interaction effect, a two-stage testing procedure should be employed.

SIMULATIONS USING REAL GWA DATA

From Illumina iControlDB (http://www.illumina.com) we retrieved 610 arrays of human array 317 K data. The Illumina iControlDB contains user submitted data of Illumina array data that were submitted to be used as "controls in case-control association studies, in which risk factors of individuals with a certain disease (cases) are compared to individuals without the disease (controls)", as well as methodological studies like the current one. From this data we removed SNPs with a minor allele frequency of under 10% and selected the remaining 10,321 SNPs on chromosome 13 for further simulations. We phased these 610 arrays using fastPhase [Scheet and Stephens, 2006] to obtain 1,220 haploid copies of chromosome 13. We then generated a huge population by randomly pairing two of these haploid copies and generated disease status using the model

$$\begin{aligned}&\text{logit}(P(Y = 1|\text{SNPs}))\\ &\quad = \beta_0 + \beta_1 U_1 + \beta_2 U_2 + \beta_3 U_1 U_2,\end{aligned}$$

where $U_1 = 1$ is at least one of the haploids for rs1751871 was a copy of the minor allele and $U_2 = 1$ is at least one of the haploids for rs9523716 was a copy of the minor allele. These two SNPs are not in

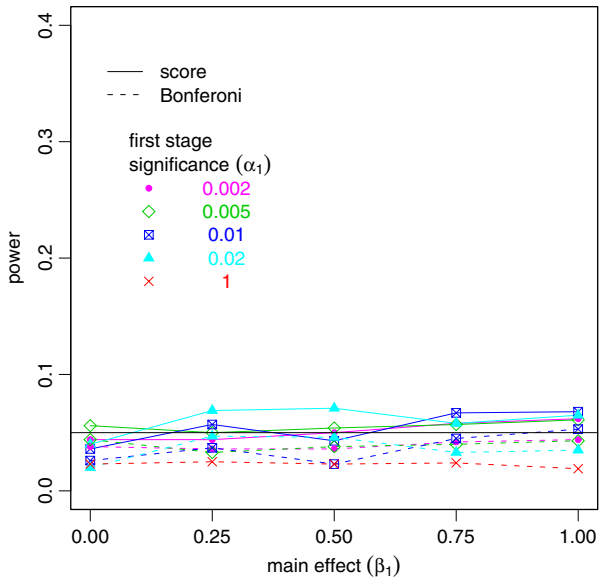

Fig. 6. Type 1 error for identifying an interaction when there is in fact no interaction effect, but there is a dominant main effect for two single nucleotide polymorphisms in data generated from Illumina HapMap 317 K arrays, for various levels of two-stage testing ($\alpha_1 < 1$) and one-stage testing ($\alpha_1 = 1$). [Color figure can be viewed in the online edition which is available at www.interscience.wiley.com]

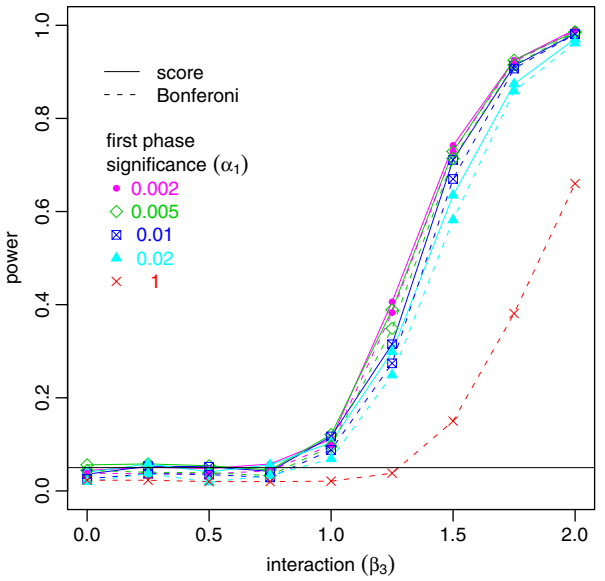

Fig. 7. Power for identifying an interaction when there is no main effect, for various interaction effects between two single nucleotide polymorphisms for two single nucleotide polymorphisms in data generated from Illumina HapMap 317 K arrays for various levels of two-stage testing ($\alpha_1 < 1$) and one-stage testing ($\alpha_1 = 1$). [Color figure can be viewed in the online edition which is available at www.interscience.wiley.com]

LD ($r^2 = 0.01$); the minor allele frequencies are 0.13 (rs1751871) and 0.25 (rs9523716).

