vqtl: An R package for Mean-Variance QTL Mapping

Robert W. Corty* **and William Valdar***,[1]

*Department of Genetics, University of North Carolina at Chapel Hill

ABSTRACT Most existing methods for QTL mapping in experimental crosses assume that the residual variance is constant across all individuals. But common situations violate this assumption. For many phenotypes, one sex is more variable than the other, some experimenters make more precise measurements than others, and specific genetic factors influence environmental sensitivity. In these cases, mean-variance QTL mapping provides higher power and better protection against false positives. It also allows for detection of QTL that influence phenotype variance, termed vQTL.

We present R package vqtl. This package makes it easy for geneticists to apply the mean-variance QTL mapping approach, control family-wide error rate (FWER), and visualize and interpret their results. Because this package is interoperable with the popular R/qtl package and uses many of the same data structures and input patterns, it will be easy for geneticists to analyze future experiments with R/vqtl as well as re-analyze past experiments, possibly discovering new QTL.

KEYWORDS

QTL mapping, variance heterogeneity

QTL mapping studies have provided important insights on nearly every trait of interest in human health and disease. Advances in breeding, phenotyping (Yang *et al.* 2014), and genotyping (Williams *et al.* 1990) model organisms as well as statistical methods (Lander and Botstein 1989) and software tools (Broman *et al.* 2003) have supported these discoveries.

One common assumption in the design and analysis of these studies is that, across all organisms in a study population, the residual variance is constant. Said another way, it is typically assumed that nothing – neither environmental factors nor genetic factors – influences the residual variance of the trait. In the companion piece [Corty2017a], we describe the mean-variance QTL mapping approach which allows for QTL mapping without the constant variance assumption, compare its behavior to other QTL mapping approaches, and illustrate the discovery of a new QTL from an existing study.

In support of other researchers who may be interested in mvQTL mapping, we created R package vqtl, which provides functions for conducting genome scans, assessing the statistical significance of results, and visualizing and interpreting significant findings. R package vqtl uses the same cross data structure as the popular qtl package and is available on CRAN, so it is easy to get started.

Here, we demonstrate typical use of the vqtl package. The code used to simulate the phenotypes, calculate the statistics, estimate their significance, and visualize significant results is available at github.com/rcorty.

Copyright © 2017 Robert Corty *et al.*

Manuscript compiled: Sunday 7th May, 2017

[1]Correspondence e-mail: william.valdar@unc.edu

SIMULATION OF AN ILLUSTRATIVE DATASET

We used the popular R/qtl package to simulate an example experimental cross. This cross consisted of 200 male and 200 female F2 offspring, with 3 chromosomes of length 100 cM, each tagged by 11 equally-spaced markers and estimated genotype probabilities at 2cM intervals with R/qtl's hidden Markov model.

We simulated four phenotypes:

1. phenotype1 consists only of random noise and will serve as an example of negative results for all tests.

2. phenotype2 is influenced by the 6th (middle) marker on chromosome one. The marker influences the mean of the phenotype, but not the variance, so it will serve as an example of a pure "mQTL".

3. phenotype3 is influenced by the 6th (middle) marker on chromosome two. The marker influences the variance of the phenotype, but not the mean, so it will serve as an example of a pure "vQTL".

4. phenotype4 is influenced by the 6th (middle) marker on chromosome three. The marker influences both the mean and the

