Genet. Res., Camb. (2004), **84**, *pp.* 103–108. *With 5 figures.* © 2004 Cambridge University Press
DOI: 10.1017/S0016672304007086 Printed in the United Kingdom

Power calculations for selective genotyping in QTL mapping in backcross mice

NUSRAT RABBEE[1,*], DAVID SPECA[2], NICOLA J. ARMSTRONG[3] AND TERENCE P. SPEED[1,4]

[1] *Department of Statistics, University of California, Berkeley, 367 Evans Hall, Berkeley, CA 94720, USA*
[2] *Department of Neurology and the Ernest Gallo Clinic and Research Center, University of California at San Francisco, 5858 Horton Street, Emeryville, CA 94608, USA*
[3] *Eurandom, Eindhoven, The Netherlands*
[4] *Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia*

(*Received 9 October 2003 and in revised form 29 March and 2 August 2004*)

Summary

Selective genotyping concerns the genotyping of a portion of individuals chosen on the basis of their phenotypic values. Often individuals are selected for genotyping from the high and low extremes of the phenotypic distribution. This procedure yields savings in cost and time by decreasing the total number of individuals genotyped. Previous work by Darvasi *et al.* (1993) has shown that the power to detect a QTL by genotyping 40–50% of a population is roughly equivalent to genotyping the entire sample. However, these power studies have not accounted for different strategies of analysing the data when phenotypes of individuals in the middle are excluded, nor have they investigated the genome-wide type I error rate under these different strategies or different selection percentages. Further, these simulation studies have not considered markers over the entire genome. In this paper, we present simulation studies of power for the maximum likelihood approach to QTL mapping by Lander & Botstein (1989) in the context of selective genotyping. We calculate the power of selectively genotyping the individuals from the middle of the phenotypic distribution when performing QTL mapping over the whole mouse genome.

1. Introduction

Experimental crosses are used routinely to map the genetic loci (QTL) contributing to variation in quantitative trait values. The objective is to identify regions in the genome where there exists an association between the phenotype and marker genotypes. *Selective genotyping* is a strategy in which only a portion of individuals are genotyped; the subset is chosen on the basis of the individuals' phenotypes (generally those with extremely high or low phenotypic values). The advantage of this is that fewer individuals can be genotyped while hardly changing the probability of identifying putative QTLs. Selective genotyping yields significant savings since genotyping is expensive.

Lander & Botstein (1989) used expected LOD scores to make a strong case for selective genotyping and recommended that the phenotypes of the ungenotyped individuals be included as missing values in order to prevent bias in parameter estimation. Power under selective genotyping for a *genome-wide* scan in QTL mapping experiments with backcrosses has not been studied. This is the typical context in all mapping experiments, including ENU mutagenesis (described below). Further, the type I error rate under the different strategies of selective genotyping has not been investigated in this context. ENU mutagenesis is a *phenotype driven* or *forward* mutagenesis approach which is performed for identifying new genes and new roles for known genes associated with disease phenotypes. Mutations are randomly introduced into an organism's genome, and then its progeny are tested for specific defects in a biological process. Studying the mouse mutants reveals the underlying gene responsible for the disease phenotype (Brown & Hardisty, 2003). The alkylating agent

* Corresponding author. Tel: +1 (510) 6429959. Fax: +1 (510) 6427892. e-mail: nrabbee@stat.berkeley.edu

