BIOINFORMATICS APPLICATIONS NOTE Vol. 19 no. 1 2003
Pages 149–150

Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits

S. Purcell[1,*], S. S. Cherny[2] and P. C. Sham[1]

[1]Social, Genetics and Developmental Psychiatry Research Centre, Institute of Psychiatry, King's College London, De Crespigny Park, London SE5 8AF, UK and [2]Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK

Received on May 16, 2002; revised on July 11, 2002; accepted on July 15, 2002

ABSTRACT

Summary: A website for performing power calculations for the design of linkage and association genetic mapping studies of complex traits.

Availibility: The package is made available at http://statgen.iop.kcl.ac.uk/gpc/

Contact: s.purcell@iop.kcl.ac.uk

An essential first step in the planning of any scientific study is to assess how many samples must be collected in order to achieve sufficient power to detect the hypothesized effect. In human genetics this requirement is particularly salient, given the costs involved in phenotyping and genotyping individuals. Futhermore, it is virtually impossible to submit a grant proposal to study a particular disease or trait without inclusion of detailed power calculations to show that the proposed research is likely to succeed, provided there is indeed a gene to be found. After conducting a study, power analysis can also shed light on negative results by indicating whether the study was underpowered, or what the smallest detectable effect size would be given the actual sample size. The power of a study is the probability of successfully detecting an effect of a particular size: if β is the probability of a false-negative (type II) error, then power is $1 - \beta$. Power depends on several factors: magnitude of effect, sample size, N, and required level of statistical significance, α (the false-positive, or type I, error rate). Although N and α are determined by the experimenter, many of the factors that contribute to the effect size are typically unknown. In order to compute power, we are therefore required to make assumptions regarding what we expect to find. For mapping loci, such factors include the proportion of variance explained by the trait locus, gene action, and marker heterozygosity and density. Although there is no shortage of statistical genetic literature to aid the researcher in performing such calculations (Cardon and Fulker, 1994; Carey and Williamson, 1991; Nance and Neale, 1989; Neale *et al.*, 1994; Schmitz *et al.*, 1998; Sham *et al.*, 2000; Suarez *et al.*, 1982) there are few software tools to make the task practical. The present paper describes an easy- to-use website which allows the researcher to quickly perform such necessary calculations.

Methods to map loci influencing complex traits fall into two broad classes: linkage and association. Linkage relies on correlating sharing of chromosomal segments among relatives with their similarity on a trait whereas association directly relates genotype to phenotype. Variance components models provide a powerful framework for both linkage and association mapping (Almasy and Blangero, 1998; Fulker and Cherny, 1996; Pratt *et al.*, 2000; Fulker *et al.*, 1999). It has been shown that maximum-likelihood variance components approaches to linkage mapping of quantitative trait loci (QTL), which utilize the full familial covariance structure, are more powerful than simple regression-based methods (Fulker and Cherny, 1996) Additionally, for association mapping, a powerful variance components approach has been presented (Fulker *et al.*, 1999) which allows simultaneous modelling of linkage and association while controlling for population stratification effects in sibship data, by considering both between-sibship and within-sibship variation. This method has been made accessible by the release of QTDT (Abecasis *et al.*, 2000a) and has been generalized to deal with extended families rather than just sibships (Abecasis *et al.*, 2000b).

The computationally intensive approach to power calculation is to simulate hundreds or thousands of replicate samples under a specifed set of population parameters. The proportion of replicates in which an effect is detected (the test statistic falling above a specified threshold) provides an estimate of power. Recently, however, closed-form analytic power equations have been presented for

*To whom correspondence should be addressed.

variance components methods of linkage and association mapping (Sham *et al.*, 2000). The use of such equations greatly speeds up power calculation and allows a more comprehensive exploration of the parameter space (e.g. different models and sample types). The Genetic Power Calculator (GPC) implements these power equations and others, for both linkage and association methods using either qualitative (e.g. the presence or absence of a disease) or quantitative (e.g. a score on a personality inventory) traits.

