

In Silico Mapping of Mouse Quantitative Trait Loci

An inbred strain survey is an essential first step to serious genetic analysis of a complex, quantitative trait. Thus, the ability to use mean phenotypic values from inbred strains to map likely genomic locations of quantitative trait loci (QTLs) “in silico,” as recently proposed by Grupe *et al.* (1), in theory constitutes a major advance, although mapping techniques based on fixed genotypes (e.g., use of recombinant inbred strains) are not entirely novel. Unfortunately, we believe that this method is too good to be true. We created a spreadsheet that contains calculated genotypic differences and phenotypic differences between strains and that performs the correlation analysis as described by Grupe *et al.* (2). To date, we have been unable to replicate their findings—using either the traits they examined or our own experimentally confirmed QTLs—when the CAST/Ei strain, the source of most of the single nucleotide polymorphisms (SNPs), is omitted. In view of its genetic outlier status, whether or not inclusion of this strain is appropriate in such an analysis is a matter of debate; as a practical matter, however, these mice are far too hyperactive to be useful for most traits involving behavioral testing.

Although we believe that detecting QTLs from inbred haplotypes may indeed be possible, we have been unsuccessful thus far in our own independent attempts to develop an in silico mapping algorithm, due to statistical power limitations. Grupe *et al.* artificially “solved” this problem by taking eight strain observations and creating pairwise differences resulting in correlation analyses with 27 degrees of freedom. Further, their method is biased by the lack of independence of linked markers. Although this autocorrelation can be resolved (3), no attempt was made to do so, resulting in a large proportion of false positive results. Peak detection is achieved by an arbitrary method of determining significance thresholds, in which the user selects the percentage of results that are considered peaks of linkage. While this method appears to resemble a permutation analysis, by not considering all possible results it does not control Type I or Type II errors. Although a threshold can be selected that optimizes the error rate of this method for previously mapped traits, the authors have provided no objective procedure for selection of a threshold for novel traits. Fisher’s exact test was performed to demonstrate the adequacy of the error rate of this method as compared to standard mapping techniques; however, significance of the results was likely due to the easily replicable

high number of negative results relative to positive results predicted across the genome.

Practical in silico mapping may still be possible if these issues of statistical power, error control, and genetic variability are resolved. However, the present method does not appear to be a viable alternative to current QTL methodology. We fear that funding for, and attention to, standard QTL mapping studies might be jeopardized by the assumption that they have become moot. For the time being, in vivo mapping still appears indispensable.

Elissa J. Chesler

*Department of Psychology and
Neuroscience Program*

*University of Illinois at Urbana-Champaign
Champaign, IL 61820, USA*

Sandra L. Rodriguez-Zas

*Department of Animal Sciences
University of Illinois at Urbana-Champaign*

Jeffrey S. Mogil

*Department of Psychology
McGill University*

*Montreal, QC H3A 1B1, Canada
E-mail: jeff@hebb.psych.mcgill.ca*

References and Notes

1. A. Grupe *et al.*, *Science* **292**, 1915 (2001).
2. The spreadsheet can be found at www.nervenet.org/xlfiles/SNP/CheslerSNPMapper.xls.
3. E. S. Lander, L. Kruglyak, *Nature Genet.* **11**, 241 (1995).

31 August 2001; accepted 3 December 2001

Grupe *et al.* (1) recently published a work suggesting an efficient method to detect QTLs by using known phenotypes and genotypes of inbred strains of mice and a mathematical algorithm. However, the method of Grupe *et al.* is not of practical use for most relevant cases of QTL mapping.

To evaluate the method, I calculated the number of inbred strains necessary to detect a hypothetical QTL affecting a quantitative trait with heritability of 50%. To strengthen the argument, I have assumed a series of optimal conditions that favor the approach of Grupe *et al.*: (i) The functional SNP at the QTL is the one being tested. (ii) The allelic frequencies at that SNP are 50% for each allele. (iii) The proportion of variance explained by the QTL is not decreased due to the increased genetic variation arising from the heterogeneity of multiple strains. (iv) The phenotypic values for all strains are assumed to be known without any error variance (equivalent to being determined by an infinite number of animals for each strain). Thus, the

proportion of variance explained by the QTL is twice that of the same QTL in an F_2 population. With an appropriate Type I error threshold (2) and with the standard equation $n = Z^2_{1-\alpha/2}/V_{QTL}$ (3), the number of inbred strains required to detect a QTL with 50% power can be calculated for a range of proportions of variance explained by the QTL (V_{QTL}). QTLs are often found to be of a magnitude of $V_{QTL} = 5 - 20\%$ in an F_2 population, thus corresponding to 10% and 40% with inbred strains under the assumptions above. Based on the above calculations, detection of a QTL will require approximately between 40 and 150 inbred strains. It should be further noted that for most instances optimal conditions will not apply; thus, these numbers represent the lower limit. The number of strains available with known genotypic and phenotypic information is well below these numbers. For most of the traits considered by Grupe *et al.* (1), the number of strains used was not more than eight, and in some cases only four strains were considered. These numbers are insufficient to provide useful information.