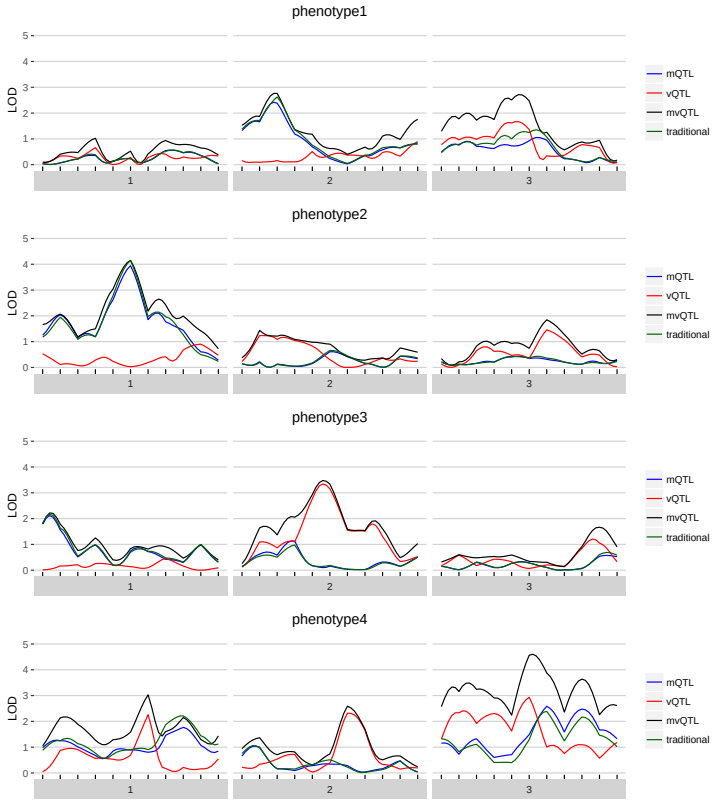

Figure 1 For each of the four simulated phenotypes, the genome scan shows the LOD score of each test – mean, variance, and joint – in blue, red, and black, respectively. The traditional test is in green and globally similar to the mean test.

variance of the phenotype, so it will serve as an example of a mixed "mvQTL".

We additionally consider `phenotype1x` through `phenotype4x`, which have the same type of genetic effects as `phenotype1` through `phenotype4`, but additionally have covariate effects on phenotype variance. All the same analyses and plots that are shown for `phenotype1` through `phenotype4` are shown for `phenotype1x` through `phenotype4x` in the appendix.

GENOME SCANS

The central function for genetic mapping in package `qtl` is `scanone` (Broman *et al.* 2003). Analogously, the central function for genetic mapping in package `mvqtl` is `scanonevar`. It takes three required inputs:

1. `cross` contains the genetic and phenotypic information from an exerimental cross. This object can be the same `cross` object used in package `qtl`.

2. `mean.formula` specifies the phenotype to be mapped, the covariates to be corrected for, and the QTL terms to be fitted (`additive` and `dominance` components by default). The `mean.formula` uses the standard `R` formula notation.

3. `var.formula` specifies the covariates to be corrected for as well as the QTL terms to be fitted (`additive` and `dominance` components by default) in modeling the residual variance. The `var.formula` also uses the standard `R` formula notation.

Optional argument `chrs` is used to specify a subset of chromosomes to be scanned, defaulting to all chromosomes. Optional argument `return.covar.effects` is used to specify whether or not fitted effects of all covariates should be returned as part of the scan result, defaulting to `FALSE`.

Unlike `scanone`, which only tests for association between each locus and the phenotype mean, `scanonevar` computes three tests for each locus – association with phenotype mean, association with phenotype variance, and joint association with phenotype mean and variance. The statistic for each of these associations is a LOD score, the log of the ratio of the likelihood of the alternative model to the null model. The results of `scanonevar` on each of the four described phenotypes is shown in figure 1. The details of the null and alternative models used in each of the three tests can be found in the companion article [Corty2017a, Corty2017b].

The LOD Score – Problems and an Alternative

The traditional test statistic in QTL mapping is the LOD score, but its interpretation can be difficult for two reasons: (1) LOD scores must be interpreted differently in autosomes and sex chromosomes. The sex chromosomes are typically fit with fewer parameters, depressing the expected value of the LOD scores under the null, and thus increasing the significance of any observed LOD score. (2) LOD scores from different tests must be interpreted differently. For example, the LOD score of the mvQTL test is always higher than the LOD score of the mQTL and vQTL tests. This relationship is due to the nested nature of the mvQTL test and the other two tests. All three tests use the same alternative model, but the null model in the mvQTL test imposes all the constraints of the null of the mQTL test and all the constraints of the null of the vQTL test, so the LOD score of the mvQTL test must, by definition, be greater than the LOD score of the other tests.