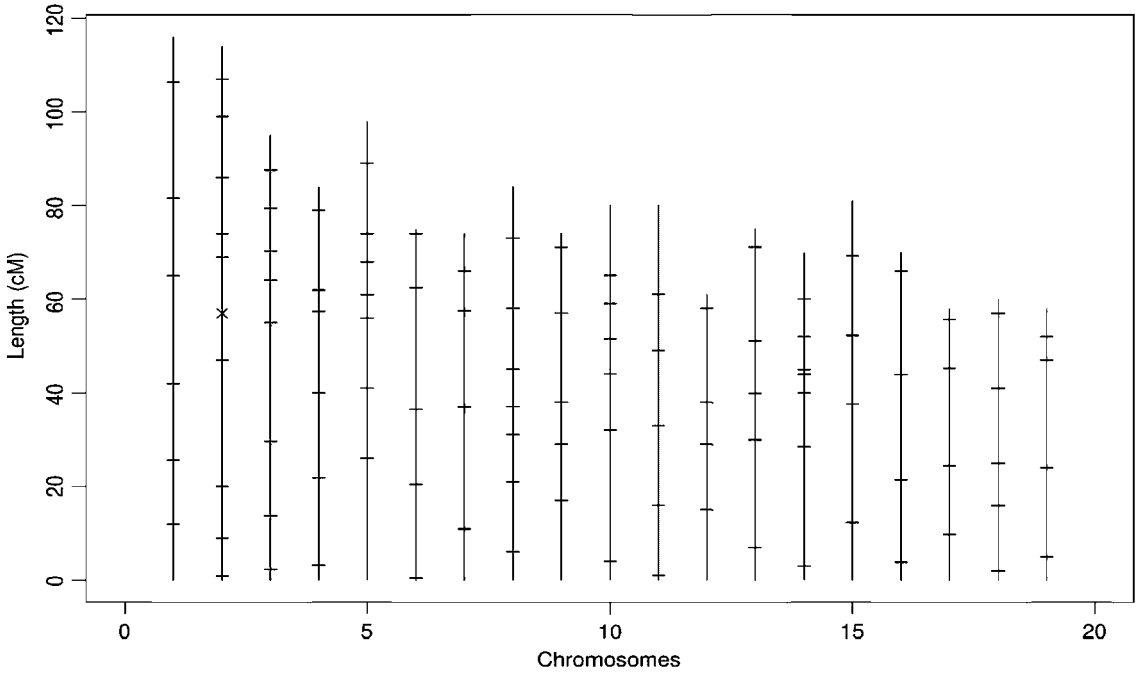

Fig. 1. Marker placement in the mouse genome; 105 markers.

ethylnitrosurea (ENU) is currently the most powerful mutagen for the production of mutants in mice. In mutagenesis experiments, male mice are injected with ENU and then mated to females in order to produce offspring which carry ENU mutations.

In this paper, we present a standard power analysis under *selective genotyping* for (a) performing genome-wide scans with sample sizes and marker spacings typically present in ENU mutagenesis, (b) exploring type I error rates, LOD thresholds and power under different selection strategies and different percentages of selective genotyping, (c) contrasting power obtained under maximum likelihood (ML) methods and ANOVA or regression at markers, and (d) providing software for investigators in order to calculate sample size with relevant experimental parameters. To our knowledge existing power studies do not address the concerns listed above. Darvasi & Soller (1992) presented power calculations for selective genotyping in a *single marker-QTL linkage* setting when no recombination is present.

Most mapping experiments today involve multiple markers. Subsequently, Darvasi *et al.* (1993) developed simulation mechanisms for investigating power when infinitely dense markers are present on a *single chromosome*. This investigation, however, did not address selective genotyping. Next, assuming a single QTL located at the centre of a chromosome with infinitely dense markers, Darvasi (1997) used simulation to investigate the effects of selective genotyping. Again, the context of a *genome-wide scan*, small sample sizes and different strategies of selective genotyping were not considered.

In the context of rapid genotyping in ENU mutagenesis experiments, the investigator typically breeds about $n=100$ to $n=150$ offspring. The individuals with phenotypes in the middle of the distribution are excluded from genotyping. Only the individuals at the extremes are genotyped to reduce cost. Each individual is genotyped with moderately spaced markers, say 100–150, over the entire mouse genome. After the markers are genotyped for each individual animal, the data are analyzed by standard QTL mapping software, such as Mapmaker (Lander *et al.*, 1987) or R/qtl (Broman *et al.*, 2003), searching for a QTL-marker linkage somewhere on the genome.

We begin by introducing the simulation we use for investigating power under selective genotyping in backcross experiments and mention the methods for analysis in Section 2. In Section 3, we present our results. Finally, in Section 4, we discuss our findings.