In this comment, I claim that the method presented for in silico mapping is mostly irrelevant. Nevertheless, the authors have presented results indicating the utility of their research. How can we thus explain the discordance between the theoretical expectation and the experimental results presented by the authors? Grupe *et al.* (1) did not provide enough details on all traits analyzed to entirely solve the discrepancy. Nevertheless, enough information is provided to indicate the source of the errors and misinterpretations.

First, I considered the two traits on which complete information has been provided, the major histocompatibility complex (MHC) K locus and airway hyperresponsiveness (AHR). For the MHC, the authors presented a diagram [figure 2A in (1)] with four peaks crossing their threshold, but in the text and in table 1, only one is reported. Furthermore, this locus is not a QTL, but rather a monogenic trait (for which the method suggested may indeed be applicable with great efficiency). For AHR, the authors present four QTLs previously discovered by conventional methods. This, however, is only a selected subset out of the QTLs reported in the literature. For example, the authors have chosen to include the QTL on chromosome (chr) 7 with a LOD score of 1.9 (4), but did not include the QTLs on chr 9 and 17, with LOD scores of 2.5 and 2.1, respectively (5). Another study (6) also identified a QTL on chr 6 with a LOD score >3.0 , which was not included in the analysis. The three loci that were not included in the analysis (chr 6, 9, and 17) showed no correlation in figure 2B (1).

Similar errors are also present in other traits, as presented in the supplementary ma-

TECHNICAL COMMENTS

terial on the Web (8). For example: (i) The QTL on chr 11 for alcohol preference (7) was not mentioned. (ii) For lymphoma, the authors cited Mucenski *et al.* (9) and included two QTLs from that study, although the research was not a QTL mapping report and significant evidence for those loci was not given. (iii) A QTL for PKC activity was reported on chr 11 citing Dwyer-Nield *et al.* (10), who also reported an even stronger QTL for PKC activity on chr 3. (iv) For PKC content, however, a QTL was reported on chr 3, whereas Dwyer-Nield *et al.* (10) reported an even stronger QTL on chr 11.

The success rate reported in identifying 15 out of 26 QTLs is not that impressive in light of the biased method of counting as outlined above and in light of the high false positive rate, approximately 0.06 (obtained by 24/400). Consequently, following an *in silico* QTL mapping experiment, one would still need a traditional QTL mapping approach to sort out the false positives and identify additional QTLs. The authors did not present compelling evidence to support their statement on the reduction of experimental time from many months to milliseconds. Nevertheless, the method itself is of interest, innovative and can be, in my view, of relevance in two instances: (i) for the analysis of genes explaining most of the genetic variation (usually monogenic traits), and (ii) as a preliminary efficient scan prior to the initiation of a traditional QTL study.

Ariel Darvasi

*Department of Evolution Systematics
and Ecology
The Hebrew University of Jerusalem
Jerusalem 91904, Israel
and IDgene Pharmaceuticals Ltd.
Beit Ofer, 5 Heftzadi St.
Post Office Box 34478
Jerusalem 91344, Israel
E-mail: arield@cc.huji.ac.il*

References and Notes

1. A. Grupe *et al.* *Science* **292**, 1915 (2001).
2. E. S. Lander, N. J. Schork, *Science* **265**, 2037 (1994).
3. A. Darvasi, *Nature Genet.* **18**, 19 (1998).
4. S. L. Ewart *et al.*, *Am. J. Respir. Cell Mol. Biol.* **23**, 537 (2000).
5. Y. Zhang *et al.*, *Hum. Mol. Genet.* **8**, 601 (1999).
6. G. T. De Sanctis *et al.*, *Am. J. Physiol.* **277**, L1118 (1999).
7. K. J. Buck *et al.*, *J. Neurosci.* **17**, 3946 (1997).
8. www.sciencemag.org/cgi/content/full/292/5523/1915/DC1
9. M. L. Mucenski *et al.*, *Mol. Cell. Biol.* **6**, 4236 (1986).
10. L. D. Dwyer-Nield *et al.*, *Am. J. Physiol. Lung Cell Mol. Physiol.* **279**, L326 (2000).

3 October 2001; accepted 3 December 2001

Response: In response to the comments of Chesler *et al.*, two key points must be emphasized. (i) The concept underlying our computational algorithm is sound and relatively simple. The program searches the SNP

database for genomic regions where allelic sharing is concordant with the phenotypic differences among the strains. (ii) The method was shown to correctly predict the chromosomal region for the MHC and the regions identified by QTL mapping for a number of traits, including airway hyperresponsiveness and alcohol withdrawal.