These difficulties in interpreting the LOD score are evident in figure 1. It seems that there are no important signials in the genome scan of `phenotype1` and it is visually clear that the most interesting signals for `phenotype2`, `phenotype3`, and `phenotype4` are on chromosomes one two and three, respectively. But important questions remain: (1) If we didn't have the genome scans from `phenotype2` - `phenotype4` available for comparison, would we be confident there are no statistically significant signals related to `phenotype1`? (2) How can we compare the results of the mvQTL test to the results of the other tests? (3) How could we compare the results of tests on autosomes to tests on sex chromosomes (if they were present)? (4) How often do we expect to observe results of this magnitude or greater when there is no true association, due simply to sampling variation and the multiplicity of tests conducted?

To put all genetic loci and all three tests on a level playing field, we consider two types of p-values: (1) Asymptotic p-values are calculated by `scanonevar` using the χ^2 distribution with the appropriate degrees of freedom for each locus. Though these p-values overcome the questions 1 through 3 of working with LOD scores described above, they leave problem 4 unresolved. (2) Empirical, family-wide error rate (FWER)-corrected p-values resolve all the above issues with interpreting LOD scores, thus they are the recommended means of assessing the significance of QTL mapping results. Methods for calculating them are described in the next section.

The object returned by the `scanonevar` function has class `scanonevar`. Calling `plot` on this object produces a publication-quality plot that shows the three association statistics at each locus. Calling `summary` on this object produces a summary of how the scan was conducted and what the results were. With this example

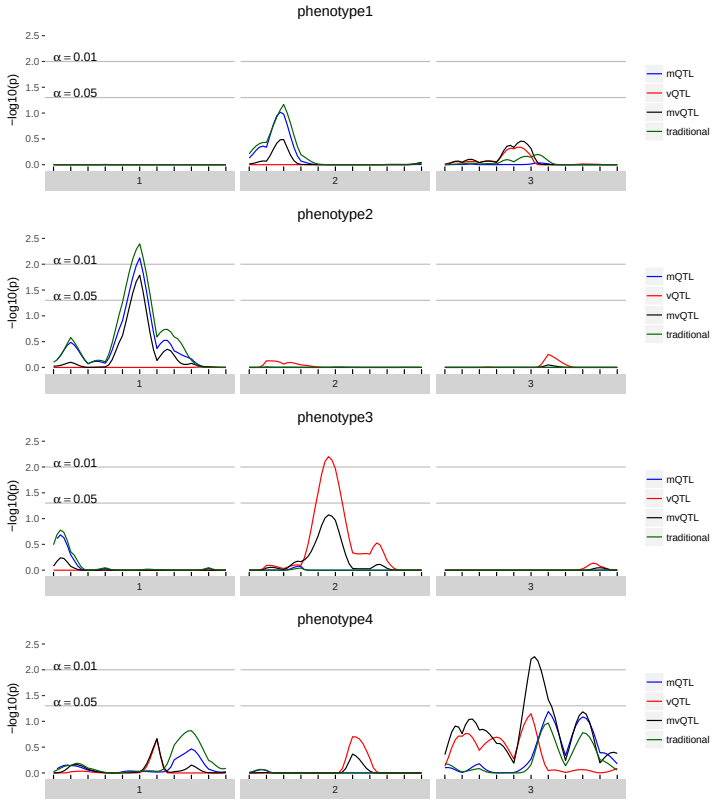

Figure 2 For each of the four simulated phenotypes, the genome scan shows the -log10 of the FWER-corrected *p*-value of each test – mean, variance, and joint – in blue, red, and black, respectively. Thus, a value of 3 implies that the quantity of evidence against the null is such that we expect to see this much or more evidence once per thousand genome scans when there is no true effect.

dataset, it takes five seconds to run one genome scan on a Intel Core i5.

