Commentary

From mouse to human: Fine mapping of quantitative trait loci in a model organism

Mary Sara McPeek*

Departments of Statistics and Human Genetics, University of Chicago, 5734 South University Avenue, Chicago, IL 60637

The classical theory of inheritance of quantitative traits, or traits exhibiting continuous variation, specifies that a very large number of genetic determinants each contribute a small amount to the phenotype or trait value (1). Such quantitative traits present a much greater challenge for mapping and identification of genetic determinants than do simple Mendelian traits, especially in humans and other natural populations. Success in mapping quantitative trait loci (QTL) in humans has so far been limited (2). Model organisms, such as mouse, have the tremendous advantage that breeding can be controlled by the experimenter. Furthermore, one can expect that some genetic determinants of quantitative traits in mouse will have homologues in humans. In a paper in this issue of PNAS, Mott *et al.* (3) describe a promising technique for QTL localization in mouse by using a genetically heterogeneous stock (HS).

The problem of mapping genetic determinants for observed traits can be divided into stages, typically including initial detection, localization or fine mapping, and positional cloning. Each stage requires distinct strategies and methods of analysis. For initial detection of loci influencing a trait, study designs in which related individuals are separated by a small number of meioses can be advantageous, because the lower number of recombinations between relatives makes it more likely to detect a gene even with coarse marker resolution. Such study designs include family studies in humans and backcross and intercross designs in model organisms. Initial detection of a genetic locus generally does not yield much precision on its location, leaving too vast a region to explore for positional cloning. Localization or fine mapping is the process of refining the location of the genetic determinant or determinants, and for this purpose, study designs with an increased number of meioses between individuals, narrowing the expected preserved segment lengths and hence increasing the resolution, are desirable. In contrast to linkage studies, linkage disequilibrium or association mapping is based on sharing of alleles or haplotypes by individuals sampled from a population who are presumably only very distantly related and so are separated by many more meioses than individuals in typical linkage studies. For that reason, association mapping has always held out the promise of allowing more precise localization of genetic determinants.

Association mapping in humans and other natural populations is confronted with several thorny problems. First, a mutation at a locus of interest may have arisen more than once in the history of a population and so may be associated with multiple ancestral haplotypes. Furthermore, it may show allelic heterogeneity. Second, the extent of linkage disequilibrium around the variant is strongly affected by the age of the mutation and by the demographic history of the population from which it is sampled, about which there is often little information. Third, the ancestral haplotype on which a mutation occurred is generally unknown, as are the frequencies of alleles or haplotypes in the population at the time of the mutation. Fourth, it is necessary to ascertain appropriate controls while minimizing confounding caused by population stratification and environmental variables.

In a model organism such as mouse, there is much greater opportunity to exert control over, for example, population demography, genetic heterogeneity, and environment. Darvasi (4) gives a detailed overview of a number of study designs for QTL mapping and tradeoffs among them in terms of cost and resolution. Typical designs involve a pair of inbred lines crossed in various ways, which can be conveniently analyzed (5–8). One disadvantage of this lack of genetic heterogeneity is that some QTL of interest may not be segregating in the pair of inbred lines. In contrast, in the HS design used in Mott *et al.* (3), the founding individuals of the HS are 40 mating pairs that are the result of an 8-way cross of 8 distinct inbred lines. Each of the 60 generations bred so far consists of 40 pairs obtained from random mating among those individuals in the previous generation that have no grandparents in common (9). The HS provides the possibility of mapping QTL segregating in any one of the eight lines.

Mott *et al.* (3) localize five previously detected QTL for fearfulness in HS mice. They analyze the data by calculating, for various points along a chromosome, the probability that the haplotypes in an individual were inherited from each of the various inbred strains, making use of information on multiple markers through the hidden Markov method (10). From this information, it is possible to test for heterogeneity of founder strain effects. The novel analysis method of Mott *et al.* (3) shares some elements in common with recent association mapping methods developed for humans (11, 12) but with a study design that makes the problem much more tractable. For instance, all of the ancestral haplotypes are available for study, and the number of generations since the ancestral strain are known, as well as the mating pattern and population history. These advantages make the approach particularly promising. Mott *et al.* (3) compare single-point and multipoint analyses. Their results clearly highlight both the tremendous power improvements that can be obtained with multipoint methods as well as their main drawback: sensitivity to map and genotyping errors. Mott *et al.* (3) deal with the latter problem by artificially setting the number of generations higher to decrease dependence among nearby markers.