The spreadsheet implementation of our method by Chesler *et al.* is generally correct, but a significant error can easily be generated when using their spreadsheet. Users have to manually remove phenotypic differences that were calculated with strains that have an unknown phenotype. If the data are not removed, a zero is assumed for the strain without phenotypic data, which results in the generation of an incorrect phenotypic difference matrix. This drastically compromises the results and appears to have led to the erroneous conclusions of Chesler *et al.* about our method. When this error is corrected, their spreadsheet method reproduces our published results for five traits with identical phenotypic data. We performed *in silico* mapping for a total of 10 phenotypic traits (1). In their attempt to reproduce our results, Chesler *et al.* analyzed seven of these traits and altered the input phenotypic data for two of them. For the remaining five traits, the experimentally verified QTL is consistently within the top 10% of regions predicted by their implementation. Because the CAST/Ei strain-specific SNPs are removed in their implementation, they confirm that *in silico* mapping is possible with even fewer SNPs than were used in our study. The only discordant results arise from the two traits for which different phenotypic input data were used.

The comments of Chesler *et al.* on the statistical aspects of our method are also misleading. Because *in silico* mapping is by definition an artificial process, we used artificial methods to make our computational predictions. The problem of uneven distribution of SNP markers was recognized and will lessen as the number of SNPs in the database increases. The computational prediction method does entail Type I and Type II errors, which also occur with *in vivo* mapping studies. However, our sensitivity analysis showed the effect of manipulating the threshold of significance and enabled evaluation of acceptable false positive and negative results.

Chesler *et al.* also misinterpret the impact of our study on funding for mouse genetics and experimental QTL analysis. There had been widespread concern in the scientific community and among funding agencies that QTL analysis was expensive, lengthy, and unproductive. Rather than reducing funding, our computational method

and SNP database should markedly increase interest in and productivity from mouse genetic research. The high-throughput genotyping method and SNP database we described were developed to facilitate experimental QTL mapping, so we certainly agree that this is important.

Darvasi raises two significant concerns about our computational method for predicting chromosomal regions regulating complex traits. He provides a theoretical framework indicating that the method cannot work, and alleges that a "biased method of counting" was used for assessing the accuracy of the computational predictions. Before addressing his theoretical concerns, we will demonstrate the absence of the alleged bias by pointing out key errors in Darvasi's comments, using the two phenotypic traits (MHC and AHR) where he provided detail comments.

Darvasi indicates that we incorrectly represented the MHC analysis. As indicated in figure 2 of Grupe *et al.* (1), the computational method identified four chromosomal regions that were most highly correlated with the phenotypic matrix at a 10% cutoff. The region containing the MHC was the highest prediction, a full two standard deviations above any other predicted region. In table 1 of (1), we indicated that the chromosomal region containing the MHC was correctly identified when 2% of the genome was within the computationally predicted regions. There was no discrepancy or deception in this presentation. A different cutoff value for each trait was provided in table 1 (1). The cutoff value was the percent of the mouse genome included within the computationally predicted regions containing the correct (or experimentally verified) chromosomal region. This representation was used to consolidate many different tables, each using a different cutoff value (range 5 to 30%), into a single table. In contrast to what Darvasi assumes, and as clearly stated in the paper, the performance of the computational method was assessed using a constant cutoff for all 10 traits examined. At a 10% cutoff value, we did indeed find that 15 of 26 experimentally identified QTL intervals were correctly identified by the computational method (1).

Darvasi also alleges that additional bias was introduced through use of a "selected subset" of literature QTLs as experimentally verified intervals and provides the AHR trait as a specific example. He is concerned that three published QTL intervals (chromosomes 6, 9, and 17) were selectively excluded, because they were not predicted by the computational method; meanwhile, a chromosome 7 QTL (LOD 1.9), which was predicted by the computational method, was included in the analysis. Darvasi criticizes us for not including a QTL on chro-

TECHNICAL COMMENTS

mosome 6 (LOD score 3), identified by De Sanctis *et al.* (2), in our analysis. Whereas they analyzed basal (noninflammatory) airway responses, our analysis focused on antigen-induced (inflammatory) airway responses. Because the chromosome 6 interval does not regulate antigen-induced airway responsiveness, it was appropriately excluded from our analysis. The genetic elements regulating basal airway responsiveness are distinct from those regulating antigen-induced responses.

