Genetic Epidemiology 31: 358–362 (2007)

Genetic Model Testing and Statistical Power in Population-Based Association Studies of Quantitative Traits

Guillaume Lettre,[1,2] Christoph Lange,[3] and Joel N. Hirschhorn[1,2,4]*

[1]*Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts*
[2]*Divisions of Genetics and Endocrinology and Program in Genomics, Children's Hospital, Boston, Massachusetts*
[3]*Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts*
[4]*Department of Genetics, Harvard Medical School, Boston, Massachusetts*

The optimal method for considering different genetic models in association studies is not clear. We compared analytical strategies that use different genetic models to analyze genotype–phenotype information from association studies of quantitative traits in unrelated individuals. We created simulated datasets where the minor alleles are causal with an additive, dominant, or recessive mode of inheritance over a range of allele frequencies. We then computed power to detect these causal alleles using one or a combination of statistical models in a standard regression framework, including corrections for the multiple testing incurred by analyzing multiple models. Our results show that, as expected, maximal power is achieved when we test a single genetic model that matches the actual underlying mode of inheritance of the causal allele. When the inheritance pattern of the causal allele is unknown, the co-dominant model, a single two degrees of freedom test, has good overall performance in any of the three simple modes of inheritance simulated. Alternatively, it is slightly more powerful to analyze all three genetic models together (additive, dominant, and recessive), but only if the significance thresholds used to correct for analyzing multiple models are appropriately determined (such as by permutation). Finally, a commonly employed approach, testing the additive model alone, performs poorly for recessive causal alleles when the minor allele frequency is not close to 50%. Our observations were confirmed by analyzing an existing genetic association dataset in which we detect the effect of a *KCNJ11* variant on insulinogenic index in unrelated non-diabetic individuals. *Genet. Epidemiol.* 31:358–362, 2007.

Key words: Co-dominant; linear regression; permutation; Bonferroni; QTL

*Correspondence to: Joel N. Hirschhorn, Division of Genetics, Children's Hospital, 300 Longwood Avenue, Boston, MA 02115.
E-mail: joelh@broad.mit.edu
Received 21 November 2006; Revised 24 January 2007; Accepted 4 February 2007
Published online 13 March 2007 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/gepi.20217

INTRODUCTION

In genetic association studies, statistical power to detect disease susceptibility loci (DSLs) depends, among other factors, on the genetic models tested in the analysis: maximum power is reached when the 'true' mode of inheritance of the DSL and the genetic model used in the analysis are concordant. One apparent solution to this problem is to test several genetic models, but this increases the multiple testing burden, which may decrease, rather than increase power, once appropriate statistical thresholds are applied to take into account the multiple hypothesis testing. It is therefore of interest to determine which genetic model, or combination of models, maximizes power to detect DSLs in genetic association studies. This question has been addressed previously for family-based association studies [Laird and Lange, 2006; Lange et al., 2002; Lange and Laird, 2002]. The issue of genetic model testing has also been explored for dichotomous (case-control design) and quantitative traits using unrelated individuals, but in the setting of an analytic approach (likelihood testing) that is more complex than the analytic methods commonly implemented in the association study literature [Wang and Sheffield, 2005]. Here, we consider population-based association studies of quantitative traits, tested for association under a simple, widely used and flexible framework (linear regression). For this study design, we evaluate how statistical power varies depending on the

genetic model(s) tested, the mode of inheritance of the DSL (additive, dominant, or recessive), and its allele frequency, and draw conclusions regarding optimal strategies for testing different genetic models.

METHODS

Our simulations were performed using the statistical package R 2.4.0, assuming a continuous and normally distributed phenotype, a sample size of 1,000 unrelated individuals, and varying the minor allele frequency (MAF) of the bi-allelic DSLs from 5 to 50%. Essentially identical conclusions were drawn from simulations using a sample size of 200 unrelated individuals, with an expected slight increase in the effect of sampling variation at low allele frequencies (data not shown). Analysis of linear regression results with the *F*-test was used to assess significance of phenotype–genotype associations. Where we tested more than one genetic model, the best *F* statistic observed under any model (F_{max}) was used as an association test statistic.