ASSESS THE SIGNIFICANCE OF RESULTS

The *p*-value assigned to each locus by calculating the position of its calculated LOD score in its null distribution answers the question, "How probable is it to observe this much of a deviation from the null at this locus, given that there is no true effect?" This question is not entirely relevant for genetic mapping, however, because we typically test for association at many loci, with no special interest in any individual locus *a priori*. Thus, a more appropriate *p*-value would answer the question, "How probable is it to observe this much or more deviation from the null at *some* locus in the genome, given that there are no true effects?" This is precisely the reasoning behind the family-wide-error-rate (FWER) controlling procedure we illustrate below.

To calculate a FWER-controlling *p*-value, one typically must estimate the effective number of statistical tests conducted. The effective number of tests in a genome scan, however, is difficult to estimate. One lower bound is the number of chromosomes. Due to the randomization in meiosis no two non-syntenic loci are correlated in an experimental cross and therefore tests on different chromosomes are always independent. But there are many non-identical tests conducted on each chromosome, so the number of chromosomes is an under-estimate. One upper bound on the effective number of tests is the total number of loci. But, loci on the same chromosome are often in linkage disequilibrium, therefore the tests are not independent and the total number of loci is an over-estimate (Lander and Botstein 1989).

Like previous work on permutation-based thresholds for genetic mapping (Churchill and Doerge 1994; Carlborg and Andersson 2002), our permutation approach sidesteps estimation of the effective number of tests. We conduct many permuted genomes scans, each executing a carefully-constructed model comparison to maintain all mean and variance effects of covariates and any non-focal genetic effects on mean and variance. The details of these model comparisons are provided in the companion article [Corty2017a]. For each test, we extract the highest observed value of the test statistic from each permutation scan and use those to model a generalized extreme value (GEV) density (Stephenson 2002). The observed LOD scores from the genome scan are then transformed by the cumulative distribution function of the extreme value density to estimate the FWER-controlling *p*-values. This approach is implemented in the function, `scanonevar.perm`, which requires two inputs:

1. `sov` is the `scanonevar` object, the statistical significance of which will be assessed through permutation.

2. `n.perms` is the number of permutations to conduct.

The object returned by `scanonevar.perm` is a `scanonevar` object with one important additional piece of information, an empirical *p*-value for each test at each locus. These *p*-values are FWER-corrected, so a value of 0.05 for a specific test at a specific locus implies that in 5% of similar experiments where there is no true genotype-phenotype association, we would expect to observe *some* locus with this much or more evidence of association. Additionally, the returned object contains a list of the per-genome-scan maximum observed LOD for each test and each chromosome type.

Accurate estimation of the FWER-controlled *p*-values requires many permutation scans. We recommend at least 100, and rarely more than 1000 (Churchill and Doerge 1994; Carlborg and Andersson 2002). These permutation scans can be run on multiple processors by specifying the optional `n.cores` argument. On an Intel Core i5, running 100 permutations on this dataset packtakes about five minutes. When many phenotypes are studied, or if faster runtimes are needed, these permutation scans can be broken into groups with different values for `random.seed`, run on separate computers, and combined with the `c` function. This function combines the permutations from all the inputted scans, re-evaluates the observed LOD scores in the context of all available permutations, and returns a new `scanonevar` object with more precisely estimated empirical *p*-values.

INVESTIGATE SIGNIFICANT FINDINGS

Having identified some QTL, we want to visualize the allele effects at those loci. Because the `vqtl` package models both mean and variance effects, existing plotting utilities aren't able to display the entirety of the modeling results. To investigate the results of a `vqtl` scan at one particular locus, we developed the `mean_var_plot`. This plot shows information about the phenotype mean on the horizontal axis and information about the phenotype variance on the vertical axis. There are both "model-free" and "model-based" versions of this plotting utility. Here we show only the "model-based" version.