In Figure 6 we show the results for 1,000 simulations with $\beta_0 = -2$, $\beta_3 = 0$, and $\beta_2 = \beta_1$, for various values of α_1. The results in this figure suggest that for the data from actual GWAs, just as for the simulated data, the type 1 error is well controlled. In Figure 7 we show the results for 1,000 simulations with $\beta_0 = -2$, $\beta_1 = \beta_2 = 0$, for various values of β_3 and α_1. The results in this figure suggest that the two-stage procedure indeed improves the power to identify the interactions dramatically compared to a one-stage procedure, and that the score approach yields almost the same results as the (cheaper) Bonferoni correction.

APPROXIMATE POWER CALCULATIONS

Using the algorithm in Appendix A we can approximate the power of identifying interactions in genome-wide association studies. For example, in Table I we give the power for identifying the specific interaction, associated with the parameter β_3 in model (7) where both involved SNPs have at least one variant allele for 40% of the samples (corresponding to a minor allele frequency of 0.225) or 20% of the samples (corresponding to a minor allele frequency of 0.106) for a case-control study with 2,000 cases and 2,000 controls, measuring 500,000 SNPs, for a variety of two-stage procedures, allowing three false positives under independence.

TABLE I. Power of identifying a specific interaction, associated with the parameter β_3 in model (7) where both involved SNPs have at least one variant allele for 40 or 20% of the samples (corresponding to a minor allele frequency of 0.225 and 0.106, respectively) for a case-control study with 2,000 cases and 2,000 controls, measuring 500,000 SNPs, for a variety of two-stage procedures, allowing three false positives

		First stage significance level α_1							
β_3	Odds ratio	0.0001	0.001	0.002	0.005	0.01	0.02	0.05	1
$P(\text{SNP} > 0) = 0.4 \Rightarrow$ minor allele frequency 0.225									
0.5	1.65	0.09	0.11	0.10	0.08	0.06	0.04	0.02	0.00
0.6	1.82	0.37	0.44	0.40	0.32	0.26	0.19	0.12	0.02
0.7	2.01	0.75	0.80	0.75	0.66	0.56	0.47	0.35	0.10
0.8	2.23	0.95	0.96	0.94	0.88	0.82	0.75	0.64	0.31
0.9	2.46	0.99	0.99	0.99	0.97	0.95	0.92	0.87	0.60
1.0	2.72	1.00	1.00	1.00	1.00	0.99	0.98	0.97	0.84
$P(\text{SNP} > 0) = 0.2 \Rightarrow$ minor allele frequency 0.106									
1.0	2.72	0.09	0.24	0.29	0.34	0.35	0.34	0.29	0.09
1.1	3.00	0.21	0.45	0.52	0.58	0.59	0.58	0.51	0.21
1.2	3.32	0.39	0.66	0.73	0.79	0.80	0.78	0.72	0.38
1.3	3.67	0.60	0.83	0.88	0.91	0.92	0.91	0.87	0.58
1.4	4.06	0.78	0.93	0.95	0.97	0.98	0.97	0.95	0.76

SNP, single nucleotide polymorphism.

We note from this table that a two-stage procedure can considerably increase the power of identifying an interaction. In particular, the optimal fraction of SNPs to consider for testing for interactions appears to be about 0.001 for the higher minor allele

TABLE II. Power of identifying a specific interaction, associated with the parameter β_3 in model (7) where both involved SNPs have at least one variant allele for 40 or 20% of the samples (corresponding to a minor allele frequency of 0.225 and 0.106, respectively) for a case-control study with 5,000 cases and 5,000 controls, measuring 500,000 SNPs, for a variety of two-stage procedures, allowing three false positives

		First stage significance level α_1							
β_3	Odds ratio	0.0001	0.001	0.002	0.005	0.01	0.02	0.05	1
$P(\text{SNP}>0)=0.4 \Rightarrow$ minor allele frequency 0.225									
0.4	1.49	0.41	0.52	0.49	0.41	0.34	0.27	0.18	0.03
0.5	1.65	0.89	0.94	0.92	0.87	0.81	0.74	0.62	0.27
0.6	1.82	0.99	1.00	1.00	0.99	0.98	0.97	0.94	0.72
0.7	2.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96
$P(\text{SNP}>0)=0.2 \Rightarrow$ minor allele frequency 0.106									
0.6	1.82	0.02	0.09	0.12	0.15	0.17	0.18	0.16	0.05
0.7	2.01	0.11	0.30	0.38	0.46	0.50	0.51	0.49	0.21
0.8	2.23	0.32	0.60	0.68	0.77	0.82	0.83	0.81	0.52
0.9	2.46	0.61	0.84	0.89	0.94	0.96	0.97	0.96	0.82
1.0	2.72	0.85	0.96	0.97	0.99	0.99	1.00	1.00	0.96

SNP, single nucleotide polymorphism.

frequency and 0.01 for the lower minor allele frequency in this example. If the sample size is increased to 5,000 pairs of cases and controls, there is even some power to identify interactions associated with odds ratios of 1.5 (Table II). It appears that the minor allele frequency is the dominant factor in determining what the "optimal" value of α_1 is, whereas in particular the odds ratio does not have much influence. On the basis of a more extensive set of simulations, we feel that values in the order of $\alpha_1 \sim 0.005$ are usually fairly close to the optimal value. Power calculations can be carried out in a straightforward fashion to optimize α_1 for any design and hypothesized effect size using our code, available from http://bear.fhcrc.org/~clk/soft.html.

