INVESTIGATIONS

vqtl: An R package for Mean-Variance QTL Mapping

Robert W. Corty[*,1] **and William Valdar**[*,1]
[*]University of North Carolina at Chapel Hill, Department of Genetics

ABSTRACT We present vqtl, an R package for mean-variance QTL mapping. This QTL mapping approach tests for genetic loci that influence the mean of the phenotype, termed mean QTL, the variance of the phenotype, termed variance QTL, or some combination of the two, termed mean-variance QTL. It is unique in its ability to correct for variance heterogeneity arising not only from the QTL itself but also from nuisance factors, such as sex, batch, or housing. This package provides functions to conduct genome scans, run permutations to assess the statistical significance, and make informative plots to communicate results. Because it is inter-operable with the popular qtl package and uses many of the same data structures and input patterns, it will be straightforward for geneticists to analyze future experiments with vqtl as well as re-analyze past experiments, possibly discovering new QTL.

KEYWORDS
QTL mapping, vQTL, mvQTL, variance heterogeneity, DGLM, R

INTRODUCTION

QTL mapping studies in experimental crosses have provided important insights on nearly every trait of interest in human health and disease. Advances in model organism genotyping (Williams *et al.* 1990) and phenotyping (Yang *et al.* 2014) as well as in statistical methods (Lander and Botstein 1989; Martínez and Curnow 1992) and software tools (*e.g.*, Broman *et al.* 2003; Mulligan *et al.* 2017) have supported these discoveries.

Traditional analyses have focused on discovering "mean QTL", regions of the genome where allelic variation drives heterogeneity of phenotype mean, while assuming that the residual variance, that is, the intrinsic stability or noisiness of the phenotype, is identical for each every individual in the mapping population. Recent work, however, has challenged the assumption of homogeneous variance and sought to identify an additional type of QTL — those that influence the extent of residual variance, sometimes termed "variance QTL" (vQTL) (Paré *et al.* 2010; Rönnegård and Valdar 2011, 2012; Cao *et al.* 2014).

Although detection of vQTL has started to enter the mainstream of genetic analysis (Yang *et al.* 2012; Hulse and Cai 2013; Ayroles *et al.* 2015; Wei *et al.* 2016; Wang and Payseur 2017; Wei *et al.* 2017), the statistical methods used for this purpose remain heterogeneous. We developed a standardized method for QTL mapping that can detect mQTL, vQTL, and a generalization of the two that we term "mvQTL". This approach, which we term "mean-variance QTL mapping", can be applied to intercross and backcross designs. In two companion articles, we characterize this method and competitors in the setting where a background factor drives variance heterogeneity (Corty and Valdar 2018+ [BVH]), and use it to discover new QTL from two existing data resources (Corty *et al* 2018+)

Here, we provide a practical guide to using the R package vqtl, which implements mean-variance QTL mapping. First, to generate illustrative data, we simulate an F2 intercross and four phenotypes: one phenotype determined entirely by random noise, and one with each of the three kinds of QTL. On each phenotype we then conduct a genome scan using standard approximations to interval mapping (Lander and Botstein 1989; Martínez and Curnow 1992), and mean-variance QTL mapping, which includes a test for mQTL, a test for vQTL, and a test for mvQTL. The association statistics of all four tests are then initially plotted in LOD score units, with drawbacks of this plotting unit discussed. Permutation scans are used to determine empirically adjusted p-values, and plotting in these units is shown to to make the results of the four tests more comparable. Last, we describe plots to communicate the effects that led to the detection of a QTL, and use the bootstrap to estimate its confidence interval.

EXAMPLE DATA: SIMULATED F2 INTERCROSS AND FOUR TRAITS

To illustrate the use of the vqtl package, we first simulated an example F2 intercross using the popular R/qtl package (Broman *et al.* 2003), on which vqtl is based. 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.

Copyright © 2018 by the Genetics Society of America
Manuscript compiled: Thursday 1st March, 2018
[1]Correspondence e-mail: william.valdar@unc.edu

We then generated four phenotypes:

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

2. `phenotype2` has an mQTL at the center of chromosome one.

3. `phenotype3` has a vQTL at the center of chromosome two.

4. `phenotype4` has an mvQTL at the center of chromosome three.

We additionally consider `phenotype1x` through `phenotype4x`, which have the same type of genetic effects as `phenotype1` through `phenotype4`, but have the additional feature that females have greater residual variance than males. All the same analyses and plots that are shown for `phenotype1` through `phenotype4` are shown for `phenotype1x` through `phenotype4x` in the appendix.

SCAN THE GENOME

The central function for genetic mapping in package `R/qtl` is `scanone` (Broman *et al.* 2003). Analogously, the central function for mean-variance QTL mapping in package `vqtl` is `scanonevar`, building on an early version of `scanonevar` in package `qtl`. It takes three required inputs:

1. `cross` is an object that contains the genetic and phenotypic information from an experimental cross, as defined in package `qtl`.