2. Design of simulations

We simulated the genomes of a backcross population of mice for 19 chromosomes and 105 marker assignments under a chi-square model of the recombination process (Zhao *et al.*, 1995; Armstrong, 2001). The marker placements are shown in Fig. 1 and were based on an unpublished study at the Walter and Eliza Hall Institute of Medical Research.

The QTL of interest was placed on chromosome 2. Markers flanking the QTL position on chromosome 2 are separated by 22 cM. The phenotypes were generated according to the mixture densities for the QTL

https://doi.org/10.1017/S0016672304007086 Published online by Cambridge University Press

with the means of the two groups, μ_{aa} and μ_{bb}, separated by the gene effect, δ.

If we denote the percent genotyped as p, then $p/2$ represents the percent in each symmetric tail of the phenotypic distribution. We compared three strategies for analysing data under selective genotyping that are currently used in real biological assays:

Strategy A: *All* genotypes and phenotypes are considered for the n individuals.
Strategy B: *All* phenotypes are considered for the n individuals but *only* the genotypes of individuals in the tails (p) of the phenotypic distribution are selected; the genotypes for those in the middle are entered as missing.
Strategy C: *Only* the phenotypes and the genotypes at the tails (p) are selected; the genotypes and the phenotypes for those in the middle are entered as missing.

To our knowledge, many investigators are performing QTL mapping with selective genotyping under strategy C and we should carefully compare power and type I error rate for this method relative to others.

We primarily use the method of *interval mapping* (IM) using flanking markers, proposed by Lander & Botstein (1989), for our simulation study of single QTL segregation in backcrosses. Under interval mapping, for each locus in the genome, Lander & Botstein (1989) proposed using the EM algorithm (Dempster *et al*., 1977) under a mixture model formulation to obtain the ML estimates for QTL genotype effects and variance parameters.

Under selective genotyping, we feel it is best to focus on the ML methods for QTL mapping, since analysis of variance (ANOVA) at each marker would (a) exclude the individuals with missing marker genotypes and yield biased estimates of the parameters (Lander & Botstein, 1989) under selective genotyping, (b) give less precise information about QTL location and (c) suffer in low- to moderate-density scans. ML methods based on mixture models would give more precise estimates of QTL effects, the QTL location, as well as properly accounting for selective genotyping using missing data mechanisms for parameter estimation. Since we are exploring ENU mutagenesis assays which use selective genotyping over the entire genome with the aim of mapping the gene, we would most probably use the interval mapping ML method to interpolate between markers and to accommodate selective genotyping when analysing the data. Therefore, power studies must be based on methods which will eventually be used to analyse the data after the experiment is complete. Other interval mapping methods, such as Haley & Knott (1992) regression mapping or Kearsey & Hyne (1994) marker regression, yield near-identical LOD curves, and therefore power, to the interval mapping ML method. But again, the interval mapping ML method would yield unbiased estimates for QTL effects and position under selective genotyping since it properly handles incomplete data.

Using interval mapping ML, we performed a genome-wide scan in R/qtl to (a) obtain appropriate LOD thresholds and (b) map or identify the QTL correctly under the three strategies. Note that we are defining power as the probability of the LOD score exceeding the threshold in the region between flanking markers of the putative QTL. In this regard, this is not a traditional definition of power which rejects the null hypothesis of no QTLs anywhere on the genome. We chose this definition of power to reflect the reality of gene mapping experiments designed with the aim of detecting a putative QTL.

For (a) we identified proper cutoffs for all three interval mapping strategies, as well as for ANOVA at each marker, by considering 10 000 replicates of genome-wide marker data for all 100 individuals under the null hypothesis of no QTLs. We then recorded the maximum LOD scores obtained anywhere over the genome for each replicate. For the ANOVA method, we recorded the maximum F-test value obtained over all the markers. We computed the 95th percentile of these maximum scores. We repeated this procedure for different selection percentages (p).

For (b) we performed 5000 simulations to generate power curves for detecting QTLs using the cutoffs obtained above. Here we used the criterion that the maximum LOD score in the region between markers flanking the QTL exceeded the appropriate threshold under the different strategies. When comparing results with the ANOVA method, we checked whether the F-test statistic at either the left or the right flanking marker exceeded the threshold obtained from simulations.