Significance thresholds were determined empirically. For each model or combination of models, we generated an empirical distribution of association test statistics by performing 100,000 simulations under the null hypothesis (H_0: variance explained by the causal allele = 0%). Where more than one model was used, we recorded the distribution of F_{max} for those models under the null. To assess empirical significance of an observed *F* or F_{max} statistic, we compared the observed statistic to the corresponding distribution of test statistics generated under the null. Note that when single models are tested in large samples, the significance thresholds derived by simulations are essentially identical to the asymptotic values from the *F*-distribution. However, for consistency across our simulations, we have used simulation-based thresholds throughout this article when testing single models as well as combinations of models. This empirical approach allows us to use a permutation-based approach to maintain a constant type I error rate (α), regardless of which or how many genetic models were tested. We then computed power achieved by each of several genetic models, or combinations of models, for detecting DSLs under the alternative hypothesis (H_1: variance explained by the causal allele = 1%). Power for each analytic approach was estimated by performing 10,000 simulations under H_1 for each of three underlying genetic modes of inheritance and a range of minor allele frequencies.

RESULTS AND DISCUSSION

We first considered statistical power obtained by each of four individual statistical tests: additive, dominant, or recessive, each with one degree of freedom, and co-dominant, with two degrees of freedom. Although both the additive and dominant statistical models — two often highly correlated tests (see below and data not shown) — perform well when the minor allele of the DSL has an additive or dominant genetic effect (Figs. 1A and 1B), these models do not have good power to detect a recessive causal genetic locus, especially at MAF $<20\%$ (Fig. 1C). Reciprocally, the recessive statistical model performs poorly for DSLs that act additively or dominantly (Fig. 1). The co-dominant model, however, achieved relatively good power to detect DSLs in any of the simple genetic scenarios simulated here for MAF $\geq 5\%$ (Fig. 1).

Secondly, we asked whether combining the additive and/or dominant model(s) with the recessive model, and using the best test statistic from all tested models (in this case, an *F* statistic) would be a more powerful analytical strategy than simply testing the co-dominant model for association studies of quantitative traits in unrelated individuals. Importantly, we empirically determined significance thresholds by performing simulations under a null model and recording the best *F* statistics for the models used in the analysis, as described above. In Fig. 2, we compare the power achieved when testing the additive or co-dominant model alone with the power achieved when combining the additive and/or dominant model(s) with the recessive model. In our simulations, testing the three models was slightly more powerful across the three possible modes of inheritance (Fig. 2). For instance, whereas the additive model alone gives more power than a combination of the three models to detect additive DSLs, or dominant DSLs with MAF $<25\%$, the combination of three tests is overwhelmingly more powerful at detecting recessive causal allele than the additive model alone (Fig. 2). Intuitively, this means that the gain in power when testing the additive, dominant, and recessive statistical models together in our study design is sufficient to counterbalance the