Figure 3 shows a model-based `mean_var_plot` for each of the four phenotypes. In each plot, the location of the dot shows the estimated mean and standard deviation of each genotype group, with

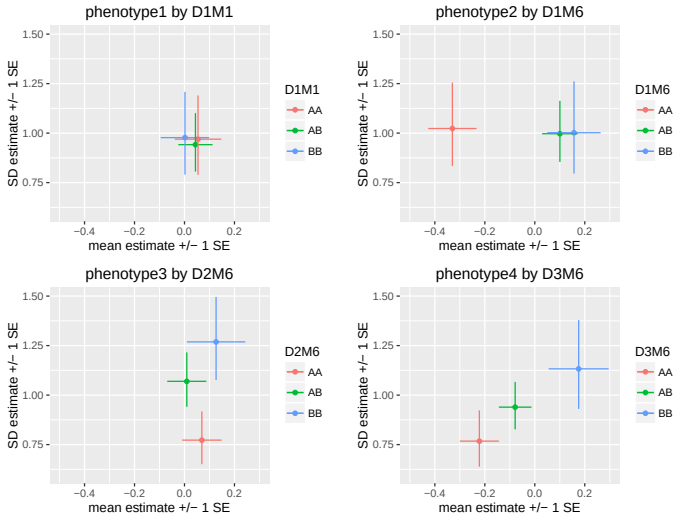

Figure 3 `mean_var_plots` show the estimated genotype effects at a locus, with mean effects on the horizontal axis and variance effects on the vertical axis. Horizontal lines indicate standard errors for mean effects and vertical lines indicate standard errors for variance effects. More information on the interpretation of these plots is provided in the 'Investigate Significant Findings' section.

the mean indicated by the horizontal position and the standard deviation indicated by the vertical position. The horizontal lines extending to the left and right from each dot show the standard error of the mean estimate, and the vertical lines extending up and down from each dot show the standard error of the standard deviation estimate. There are two types of grouping factors considered by the function `mean_var_plot_model_based`: (1) `focal.groups` are groups that are modeled and the prediction for each group is plotted. For example, a genetic marker is the `focal.group` in each plot in figure 3; `D1M1` in the top left, `D1M6` in the top right, etc. (2) `nuisance.groups` are groups that are modeled, but then averaged over before plotting. When there are many grouping factors thought to play a role in determining the mean and variance of an individual's phenotype, such as sex, treatment, and batch, we recommend putting just one or two in `focal.groups` and the others in `nuisance.groups` for clarity, cycling through which are displayed to gain a thorough understanding of the factors that determine the mean and variance of the phenotype.

For `phenotype1`, the `mean_var_plot` is shown at the first marker of the first chromosome. The estimates of the genotype effects on phenotype mean and variance are within one standard error of each other. This pattern is consistent with the fact that there are no genetic effects, and the p-value was not statistically significant at any locus.

For `phenotype2`, the `mean_var_plot` is shown at the most significant marker, the sixth marker on the first chromosome. The estimates of the genotype effects differ in the horizontal axis, but not the vertical axis. This pattern is consistent with the fact that there is a genetic effect on phenotype mean but none on phenotype variance and with the highly significant p-value for the mQTL test, but non-significant p-value for the vQTL test.

For `phenotype3`, the `mean_var_plot` is shown at the most significant marker, the sixth marker on the second chromosome. The estimates of the genotype effects differ in the vertical axis, but on on the horizontal axis. This pattern is consistent with the fact that there is a genetic effect on phenotype variance but none on phenotype mean and with the highly significant p-value for the vQTL test but non-significant p-value for the mQTL test.

For `phenotype4`, the `mean_var_plot` is shown at the most significant marker, the sixth marker on the third chromosome. The estimates fo the genotype effects differ somewhat in both the horizontal and vertical axes, but not as much on the horizontal axis as `phenotype2` and not as much on the vertical axis as `phenotype3`. This pattern is consistent with the fact that there is a genetic effect on both phenotype mean and variance and with the highly significant p-value for the mvQTL test and the marginally significant p-values for the other two tests.