Power for the variance components linkage test can be calculated for sibships of arbitrary size, under user-definable levels of the proportion of variance explained by the trait locus, acting additively and/or via dominance. The background residuals sibling correlation can be varied, as can the polymorphism information content at the locus of interest, allowing accommodation of either twopoint or multipoint linkage. The output includes a table of power for various common α levels as well as a user-selected α level, and required sample size to achieve the user-selected desired level of power, suitable for direct inclusion in a grant proposal. In addition the noncentrality parameter is provided to facilitate power calculation for samples of variable-sized sibships. For tests of association, GPC uses the variance components test described by Fulker *et al.* (1999). The user is presented with a similar set of options as for the linkage test, along with association-specific options such as the extent of trait locus-marker locus linkage disequilibrium and allele frequencies. Output is presented for between-sibship, within-sibship and combined tests of association.

For testing association in discrete (disease) traits, tools are available for both the TDT test, which employs parents and a single affected offspring (Spielman *et al.*, 1993) and the case-control design (Sham, 1998). The user is able to explore power under conditions of varying disease allele frequency, disease prevalence, and genotype relative risk. Again, output is similar to that described above. In addition, calculation of power for the quantitative TDT and quantitative case-control designs is also available. These study designs assume that cases and controls are defined as scoring above or below specific thresholds. Additional utilities made available on the website include two-locus linkage power calculations and a facility for calculating the potential informativeness of sibships for linkage, conditional on observed trait values. This index of informativeness provides a basis for efficient selective genopyting (Purcell *et al.*, 2001). Presently, there is no software widely available to employ such tests, but the situation is likely to improve in the near future. We will attempt to add additional tools for estimating power to this website as additional methods of analysis are developed and software distributed for their implementation.

ACKNOWLEDGEMENTS

Supported in part by National Institutes of Health (USA) grant EY-12562, MRC components grant G9700821 (UK) and the Wellcome Trust.

REFERENCES

Abecasis,G.R., Cardon,L.R. and Cookson,W.O. (2000a) A general test of association for quantitative traits in nuclear families. *Am. J. Hum. Genet.*, **66**, 279–292.

Abecasis,G.R., Cookson,W.O. and Cardon,L.R. (2000b) Pedigree tests of transmission desiquilibrium. *Eur. J. Hum. Genet.*, **8**, 545–551.

Almasy,L. and Blangero,J. (1998) Multipoint quantitative-trait linkage analysis in general pedigress. *Am. J. Hum. Genet.*, **62**, 1198–1211.

Cardon,L.R. and Fulker,D.W. (1994) The power of interval mapping of quantitative trait loci, using selected sib pairs. *Am. J. Hum. Genet.*, **55**, 825–833.

Carey,G. and Williamson,J. (1991) Linkage analysis of quantitative traits; incresed power by using selected samples. *Am. J. Hum. Genet.*, **49**, 786–796.

Fulker,D.W. and Cherny,S.S. (1996) An improved multipoint sib-pair analysis of quantative traits. *Behavior Genet.*, **26**, 527–532.

Fulker,D.W., Cherny,S.S., Sham,P.C. and Hewitt,J.K. (1999) Combined linkage and association sib-pair analysis for quantitative traits. *Am. J. Hum. Genet.*, **64**, 259–267.

Nance,W.E. and Neale,M.C. (1989) Partitioned twin analysis: a power study. *Behavior Genet.*, **19**, 143–150.

Neale,M.C., Eaves,L.J. and Kendler,K.S. (1994) The power of the classical twin method to resolve variation in threshold traits. *Behavior Genet.*, **24**, 239–258.

Pratt,S.C., Daly,M.J. and Kruglyak,L. (2000) Exact multipoint quantitative-trait linkage analysis in pedigress by variance components. *Am. J. Hum. Genet.*, **66**, 1153–1157.

Purcell,S., Cherny,S., Hewitt,J. and Sham,P. (2001) Optimal sibship selection for genotyping in quantitative trait locus linkage analysis. *Human Heredity*, **52**, 1–13.

Schmitz,S., Cherny,S.S. and Fulker,D.W. (1998) Increase in power through multivariate analyses. *Behavior Genet.*, **28**, 357–363.

Sham,P. (1998) *Statistics in Human Genetics*, Ist edn, Arnold, London.

Sham,P.C., Cherny,S.S., Purcell,S. and Hewitt,J.K. (2000) Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *Am. J. Hum. Genet.*, **66**, 1616–1630.

Spielman,R.S., McGinnis,R.E. and Ewens,W.J. (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.*, **52**, 506–516.

Suarez,B., O'Rourke,D. and Van Eerdewegh,P. (1982) Power of the affected-sib-pair method to defect disease susceptibility loci of small effect: an application to multiple sclerosis. *Am. J. Med. Genet.*, **12**, 309–326.