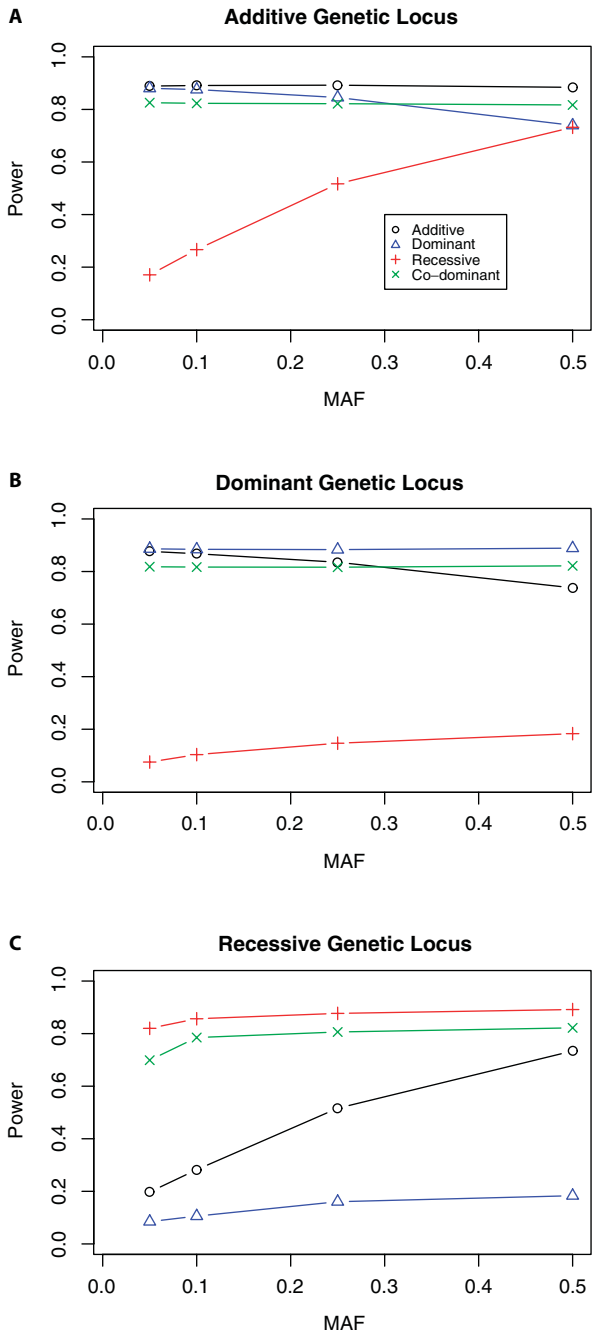

Fig. 1. Statistical power to detect (A) additive, (B) dominant, or (C) recessive disease susceptibility loci using four different statistical models (one degree of freedom: additive, dominant, recessive models; two degrees of freedom: co-dominant model) in population-based association studies of quantitative traits. Results shown were obtained using empirically determined significance thresholds as described in the text ($\alpha = 0.05$). For these simple models, the empirical thresholds were essentially identical to the asymptotic values from the appropriate *F*-distribution with one or two degrees of freedom.