2. `mean.formula` is a two-sided formula, specifying the phenotype to be mapped, the covariates to be corrected for, and the QTL terms to be fitted, with keywords `mean.QTL.add` and `mean.QTL.dom`

3. `var.formula` is a one-sided formula, specifying the variance covariates to be corrected for as well as the QTL terms to be fitted, using keywords `var.QTL.add` and `var.QTL.dom`.

For example, to scan a phenotype named `p1`, we run:

```
scanonevar(
 cross = test_cross,
 mean.formula = p1 ~ sex + mean.QTL.add + mean.QTL.dom,
 var.formula = ~ sex + var.QTL.add + var.QTL.dom
)
```

At each locus in turn, this function tests for the presence of an mQTL, a vQTL, and an mvQTL. The basis of these tests is a comparison between the fit of an alternative model of the form

$$\text{mean} = \text{covariate effects} + \text{locus effects}$$
$$\log(\text{variance}) = \text{covariate effects} + \text{locus effects}$$

with a null model that omits specific terms: for the mQTL test, the null model omits locus effects on phenotype mean; for the vQTL test, the null omits the locus effects on phenotype variance; and for the mvQTL test, the null omits locus effects on both mean and variance. (Note that the mQTL test in mean-variance QTL mapping is different from the traditional test: the traditional test does not have variance predictors of any kind in either null or alternative models.)

LOD scores and nominal p-values

Each type of test (mQTL, vQTL, and mvQTL) yields two association statistics: the LOD score, and the (nominal) p-value. The LOD is a raw measure of association equal to the base 10 logarithm of the likelihood ratio (LR) between the fitted alternative and null models. Higher values indicate greater association when considered across loci for the same type of test; but LOD scores between different types of tests, namely between mvQTL test vs either mQTL or vQTL tests, are not readily comparable. The p-value, which is comparable between different types of tests, transforms the LOD score to take account of the number of parameters being fit: it is calculated from the asymptotic distribution of $2\log_e(\text{LR})$ under the null model, namely the χ^2 distribution with degrees of freedom equal to the difference in the number of parameters between the alternative and null models. With this example dataset, it takes five seconds to run one genome scan on a Intel Core i5.

The p-values described above, however, are nominal: they do not take into account multiple testing across the genome. They also rely on asymptotic theory that assumes the underlying phenotype being residually normal; this may not always be the case and when violated will lead to inflated significance. More robust p-values that are corrected for genomewide significance via control of the family-wise error rate (FWER) can be obtained empirically, through a permutation procedure described below.

Robust, genomewide-adjusted p-values

To calculate the empirical, FWER-controlled p-value of each test at each locus we advocate use of a permutation procedure [Corty and Valdar 2018+ [BVH]]. Like previous work on permutation-based thresholds for genetic mapping (Churchill and Doerge 1994; Carlborg and Andersson 2002), this procedure sidesteps the need to explicitly estimate the effective number of tests.

In brief, this approach involves conducting many genomes scans on pseudo-null data generated through permutation to maintain as much of the character of the data as possible, while breaking the tested phenotype-genotype association. Specifically, the design matrix of the QTL is permuted in the mean portion of the mQTL alternative model, the variance portion of the vQTL alternative model, and in both portions of the mvQTL alternative model.

For each test (mQTL, vQTL, and mvQTL), the highest observed test statistic is extracted from each permutation scan and the collection of statistics that results is used to fit a generalized extreme value (GEV) density (Stephenson 2002; Dudbridge and Koeleman 2004; Valdar *et al.* 2006). 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 two additional pieces of information: an empirical p-value for each test at each locus and the per-permutation maxima that were used to calculate those p-values. These FWER-corrected p-values are straightforwardly interpretable: $p = 0.05$ 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 in this test.