For our simulations we considered $p = (10, 20, \ldots, 90)$ as selection percentages with $p = 100\%$ reducing to strategy A above. For the gene effect, we present simulations for $\delta = (1, 2, \ldots, 5)$ with $\delta = 0$ reducing to estimate the type I error rate.

3. Results

Fig. 2 clearly shows that, for $p = 50\%$, the LOD thresholds needed to ensure that the rate of false positives is under 0·05 varies for the different strategies under selective genotyping. Lander & Kruglyak (1995) suggested the LOD threshold of approximately 2·8 for a genome-wide scan with moderately spaced markers, as in our simulation setting. However, we see that if QTL mapping is performed excluding the middle 50% phenotypes and genotypes of all the individuals (strategy C), then a threshold of 3·1 is appropriate for $n = 100$.

Furthermore, we found that the type I error rate varies also with the percent genotyped (p). Fig. 3

https://doi.org/10.1017/S0016672304007086 Published online by Cambridge University Press

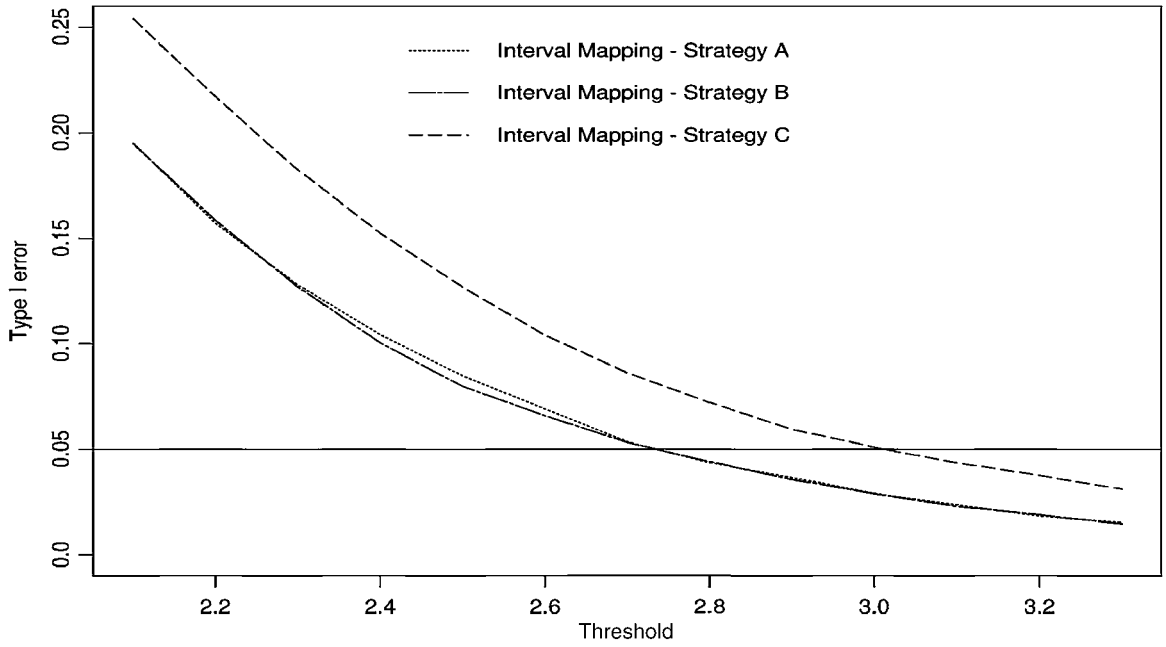

Fig. 2. Genome-wide type I errors for given LOD thresholds: selective genotyping under three different strategies; mean = 8, $\sigma = 0{\cdot}32$, $n = 100$, $p = 50\,\%$.