Darvasi also asks why published QTL intervals on chromosomes 10 and 11 were included, whereas two other QTL intervals (chromosomes 9 and 17) found in the same study (3) were excluded. Because our study was the first analysis of this type, our threshold for inclusion of published, experimentally verified QTL intervals was not determined by accepted criteria. It is likely that inclusion criteria will be more rigid in subsequent analysis. However, we based our decision on comments within the paper (3) indicating that “linkages to chromosome 10 and 11 were significant” but that linkage to chromosomes 9 and 17 “would be classified as ‘suggestive.’” Nevertheless, inclusion or exclusion of the chromosome 9 and 17 QTLs in the analysis would not significantly alter the fact that our computational method performed exceedingly well in predicting chromosomal regions regulating allergen-induced AHR. Two of us (Peltz and Grupe) were co-authors on the study (4) that identified the chromosome 7 QTL included in our analysis, and we can definitely state that QTL intervals on chromosomes 2 and 7 were the only ones identified in that study examining allergen-induced AHR. In contrast to the study of Zhang *et al.* (3), which analyzed F_2 progeny, our study analyzed BC1 mice (4). LOD scores arising from analysis of BC1 progeny tend to be lower than F_2 mice. Because of this, the chromosome 7 locus was included within the experimentally verified intervals.

We next address the apparent paradox that the computational prediction method appears to correctly predict chromosomal regions identified by experimental analysis, despite Darvasi’s suggestion that this method cannot provide useful information in these situations. It is likely that his underlying assumptions are not applicable to our

in silico prediction method, which is quite distinct from conventional QTL analysis. For example, we did not assume, as Darvasi indicates, that “[t]he functional SNP at the QTL is the one being tested.” In contrast, our computational program identified genomic regions, irrespective of whether the “functional SNP” was in the database, among the mouse strains analyzed in which allelic sharing was concordant with the phenotypic differences.

Also, the equation Darvasi uses to calculate the number of inbred strains required ($n = Z^2_{1-\alpha/2}/N_{QTL}$) and the significance thresholds applied are not appropriate for our computational prediction method, for several reasons. (i) The Darvasi equation for n presumes Lander and Schork significance levels for an F_2 (5×10^{-5}), which leads to his calculation that 42 strains were required for a QTL accounting for 40% of the trait variance. However, his criterion for statistical significance (5×10^{-5}) is based on an infinite number of genotypes (animals) using an infinite density of markers, which is not applicable to our computational method. A permutation test would undoubtedly estimate a much more relaxed criterion for significance. (ii) Darvasi’s reasoning does not take into account that in an F_2 analysis each mouse has a unique genotype, whereas among inbred strains each genotype can be replicated any number of times. This greatly reduces the environmental sources of variation and makes the proportion of the trait variance due to a chromosomal region much higher among inbred strains than the same QTL would have in an F_2 . Darvasi states that the difference is twofold because of the absence of heterozygotes that make up one-half of F_2 populations (they contribute little to QTL detection), but the difference is likely to be much larger than twofold. (iii) It is not necessary for statistical significance to be attained in order for the method to be useful.

In summary, our program identified genome segments that are likely to contribute to quantitative traits through examining phenotypic differences among inbred strains, and does not require lengthy breeding and genotyping experiments. The usefulness of standard inbred strains as a QTL mapping resource has been almost entirely overlooked in the past. With the advent of the mouse phenome project, which will provide data for hundreds of medically important traits across the more commonly used inbred strains, our

method is one that can mine this wealth of phenotypic information for QTL information within and between traits. We did not claim that this approach would replace traditional QTL analysis for confirming the identity of such genome segments. Indeed, we presented genotyping tools for improving traditional QTL analysis in the same paper. Because QTLs of large effect are likely to be detected by the computational method, conventional crosses may still be needed for many traits. But these considerations do not make our method “irrelevant.” Quite the contrary: Every new source of QTL information is valuable, especially when the source we used has been underutilized by the QTL research community in the past. We hope that publication of our computational method will lead to additional testing and improved understanding of how it works, and will inspire others to develop even better computational methods and databases in the future.

Jonathan Usuka
Andrew Grupe

Department of Genetics & Genomics
Roche Bioscience
Palo Alto, CA 94303, USA

Soren Germer
Roche Molecular Systems
Alameda, CA 94501, USA

Dee Aud
Department of Genetics & Genomics
Roche Bioscience

John K. Belknap
Robert F. Klein
Oregon Health Sciences University
and Portland Veterans Affairs
Medical Center
Portland, OR 97201, USA

Mandeep K. Ahluwalia
Russell Higuchi
Roche Molecular Systems

Gary Peltz
Department of Genetics & Genomics
Roche Bioscience
E-mail: gary.peltz@roche.com

References

1. A. Grupe *et al.*, *Science* **292**, 1915 (2001).
2. G. T. De Sanctis *et al.*, *Am. J. Physiol.* **277**, L1118 (1999).
3. Y. Zhang *et al.*, *Hum. Mol. Genet.* **8**, 601 (1999).
4. S. L. Ewart *et al.*, *Am. J. Respir. Cell Mol. Biol.* **232**, 537 (2000).

20 September 2001; accepted 3 December 2001