Additional plotting utilities, `phenotype_plot`, `effects_plot` and `mean_var_plot_model_free` are described in the online vignette for the `vqtl` package, available at `github.com/rcorty`.

CONCLUSION

We have demonstrated typical usage of the new `vqtl` `R` package for QTL mapping in experimental crosses. This package is most appropriate for crosses and phenotypes where nuisance covariates or genetic factors are known or suspected to influence phenotype variance. In the case of genetic factors, they can be mapped. In the case of nuisance covariates, they can be accommodated.

The central function of this package, `scanonevar`, carries out the initial genome scan. The permutation function, `scanonevar.perm`, executes a carefully-constructed set of permutations to empirically estimate a FWER-controlling p-value for each observed LOD score. A suite of plotting functions, e.g. the `mean_var_plot`, allows geneticists to evaluate the allelic and covariate effects that underlie detected signals.

LITERATURE CITED

Broman, K. W., H. Wu, Ś. Sen, and G. A. Churchill, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics **19**: 889–890.

Carlborg, O. and L. Andersson, 2002 Use of randomization testing to detect multiple epistatic QTLs. Genet. Res. **79**: 175–184.

Churchill, G. A. and R. W. Doerge, 1994 Empirical Threshold Values for Quantitative Trait Mapping. Genetics **138**: 963–971.

Lander, E. S. and S. Botstein, 1989 Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics **121**: 185.

Stephenson, A. G., 2002 evd: Extreme Value Distributions. R News **2**: 0.

Williams, J. G., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey, 1990 DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. **18**: 6531–6535.

Yang, W., Z. Guo, C. Huang, L. Duan, G. Chen, N. Jiang, W. Fang, H. Feng, W. Xie, X. Lian, G. Wang, Q. Luo, Q. Zhang, Q. Liu, and L. Xiong, 2014 Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. Nat. Commun. **5**: 5087.

APPENDIX

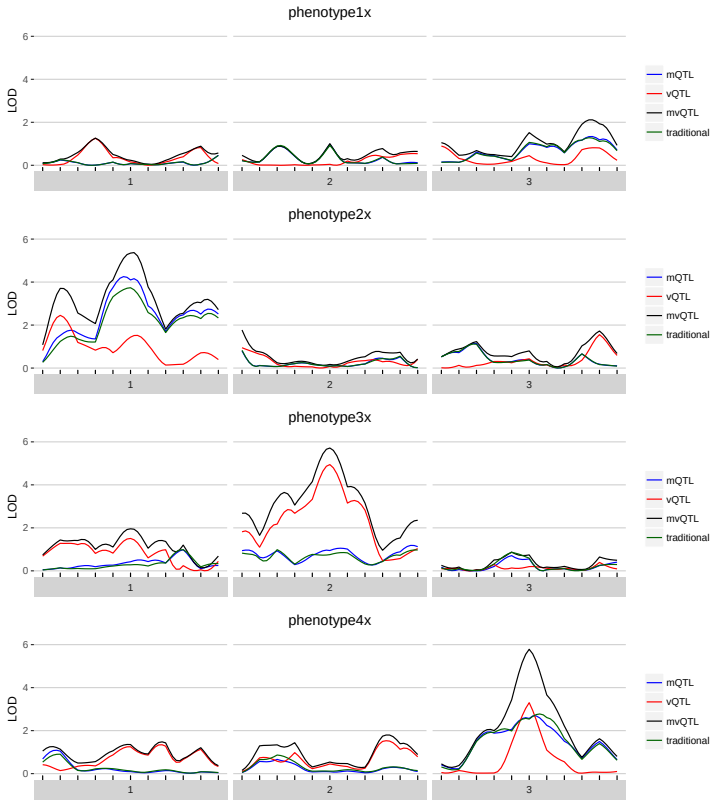

Figure 4 For each of the four simulated phenotypes, the genome scan shows the LOD score of each test – mean, variance, and joint – in blue, red, and black, respectively. The traditional test is in green and globally similar to the mean test.