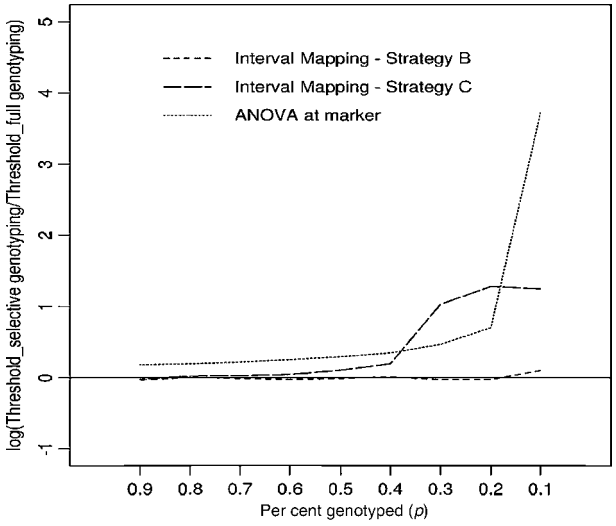

Fig. 3. Log (threshold ratio) versus percent genotyped (p): selective genotyping under different strategies; mean = 8, $\sigma = 0{\cdot}32$, $n = 100$.

shows the log ratio of the selective genotyping threshold to the full genotyping threshold for interval mapping (strategies B and C) and ANOVA at marker locations, under different values for p. When less than 50 % are genotyped at the tails, the type I error inflation rate is very high for interval mapping (strategy C) and the ANOVA method. Therefore, if either interval mapping strategy C or the ANOVA method is used to analyse the data, then the appropriate LOD threshold for the given p value must be used.

The results for estimating power as a function of the QTL effect (δ) are displayed in Fig. 4, with the middle 50 % not genotyped, for interval mapping strategies A, B and C. Since the curves are very close to each other in this figure, we note that the loss of power is negligible for all the three selection strategies under interval mapping, after proper calibration with thresholds.

Fig. 5 provides an example of how percent genotyped (p) affects power under the three selection strategies under interval mapping, as well as under the ANOVA method, for $\delta = 0{\cdot}25$, $\sigma = 0{\cdot}32$, $n = 100$. Here power under strategy A holds steady at about 0·44; however, power under strategies B and C starts to decline when $p < 60\,\%$. After proper calibration with thresholds, the two strategies yield similar estimates of power, with strategy B giving slightly higher power when the percent genotyped (p) is low. Fig. 5 also shows that power is lowest when selective genotyping is carried out using the ANOVA method.

Table 1 explores power loss for different QTL effects (δ) and the percent genotyped (p). Again we see that the loss of power is negligible for selective genotyping (strategy A vs strategy B) when percent genotyped is more than 50 %, for all values of δ. However, power rapidly decreases when p falls below 50 %. Finally, with less than 50 % genotyped, increasing values of δ give rise to power loss.

4. Discussion

In this paper, we have presented a study of the power to detect a single QTL with selective genotyping for backcross mice using simulated genome-wide scans. We have used the interval mapping methodology,

https://doi.org/10.1017/S0016672304007086 Published online by Cambridge University Press