There are several statistical issues related to HS mapping that have yet to be addressed. First, when individuals are sampled from a small isolated population, the problem arises of taking account of the varying degrees of relationship among them. For quantitative traits, phenotypes for a sample from such a group of related individuals are typically modeled by a multivariate normal distribution, with co-

See companion article on page 12649.

*E-mail: mcpeek@galton.uchicago.edu.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.240463597. Article and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.240463597

variance matrix depending on the relatedness of the individuals. These covariances among individuals arise as the result of many QTL, usually assumed to be of small effect, throughout the genome, at which closer relatives will be more likely to share alleles. A hypothesis test for detection of a particular QTL would then typically be performed against this background of correlation among relatives because of many other QTL. Dependence among individuals related by a pedigree can also have an impact on such statistical procedures as the permutation test and the bootstrap. Iturria *et al.* (13) discuss a variation on a permutation test when the data consist of genotypes and phenotypes collected on a set of relatives. Mott *et al.* (3) make no use of the pedigree connecting the individuals in the study. Assuming that it is available, incorporating it into the analysis may lead to improvement.

Second, the question arises of establishing a threshold for detection when multiple hypothesis tests are performed throughout a region or the entire genome. Bonferroni correction (in which the smallest P value, among all of the hypothesis tests performed, is multiplied by the number of tests performed) becomes increasingly conservative as the marker density increases. In special cases of very simple relationships with dense markers, theoretical determination of an appropriate threshold is possible (14, 15). In the case of the HS population, determination of such a threshold by simulation, if not by theory, seems possible.

Mott *et al.*'s (3) strategy of QTL localization in HS combines elements of both linkage and association mapping. Their method of analyzing the data is intriguing and appears to lend itself to further methodological development. Their results suggest that an HS strategy gives the capability to localize QTL detected in the eight component strains and to distinguish closely linked QTL.

The support of the National Institutes of Health/National Human Genome Research Institute (HG01645) is gratefully acknowledged.

1. Fisher, R. A. (1918) *Trans. R. Soc. Edinburgh* **52,** 399–433.
2. Risch, N. J. (2000) *Nature (London)* **405,** 847–856.
3. Mott, R., Talbot, C. J., Turri, M. G., Collins, A. C. & Flint, J. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 12649–12654.
4. Darvasi, A. (1998) *Nat. Genet.* **18,** 19–24.
5. Lander, E. S. & Botstein, D. (1989) *Genetics* **121,** 185–199.
6. Zeng, Z. B. (1994) *Genetics* **136,** 1457–1468.
7. Jansen, R. C. (1992) *Theor. Appl. Genet,* **85,** 252–260.
8. Jansen, R. C. & Stam, P. (1993) *Genetics* **135,** 205–211.
9. McClearn, G. E., Wilson, J. R. & Meredith, W. (1970) in *Contributions to Behavior-Genetic Analysis: The Mouse as a Prototype*, eds. Lindzey, G. & Thiessen, D. (Appleton-Century-Crofts, New York), pp. 3–22.
10. Baum, L. E. (1972) *Inequalities* **3,** 1–8.
11. Service, S. K., Lang, D. W., Fremer, N. B. & Sandkuijl, L. A. (1999) *Am. J. Hum. Genet.* **64,** 1728–1738.
12. McPeek, M. S. & Strahs, A. (1999) *Am. J. Hum. Genet.* **65,** 858–875.
13. Iturria, S. J., Williams, J. T., Almasy, L. & Blangero, J. (1999) *Genet. Epidemiol* **17,** S169–S173.
14. Feingold, E., Brown, P. O. & Siegmund, D. (1993) *Am. J. Hum. Genet.* **53,** 234–251.
15. Lander, E. & Kruglyak, L. (1995) *Nat. Genet.* **11,** 241–247.