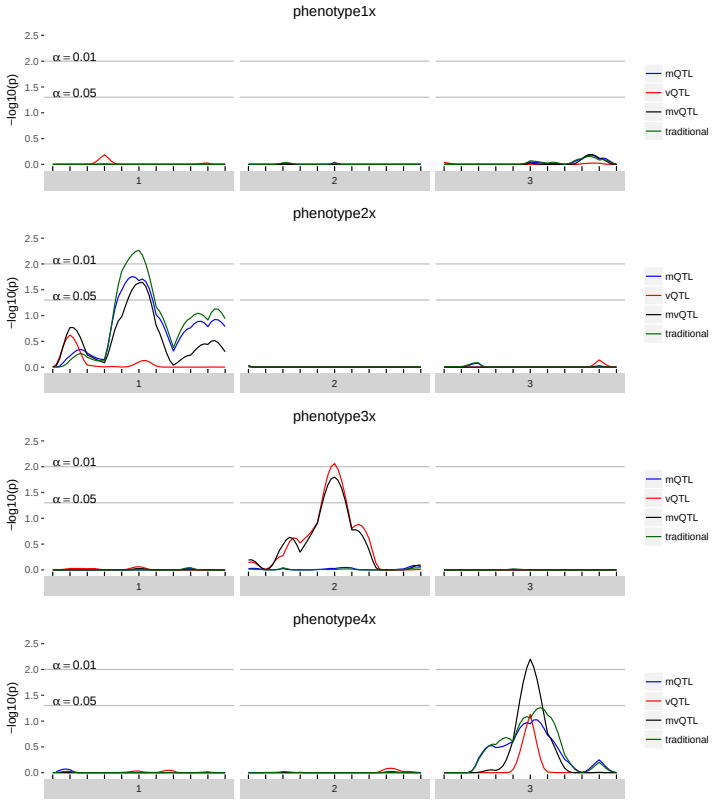

Figure 5 For each of the four simulated phenotypes, the genome scan shows the -log10 of the FWER-corrected *p*-value of each test – mean, variance, and joint – in blue, red, and black, respectively. Thus, a value of 3 implies that the quantity of evidence against the null is such that we expect to see this much or more evidence once per thousand genome scans when there is no true effect.

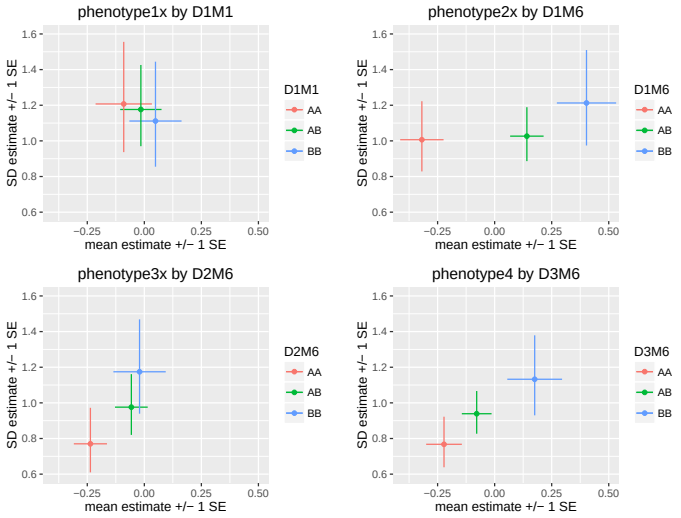

Figure 6 `mean_var_plots` show the estimated genotype effects at a locus, with mean effects on the horizontal axis and variance effects on the vertical axis. Horizontal lines indicate standard errors for mean effects and vertical lines indicate standard errors for variance effects. More information on the interpretation of these plots is provided in the 'Investigate Significant Findings' section.