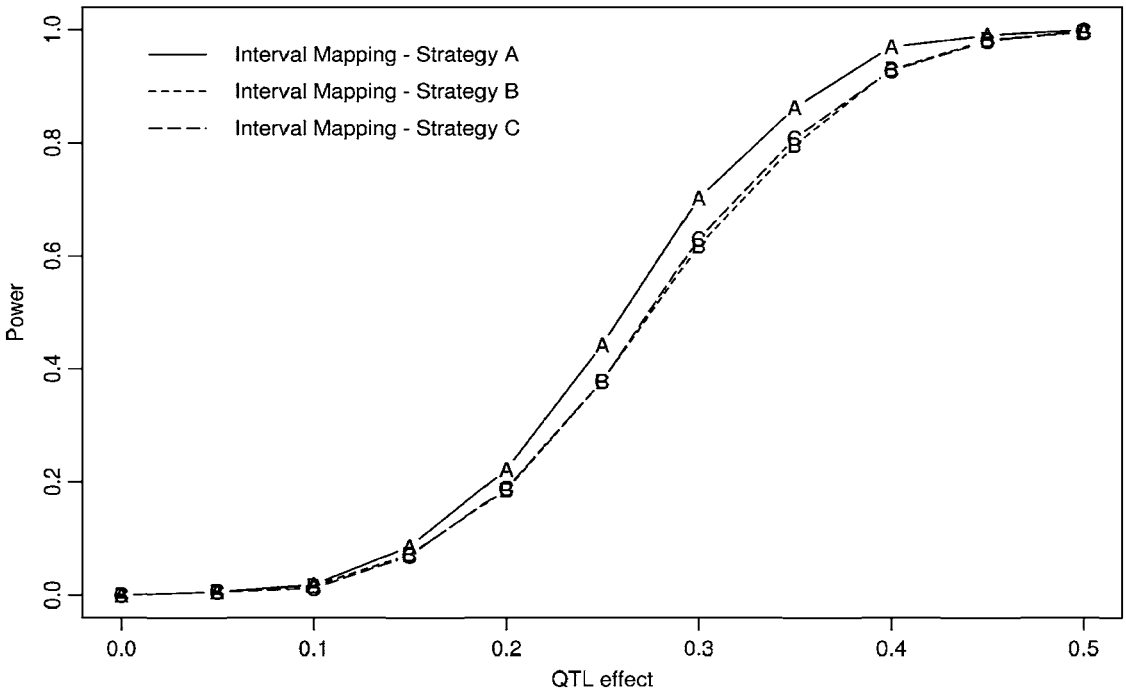

Fig. 4. Power for a given QTL effect (δ) under two different strategies; $\sigma = 0{\cdot}32$, $n = 100$, not genotyped = middle 50%.

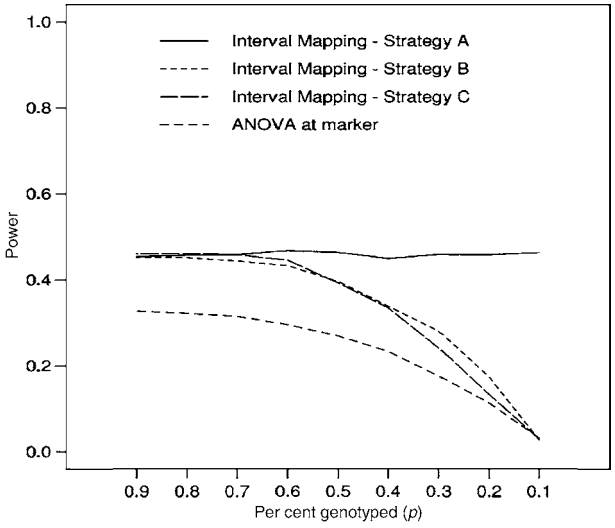

Fig. 5. Power for a given percent genotyped (p): two different strategies; mean = 8, $\sigma = 0{\cdot}32$, $n = 100$, $\delta = 0{\cdot}25$.

Table 1. *Percent loss in power: by QTL effect* (*row*) *and percent genotyped* (*columns*); *mean* = 8, $\sigma = 0{\cdot}32$, $n = 100$

	10	*20*	*30*	*40*	*50*	*60*	*70*	*80*	*90*
0.10	81	58	39	16	15	8	5	1	0
0.15	90	59	35	21	17	6	4	2	0
0.20	92	65	46	24	15	9	5	2	0
0.25	93	62	39	25	14	7	3	1	0
0.30	92	56	34	19	11	5	3	1	0

introduced by Lander & Botstein (1989), which is an improvement over previous approaches of ANOVA at each marker location, to map the putative QTL. We compared power estimates obtained for different selection strategies under interval mapping to those under the ANOVA method. We investigated the type I error rate and power under different selection percentages and different selection strategies, in order to suggest specific guidelines for performing backcross experiments.