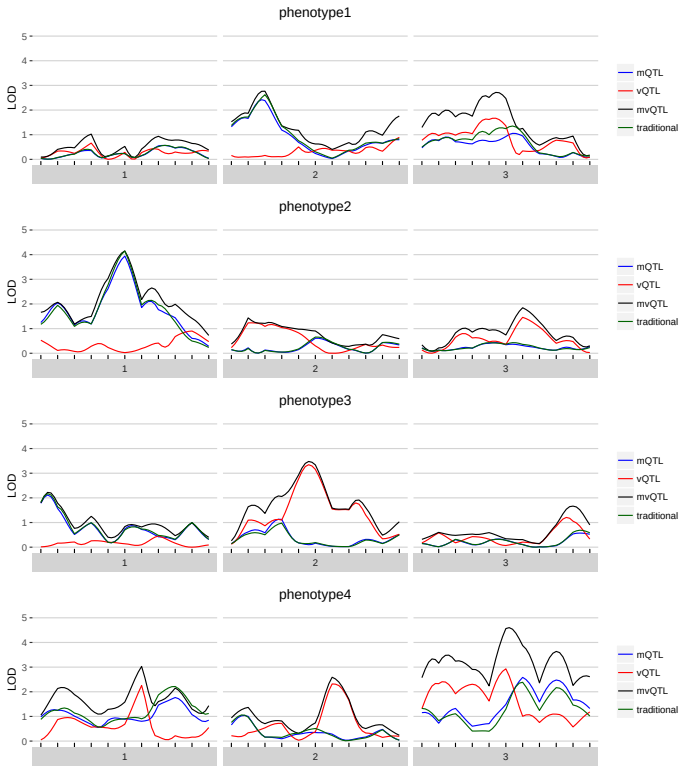

Figure 1 For each of the four simulated phenotypes, the genome scan shows the LOD score of each test — mQTL, vQTL, and mvQTL — in blue, red, and black, respectively. The traditional test is in green and globally similar to the mQTL test.

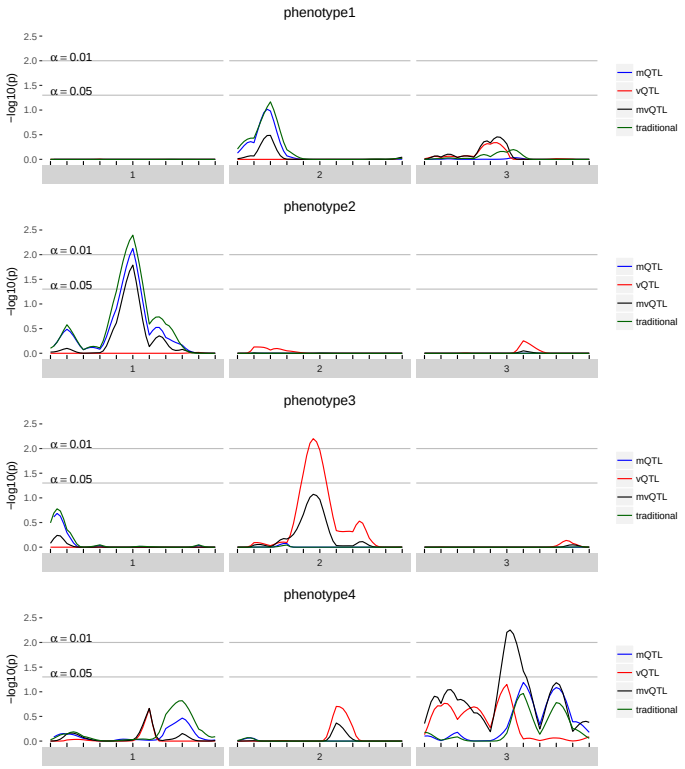

Figure 2 For each of the four simulated phenotypes, the genome scan shows the -log10 of the FWER-corrected *p*-value for each test — mQTL, vQTL, and mvQTL — in blue, red, and black, respectively. The traditional test is in green and globally similar to the mQTL test. A value of 2 implies that the quantity of evidence against the null is such that we expect to see this much or more evidence once per hundred phenotypes no QTL.

Accurate estimation of the FWER-controlled *p*-values requires many permutation scans: traditionally recommended is 1,000 (*e.g.*, Churchill and Doerge 1994; Carlborg and Andersson 2002), although the efficiency gain of using the GEV rather than raw quantiles means that fewer may be adequate in practice (Valdar *et al.* 2006). These permutation scans can be run on multiple processors by specifying the optional `n.cores` argument, which defaults to the total number of cores on the computer minus 2. On an Intel Core i5, running 100 permutations on this dataset takes 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-estimates the extreme value density, re-evaluates the observed LOD scores in the context of new extreme value density, and returns a new `scanonevar` object with more precisely estimated empirical *p*-values.

Reporting and plotting genome scans

The results of `scanonevar` can be plotted by calling `plot` on the `scanonevar` output object. This produces a publication-quality figure that shows the association of the phenotype for each location in the genome as different colors for type of test, with y-axis scale being specified by the user, via option `plotting.units` as the LOD (Figure 1), nominal *p*-value, or, provided permutations have been run, empirical, FWER-controlling *p*-value (Figure 2). Of the available y-axis scales, we recommend using the FWER-controlled *p*-values since this scale puts all tests on a level-footing (unlike the LOD), and allows direct identification of genomewide significance and thereby relevance (unlike the nominal *p*-value).

Calling `summary` on the output of `scanonevar` produces a summary of how the scan was conducted and what the results were.