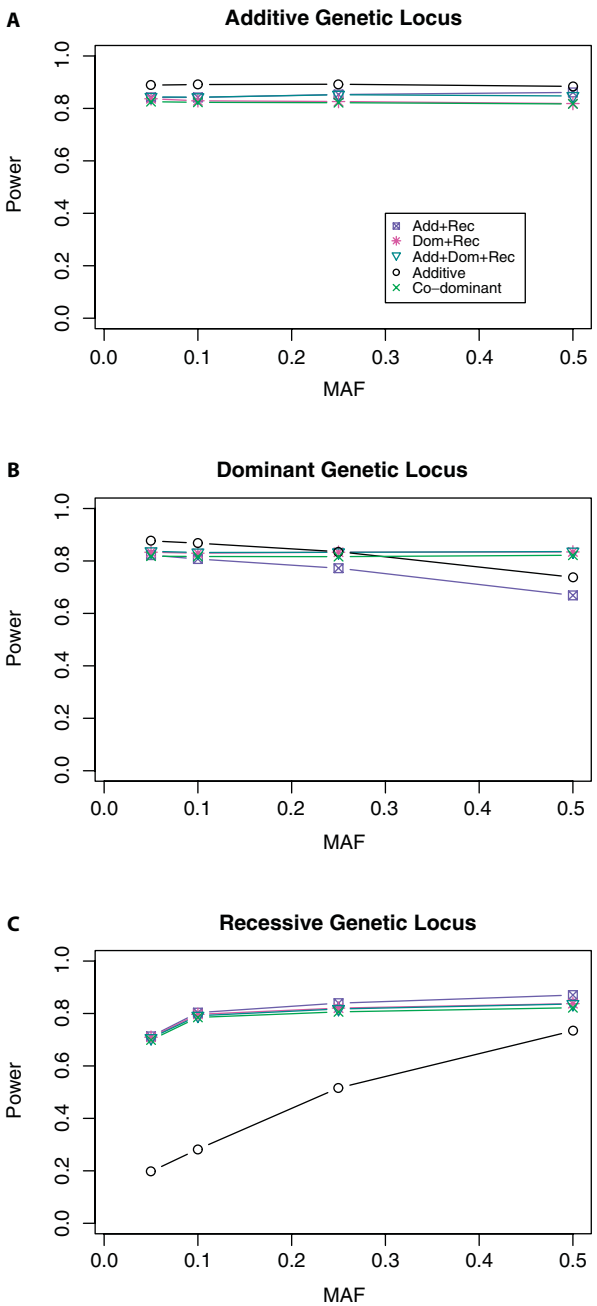

Fig. 2. Combining genetic models increases statistical power to detect (A) additive, (B) dominant, or (C) recessive disease susceptibility loci (DSLs) in association studies of quantitative traits in unrelated individuals. We compare the power achieved by single tests (additive and co-dominant models) with the power achieved when the additive and dominant models are combined with the recessive model. Combinations of models ("Add+Rec", "Dom+Rec", "Add+Dom+Rec") provide more power to detect recessive DSLs than the additive model alone. Combining the three one degree of freedom tests ("Add+Dom+Rec") is also slightly more powerful than the co-dominant alone to detect recessive DSLs. Results shown were obtained using empirically determined significance thresholds to maintain a type I error rate of 5%, as described in the text.

loss of power caused by the multiple testing problem, as long as appropriate thresholds of significance are used. We also note that testing the three models together provided slightly more power than testing the two degrees of freedom co-dominant model alone (compare "Add+Dom +Rec" with "Co-dominant" in Fig. 2), but these approaches were fairly similar.

Thirdly, we examined the impact of using empirical statistical thresholds, as opposed to simply performing a Bonferroni correction for the multiple genetic models tested. We compared the power observed when using Bonferroni-corrected significance thresholds from the one degree of freedom *F*-test statistics rather than the empirical thresholds that were determined using the permutation-based approach. We would expect Bonferroni correction to be overly stringent in this context, as none of the models tested are completely independent from each other, and the additive and dominant models are highly correlated at low minor allele frequencies. For example, the Spearman's rank correlation coefficient between the additive and dominant model for an additive DSL at MAF 10% is 0.82. Not surprisingly, using the Bonferroni-corrected thresholds led to a loss in power: in Fig. 3, all combinations of tests that used empirical significance thresholds achieved more power than the corresponding power for simple Bonferroni-corrected thresholds (compare the corresponding "Permutation" and "Bonferroni" results in Fig. 3). This difference in power was especially true when the three statistical models were tested, leading to a ~5% loss in power (compare "Add+Dom+Rec Bonferroni" and "Add+Dom+Rec Permutation" in Fig. 3).

Finally, we evaluated the effect of genetic model testing on the significance level of a real genotype-phenotype association. Florez et al. reported an association between a variant in the gene *KCNJ11* (E23K, rs5219) and a quantitative trait, insulinogenic index, in unrelated non-diabetic Scandinavians ($N = 674$); the authors obtained their most significant result when modeling a recessive locus [Florez et al., 2004]. Similarly, the recessive model was the most significant single test in our analysis ($P = 0.016$, Table I), and the result was not significant under an additive model ($P = 0.125$, Table I). We also assessed the significance level of this association if several genetic models had been tested, that is if we did not know the mode of inheritance of *KCNJ11* E23K, and we had to correct for testing multiple genetic models. Empirical *P*-values for this dataset, obtained by randomly

