A Pattern Discovery-Based Method for Detecting Multi-Locus Genetic Association

Zhong Li[1], Aris Floratos[1], David Wang[1], Andrea Califano[2]

Keywords: pattern discovery, multi-locus, genetic association, power calculation

1 Introduction.

Methods to effectively detecting global multi-locus genetic associations are becoming increasingly relevant in the genetic dissection of complex trait. Current approaches typically only consider a limited number of hypotheses, most of which are related to the effect of a single locus or of a relatively small number of loci co-localized on a chromosomal region, therefore do not accommodate the full range of genetic mechanisms that may contribute to a complex trait. We have developed a novel association analysis methodology (enGENIOUS) that is specifically designed to detect genetic associations involving multiple disease-susceptibility loci. Our approach relies on the efficient discovery of patterns comprising spatially unrestricted polymorphic markers and on the use of appropriate test statistics to evaluate pattern-trait association. Because markers within a pattern are not required to be in linkage (or even on the same chromosome), this method represents a truly global multi-locus association test. Power calculations using multi-locus disease models confirm the superior performance of out approach when compared to a frequency-based single marker analysis method. When applied on a real dataset, our method was successful in localizing a previously verified two-locus/gene interaction associated with Schizophrenia. In addition, we also identified a novel and less conspicuous association involving different markers on the same two genes, suggesting a role for genetic heterogeneity and population substructure in Schizophrenia.

2 Results.

enGENIOUS demonstrated better power than a single marker association test

A simulation-based power calculation was performed to evaluate the power of the pattern discovery-based method on multi-locus association analysis. Power was computed for two disease models: dominant-recessive and recessive-recessive, each involving two loci (markers M1 and M2). For each disease model, we considered five genotype frequency settings for disease-affected markers and two sample sizes (250 cases/250 controls or 500 cases/500 controls) (Figure 1). For each combination of disease model, frequency setting, and sample size, 500 simulated datasets were generated. As shown in Figure 1, the pattern discovery-based method consistently outperformed the single marker χ^2 test under both disease models (Figure 1A and 1B). As expected, power for both tests improved with larger sample size and higher genotype frequency. The power difference between these two methods demonstrates the advantage of using the pattern discovery-based method on association detection, especially when affected genotype frequency is low. Because no constraint has been established on the physical distance between markers M1 and M2, the same results would have been obtained even if M1 and M2 were on different chromosomes, thus providing direct evidence that the pattern discovery-based method is indeed a global association test.

Discovery of a novel gene-gene interaction in Schizophrenia patients

enGENIOUS was applied to a dataset collected from a case/control association study on Schizophrenia. The dataset contained genotypes for 28 SNP markers spanning 115 kb at 12q24 (co-localized with the DAO gene) and 266 kb at 13q34 (co-localized with the G72 gene). With the

[1] First Genetic Trust Inc., 201 Route 17 North, Suite 902, Rutherford, NJ 07070, USA. E-mail: zli@firstgenetic.net

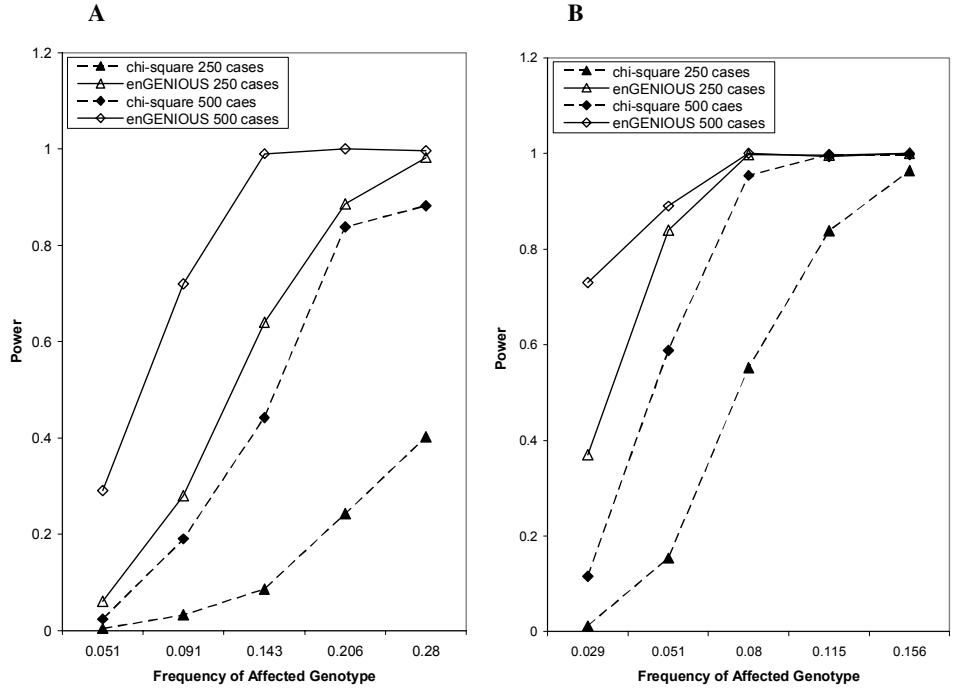
[2] Department of Biomedical Informatics, Columbia University, 1150 St. Nicholas Ave., New York, 10032, USA. E-mail: califano@dbmi.columbia.edu

association test rejection level for individual pattern as 9.49e-5 at the support constraint of 24, only two out of 5,952 identified patterns were found to be significantly associated with Schizophrenia. The most significant pattern (P=4.64e-5) included two markers, marker MDAAO-6 on chromosome 12, and M-22 on chromosome 13, the same two markers found to be associated with Schizophrenia in both genetic analysis and bioassay [1]. The second most significant pattern (P=7.33e-5) included seven markers, one marker (marker B-7) on chromosome 12, and markers B-1, B-2, B-3, B-4, B-5, and B-6 (B-1~B-6) on chromosome 13. None of the markers in this pattern was significant by itself when evaluated with the single-marker χ^2 test. To calculate the combined relative risk on DAO and G72 loci as characterized by Pattern A or Pattern B for Schizophrenia, we constructed a χ^2 test in which individuals carrying either of the two patterns or none of the two patterns were counted in cases and in controls. A highly significant P value (P=8.67E-11) indicated that the genetic interaction between DAO and G72 is strongly associated with Schizophrenia and multiple molecular mechanisms might be responsible for the susceptibility on those two loci.

3 Figures and tables.

Figure 1 Power Comparisons between a Single Marker χ^2 test vs. enGENIOUS.

Powers were compared between a single marker χ^2 test and the pattern discovery-based multi-locus association test under two disease models (dominant-recessive in A, recessive-recessive in B). A and B showed the power curves for both methods with two sample sizes (250 cases/250 controls and 500 cases/500 controls).



References

[1] Chumakov I, Blumenfeld M, Guerassimenko O, Cavarec L, Palicio M, Abderrahim H, Bougueleret L, et al. (2002) Genetic and physiological data implicating the new human gene G72 and the gene for D-amino acid oxidase in schizophrenia. *Proc Natl Acad Sci U S A* 99:13675-13680.