## Type I error rates for MIVQUE methods

We performed limited simulations to explore the Type I error rate of the nested ANOVA analysis with MIVQUE variance estimation method developed in Xu and Garland (2017). Simulations were performed by Xu and Garland (2017) using a β distribution with average allele frequency of 0.4 and 0.6 for HR and C linetypes, respectively, producing a power of 0.5541. Several data sets were constructed under the null hypothesis of no differentiation between the four High Runner lines and the four non-selected Control lines, then analyzed using the MIVQUE method. Each locus has two alleles randomly assigned as 0 or 1 using the binary distribution. Datasets were constructed with 3 mice, 10 mice or 20 mice in each of the 8 lines. Additional simulations used 10 mice per line but only two replicate C lines and two replicate HR lines. For  = 0.05, all Type I error rates were substantially lower than expected, ranging from 0 to 0.0248 (Table S8). Given these results, it seems clear that the MIVQUE method may be underpowered in an absolute sense, even though previous simulations indicate that it has higher power than the regularized *F* test or a GLMM (Xu and Garland 2017). The Type I error rate did not vary much based on the number of mice genotyped per line, but more on the number of replicate lines. The probable explanation for this is related to degrees of freedom. As the line is the experimental unit, analyses based on nested ANOVAs with line as a random effect nested within linetype are conducted with d.f. = 1 and 6. However, when random effects are associated with little or no variance (e.g., covariance parameter estimates for the line effect are low or zero in SAS Proc Mixed), it may be more appropriate to recoup d.f. with a method such as Kenward-Rogers. Preliminary analyses of simulated data indicated that doing so increased the Type I error rate for  = 0.05, but not up to 5%. Moreover, the optimal procedure for recouping d.f. in such cases is somewhat controversial, and any attempt to do this violates the fundamental experimental design.

**Table S8 MIVQUE simulations under the null hypothesis**

|  |  |  |  |
| --- | --- | --- | --- |
| **Number of Simulated Loci** | **Mice Per Line** | **Lines (half C, half HR)** | **Type I Error Rate** |
| 23,000 | 10 | 8 | 0.011261 |
| 20,000 | 3 | 8 | 0.01225 |
| 20,000 | 10 | 4 | 0 |
| 20,000 | 10 | 16 | 0.0248 |
| 20,000 | 20 | 8 | 0.0112 |

Given the inherent low Type I error rate of this model design, permutations are an essential mechanism to determine critical thresholds and simultaneously correct for multiple testing. As demonstrated by Xu and Garland (2017), with the inclusion of permutations for multiple testing correction, the nested ANOVA proved more powerful than other conventional methods. Furthermore, the local maxima used in the present study, being chosen from regions with initially less stringent p-value cutoffs (P<0.001), are also robust to the deflated type I error because each LM need only be more significant than the surrounding loci.