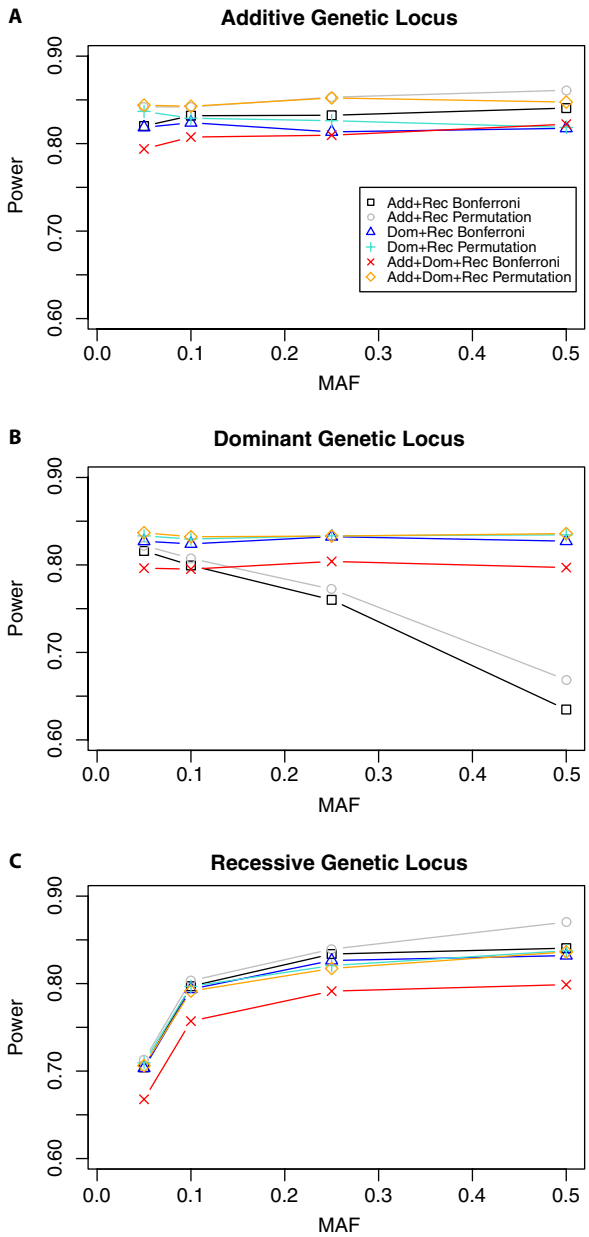

Fig. 3. Using Bonferroni-corrected asymptotic values, as opposed to empirically determined significance thresholds, reduces statistical power in population-based association studies to detect (A) additive, (B) dominant, or (C) recessive quantitative disease susceptibility loci. Permutation-based significance thresholds were obtained by generating null distributions of the best *F* statistics for each simulation, and selecting values that allow a type I error rate of 5%, as explained in the text. Bonferroni-corrected significance thresholds were obtained by dividing the type I error rate allowed (5%) by the number of models tested (by two for simulations with "Add+Rec" and "Dom+Rec", and by three for "Add+Dom+Rec") and retrieving the corresponding asymptotic values from the *F*-distribution. Bonferroni correction is too stringent and leads to a loss in power in all combinations of models tested; this is especially true when the additive, dominant, and recessive models are tested together ("Add+Dom+Rec Bonferroni" vs. "Add+Dom +Rec Permutation"). Please note the scale of the Y-axis.