DISCUSSION

In this paper we investigate the power for identifying interactions using a two-stage analysis. We found that the power of identifying interactions can be greatly improved using such an analysis. The significance for such an analysis can be controlled using a Bonferoni correction on the number of interactions actually tested or a resampling approach similar to the one proposed by Lin [2006]. Approximate power calculations for such an analysis can be computed explicitly.

Our results are somewhat contradictory to those of Marchini et al. [2005] and Evans et al. [2006]. In these papers a two-stage analysis is used to marginally identify SNPs that may have epistatic effects on a disease outcome. In their analysis they find that a two-stage analysis does not improve the power to identify SNPs marginally. Besides that the goal in our analysis is slightly different from that in these two papers; we believe that there are two reasons why we do reach the conclusion that a two-stage analysis is useful.

The main reason is that we focused on what we feel are *plausible* interactions, interactions where the effects are monotone in the number of minor alleles of both SNPs that are involved. We believe that it is often reasonable to make such an assumption about the type of interaction. Clearly, some assumption is necessary; without such an assumption the number of possible interactions is enormous, and the power to identify them is substantially reduced. In presentations we have sometimes compared this with a cake; if we want to divide the "power" over all possible interactions, nobody will get more than a crumb, and no-one will taste how good the cake is; we are better off dividing the cake among those people we believe to enjoy it. We believe that not all possible interactions are likely, in contrast Evans et al. [2006] consider all sorts of interactions (some of which they themselves label as "exotic"; see their Fig. 4).

The second reason is that it is critical that in a two-stage procedure we control the type 1 error for our testing strategy. Correcting for all possible tests in a Bonferoni manner in a two-stage procedure is unnecessarily conservative. In our two-stage procedure, where the first and second stage testings are (virtually) independent, this can be done with a Bonferoni correction on the number of tests actually carried out or with a permutation based score approach. This later approach could potentially be adapted to situations where the two stages are not independent.

In this paper we show that a two-stage procedure, which takes those issues into account, dramatically increases the power to identify interactions over a one-stage approach. The power to identify them is still limited, but with the increased size of some of the planned GWAs finding interactions is no longer out of the question. It remains an open question whether a two-stage analysis that take these issues into account would improve the power over a one-stage analysis for identifying SNPs that are associated with a disease.

ACKNOWLEDGMENTS

The authors thank Lon Cardon, Li Hsu, Hua Tang, and Yingye Zheng for many helpful discussions.

REFERENCES

Chapman J, Clayton D. 2007. Detecting association using epistatic information. Genet Epidemiol 31:776–788.

Chatterjee N, Kalaylioglu Z, Moslehi R, Peters U, Wacholder S. 2006. Powerful multilocus tests of genetic association in the presence of gene–gene and gene–environment interactions. Am J Hum Genet 78:1002–1016.

Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, The EARCH Collaborators, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen C-Y, Wu P-E, Wang H-C, Eccles D, Evans DG, Peto J, Fletcher O, Johnson N, Seal S, Stratton MR, Rahman N, Chenevix-Trench G, Bojeseu SE, Nordestgaard BG, Axelsson CK, Garcia-Closas M, Brinton L, Chanock S, Lissowska J, Peplonska B, Nevanlinna H, Fagerholm R, Eerola H, Kang D, Yoo K-Y, Noh D-Y, Ahn S-HG, Hunter DJ, Hankinson SE, Cox DG, Hall P, Wedren S, Liu J, Low Y-L, Bogdanova N, Schrmann P, Drk T, Tollenaar RAEM, Jacobi CE, Devilee P, Klijn JGM, Sigurdson AJ, Doody MM, Alexander BH, Zhang J, Cox A, Brock JW, MacPherson G, Reed MWR, Couch FJ, Goode EL, Olson JE, Meijers-Heijboer H, van den Ouweland A, Uitterlinden A, Rivadeneira F, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Hopper JL, McCredie M, Southey M, Giles GG, Schroen C, Justenhoven C, Brauch H, Hamann U, Ko Y-D. Spurdle AB, Beesley J, Chen X, kConFab and AOCS Management Group, Mannermaa A, Kosma V-M, Kataja V, Hartikainen J, Day NE, Cox DR, Ponder BAJ. 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447:1087–1093.