We have presented simulations that attempt to capture the usual realities of ENU mutagenesis assays with backcross mice. Fig. 1 shows the placement of the 105 markers. There are five key points that we wish to highlight:

Type I error rate varies with different selection strategies, as well as with different selection percentages. The strategy of excluding the phenotypes and genotypes of the individuals in the middle (strategy C) should only be used with the appropriate threshold for obtaining a genome-wide type I error rate of 0·05. With 50% genotyped at the two tails, the type I error rate is 0·075 for strategy C. While this is not a hugely significant difference from the acceptable rate of 0·05, Fig. 2 shows that the type I error rate is consistently higher for strategy C relative to strategy B or A, for a given p, for a range of threshold values. For $p < 50\%$, type I error rate is even higher for strategy C. This is an important finding of our analysis.

https://doi.org/10.1017/S0016672304007086 Published online by Cambridge University Press

Fig. 3 shows the log ratio of selective genotyping threshold to full genotyping threshold for the interval mapping method using both strategies B and C, as well as for the ANOVA method of analysis, under different selection percentages (p). Both the ANOVA method and interval mapping method (strategy C) need to be calibrated with thresholds quite different from the complete data threshold as p decreases, for experimental parameters similar to our simulation setting. The complete data threshold, however, is fine for selective genotyping under strategy B, where the phenotypes of all the individuals are included in the analysis, for different selection percentages.

Power under the ANOVA method is the lowest when compared with the power under the ML method of interval mapping for all strategies and for all selection percentages (see Fig. 5). This finding differs from the suggestion by Lynch & Walsh (1998) that t-test guidelines are also reasonable for ML interval mapping. We feel this is another important finding of our power investigation.

The **percent genotyped** (p) plays an important role in power loss under selective genotyping. Fig. 5 shows that with $p \geqslant 50\,\%$ and $\delta = 0{\cdot}25$, $\sigma = 0{\cdot}32$, power loss is negligible even for $n = 100$. Thus there is no need to increase n in this setting, since for ENU mutagenesis assays, it is often quite infeasible to increase sample size to 500 or to 1000. However, when p falls below 50 %, power begins to decrease rapidly. This is the lower limit for percent genotyped for minimally acceptable power loss within our framework.

Gene effect (δ) also determines power (see Fig. 4). Power loss depends on both the magnitude of the gene effect, δ, and the percent genotyped, p. For example, Table 1 demonstrates that with 40 % genotyped and for a small gene effect, $\delta = 0{\cdot}15$, the power loss could be 21 % for $n = 100$. We do not recommend that experimenters genotype less than 50 % in parameter settings similar to ours. Finally, for **increasing** n, the percent power loss diminishes, as expected. For example, when $\delta = 0{\cdot}25$, $\sigma = 0{\cdot}32$, $n = 150$, with 50 % genotyped, the percent power loss is 6 %, as opposed to 14 % with $n = 100$, after resetting the threshold. This indicates that the power loss under selective genotyping decreases with increasing sample size, all other parameters being held fixed.

The software for power calculations presented in this paper is implemented in R/qtl and C and is available from the first author.

We thank Karl Broman for his counsel and feedback in this study. We thank the reviewers for their thoughtful comments on the manuscript.

References

Armstrong, N. (2001). Incorporating interference into the linkage analysis of experimental crosses. PhD dissertation, Department of Statistics, University of California, Berkeley.

Broman, K., Wu, H., Sen, S. & Churchill, G. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890.

Brown, S. & Hardisty, R. (2003). Mutagenesis strategies for identifying novel loci associated with disease phenotypes. *Seminars in Cell and Developmental Biology* **14**, 19–24.

Darvasi, A. (1997). The effect of selective genotyping on QTL mapping accuracy. *Mammalian Genome* **8**, 67–68.

Darvasi, A. & Soller, M. (1992). Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. *Theoretical and Applied Genetics*. **85**, 353–359.

Darvasi, A., Weinreb, A., Minke, V., Weller, J. & Soller, M. (1993). Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic-map. *Genetics* **134**, 943–951.

Dempster, A., Laird, N. & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* **39**, 1–38.

Haley, C. S. & Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315–324.

Kearsey, M. J. & Hyne, V. (1994). QTL analysis: a simple 'marker-regression' approach. *Theoretical and Applied Genetics* **89**, 698–702.

Lander, E. S. & Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185–199.

Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E. & Newburg, L. (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174–181.

Lander, E. S. & Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* **11**, 241–247.

Lynch, M. & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*, pp. 444–447. Sunderland, MA: Sinauer Associates.

Zhao, H., Speed, T. & McPeek, M. (1995). Statistical analysis of crossover interference using the chi-square model. *Genetics* **139**, 1045–1056.