TABLE I. Genetic model testing and significance level of *KCNJ11* E23K (rs5219) on the insulinogenic index of unrelated non-diabetic Scandinavians ($N = 674$) [Florez et al., 2004]

Model(s)	Asymptotic/ Corrected *P*-value[a]	Empirical *P*-value[b]
Additive	0.125	n.a.
Dominant	0.905	n.a.
Recessive	0.016	n.a.
Co-dominant	0.044	n.a.
Add+Rec	0.032	0.028
Dom+Rec	0.032	0.030
Add+Dom+Rec	0.048	0.037

[a]For the additive, dominant, recessive, or co-dominant model alone, asymptotic *P*-values were calculated from the *F*-distribution. When several statistical models were tested, corrected *P*-values were obtained by multiplying the lowest asymptotic *P*-value from the *F*-tests by the number of models tested: for "Add +Rec" and "Dom+Rec", the *P*-value for the recessive model (i.e. the lowest *P*-value) was multiplied by two to give the corrected *P*-value, and for "Add+Dom+Rec", it was multiplied by three.
[b]Empirical *P*-values were obtained by permuting the phenotype 10,000 times, and counting the number of times *P*-values lower than the lowest asymptotic *P*-value were obtained. n.a., not applicable: empirical and asymptotic *P*-values for a single test are identical because there is no correction needed.

permuting the phenotype data, were all lower than the Bonferroni-corrected *P*-values, indicating again that Bonferroni is a too severe correction because of the dependence of the models tested (compare "Corrected" and "Empirical" *P*-values in Table I). The co-dominant model did also relatively well by itself, and would have captured the association between *KCNJ11* E23K and insulin secretion at a significance level $\alpha < 0.05$ (Co-dominant $P = 0.044$, Table I). The co-dominant model, however, did not perform as well in this dataset as a combination of the three one degree of freedom tests, as long as empirical significance thresholds were used (*P*-values for the co-dominant and "Add+Dom+Rec Empirical" strategies are, respectively, 0.044 and 0.037, Table I). Similar results were obtained when analyzing an association between insulin secretion and a variant in the gene *TCF7L2* (data not shown) [Saxena et al., 2006].

CONCLUSION

In conclusion, we have used simulations and existing phenotype–genotype datasets to compare strategies for using different genetic models to analyze results from population-based association studies of quantitative traits. Based on our results, we recommend either testing the co-dominant statistical model alone, or alternatively testing the additive, dominant, and recessive models together but using empirically determined significance thresholds to correct for testing multiple correlated genetic models. The additive model performs well to detect additive or dominant DSLs, but testing the additive model alone, a common practice in association study genetics, does not maximize power if recessive effects are important contributors to the trait being studied.

ACKNOWLEDGMENTS

We thank Jose C. Florez and Richa Saxena for the *KCNJ11* E23K and *TCF7L2* rs7903146 datasets, respectively. J.N.H. is a recipient of a March of Dimes grant ♯6-FY04-61 and an ADA Smith Family Foundation Pinnacle Program Project.

REFERENCES

Florez JC, Burtt N, de Bakker PI, Almgren P, Tuomi T, Holmkvist J, Gaudet D, Hudson TJ, Schaffner SF, Daly MJ, et al. 2004. Haplotype structure and genotype-phenotype correlations of the sulfonylurea receptor and the islet ATP-sensitive potassium channel gene region. Diabetes 53:1360–1368.

Laird NM, Lange C. 2006. Family-based designs in the age of large-scale gene-association studies. Nat Rev Genet 7:385–394.

Lange C, Laird NM. 2002. Power calculations for a general class of family-based association tests: dichotomous traits. Am J Hum Genet 71:575–584.

Lange C, DeMeo DL, Laird NM. 2002. Power and design considerations for a general class of family-based association tests: quantitative traits. Am J Hum Genet 71:1330–1341.

Saxena R, Gianniny L, Burtt NP, Lyssenko V, Giuducci C, Sjogren M, Florez JC, Almgren P, Isomaa B, Orho-Melander M, et al. 2006. Common single nucleotide polymorphisms in TCF7L2 are reproducibly associated with type 2 diabetes and reduce the insulin response to glucose in nondiabetic individuals. Diabetes 55:2890–2895.

Wang K, Sheffield VC. 2005. A constrained-likelihood approach to marker-trait association studies. Am J Hum Genet 77:768–780.