Efron B, Tibshirani RJ. 1993. An Introduction to the Bootstrap. London: Chapman & Hall.

Evans DM, Marchini J, Moris AP, Cardon LR. 2006. Two-stage two-locus models in genomewide association. PLoS Genet 2:e157.

Lin DY. 2006. Evaluating statistical significance in two-stage genomewide association studies. Am J Hum Genet 78:505–509.

Macgregor S, Khan IA. 2006. GAIA: an easy-to-use web-based application for interaction analysis of case–control data. BMC Med Genet 7:34.

Manolio TA, Collins F. 2007. Genes, environment, health, and disease: facing up to complexity. Hum Hered 63:63–66.

Marchini J, Donnelly P, Cardon LR. 2005. Genome-wide strategies for detecting multiple loci that influence complex diseases. Nat Genet 37:413–417.

Pesarin F. 2001. Multivariate Permutation Tests with Applications in Biostatistics. Chichester: Wiley.

Scheet P, Stephens M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet 78:629–644.

Scott LJ, Mohlke KL, Bonnycastle LL, Willer CL, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding C-J, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li X-Y, Conneeely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M. 2007. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316:1341–1345.

Skol AD, Scott LJ, Abecasis GR, Boehnke M. 2006. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Gen 38: 209–213.

Skol AD, Scott LJ, Abecasis GR, Boehnke M. 2007. Optimal designs for two-stage genome-wide association studies. Genet Epidemiol, in press.

The Wellcome Trust Case Control Consortium (WTCCC) 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678.

Thomas D, Xie RR, Gebregziabher M. 2004. Two-stage sampling designs for gene association studies. Genet Epidemiol 27:401–414.

Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W, Barclay E, Lubbe S, Martin L, Sellick G, Jaeger E, Hubner R, Wild R, Rowan A, Fielding S, Howarth K, the CORGI Consortium, Silver A, Atkin W, Muir K, Logan R, Kerr D, Johnstone E, Sieber O, Gray R, Thomas H, Peto J, Cazier J-B, Houlston R. 2007. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet 39:984–988.

APPENDIX A: ALGORITHM FOR A POWER CALCULATION FOR TWO-STAGE TESTING OF INTERACTIONS

Assume model (7), assume that there are n_{ca} cases and n_{co} controls.

1. Given model (7) compute $P(Y = 1|X_j = 1)$ and $P(Y = 1|X_k = 1)$.
2. Compute the power z_j (z_k) that X_j (X_k) is marginally significant in a case–control study with n_{ca} cases and n_{co} controls at the significance level α_1.
3. Compute $q(m) = P(Z = m)$, where Z has a binomial distribution with n the number of SNPs minus 2, and $p = \alpha_1$. Thus, $q(m)$ is the probability distribution of the number of other SNPs that are significant at level α_1, and that thus go on to stage 2.
4. Compute the expected variance of $\widehat{\beta}_3$, and then the probability $p(\rho)$ that $\widehat{\beta}_3$ is significant at the level ρ using a normal approximation.
5. Let $\alpha_2(m) = \alpha\binom{m+2}{2}$, the Bonferoni corrected significance level for the second stage of the analysis when the two SNPs involved with the interaction and m other SNPs are significant at level α_1.
6. We approximate the power of identifying an interaction at a global significance level α by

$$\begin{aligned} & z_k z_l \sum_m q(m) p(\alpha_2(m)) \\ & \quad + z_k z_l \sum_m q(m)(1 - p(\alpha_2(m))\alpha_2(m) \\ & \left(\binom{m+2}{2} - 1\right) + (1 - z_k z_l)\alpha. \end{aligned} \tag{8}$$

The first term in (8) is the power of identifying the $X_k X_l$ interaction (where we sum over the number of other SNPs that go to the second stage). The second term in (8) is the power of

identifying another interaction if the X_kX_l interaction is not identified, but both X_k and X_l were marginally significant at level α_1, again summing over the number of other SNPs that go to the second stage, also noting that one of the $\binom{m+2}{2}$ is X_kX_l and is already taken care of in the first term. The third term is the power of identifying an interaction if either X_k or X_l was not marginally significant at level α_1.

7. Similarly, we can approximate the power of identifying an interaction between X_k and X_l (ignoring other interactions that may be identified) by

$$z_k z_l \sum_m q(m)p(\alpha_2(m)). \tag{9}$$

The scenario where we allow for, say, F false positives can be approximated by using $\alpha_2(m) = F/\binom{m+2}{2}$ in (9), so that when $\binom{m+2}{2}$ interactions are tested F of them are significant just by chance.

The results of Equation (8) are shown in the figures; those of Equation (9) are shown in the tables.