COMMUNICATE SIGNIFICANT FINDINGS

Having identified interesting QTL, we want to visualize the their estimated genetic and covariate effects. Because the `vqtl` package models effects for both mean and variance, existing plotting utilities are not able to display the entirety of the modeling results. To understand and communicate the results of a `vqtl` scan at one particular locus, we developed the `mean_var_plot`. This plot illustrates how the mean sub-model and variance sub-model of the DGLM fit the data at a given locus.

In each `mean_var_plot` in Figure 3, the location of the dot shows the estimated mean and standard deviation of each genotype group, with 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, `D2M6` in the bottom left, and `D3M6` in the bottom

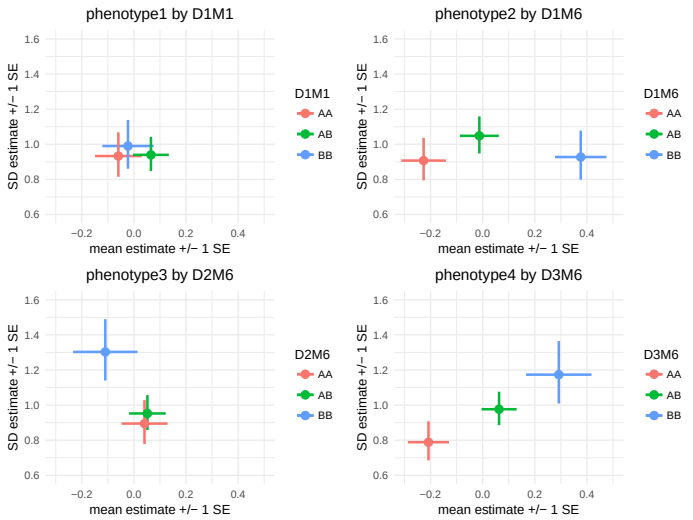

Figure 3 `mean_var_plots` show the estimated genotype effects at a locus with mean effects on the horizontal axis and variance effects on the vertical axis. Horizontal lines indicate standard errors for mean effects and vertical lines indicate standard errors for variance effects. For `phenotype1`, the pattern of overlapping estimates and standard errors 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 pattern of horizontal, but not vertical, separation visually illustrates the identified mQTL. For `phenotype3`, the pattern of vertical, but not horizontal, separation visually illustrates the identified vQTL. For `phenotype4`, the pattern of two-dimensional separation illustrates an mvQTL.

right. (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 influence the phenotype.

Additional plotting utilities, `phenotype_plot`, `effects_plot` and `mean_var_plot_model_free` are described in the online documentation, available on CRAN.

ESTABLISH A CONFIDENCE INTERVAL FOR THE QTL

Last, it is important to assess the genetic precision of a discovered QTL for bioinformatic follow-up. The function `scanonevar.boot` implements the non-parametric bootstrap (Visscher *et al.* 1996). This function takes, as arguments, a `scanonevar` object, the type of QTL detected, the name of the chromosome containing the QTL, and `num.resamples`, the number of bootstrap resamplings desired. As with `scanonevar.perm`, the `n.cores` argument can be used to spread the bootstraps over many computational cores and defaults to the number of cores available minus two, and bootstraps can be run on separate computers and combined with `c` to increase the precision of the estimate of the confidence interval.

We recommend 1000 resamples to establish 80% and 90% confidence intervals. With the datasets simulated here, it takes 20 minutes to run 1000 bootstrap resamples on an Intel core i5.

CONCLUSION

We have demonstrated typical usage of the `R` package `vqtl` for mean-variance QTL mapping in an F2 intercross. This package is appropriate for crosses and phenotypes where genetic factors or covariates or are known or suspected to influence phenotype variance. In the case of genetic factors, they can be mapped, as illustrated in Corty *et al* 2018+. In the case of covariates, they can be accommodated, which can increase power and control false positive rate, as described in Corty and Valdar 2018+ [BVH].

RESOURCES

The scripts used to simulate genotypes and phenotypes, conduct the genome scans, and plot the results are available as a public, static Zenodo repository at DOI:10.5281/zenodo.1173799. The package `vqtl` and its documentation are freely available on `CRAN` at https://CRAN.R-project.org/package=vqtl.

ACKNOWLEDGMENTS

This work was primarily funded by a National Institutes of General Medical Sciences (NIGMS) grant to WV (R01-GM104125) and a National Institutes of Mental Health grant to RWC (F30-MH108265). RWC received additional support from NIGMS grant T32-GM067553 and National Library of Medicine grant T32-LM012420.

LITERATURE CITED

Ayroles, J. F., S. M. Buchanan, C. O'Leary, K. Skutt-Kakaria, J. K. Grenier, *et al.*, 2015 Behavioral idiosyncrasy reveals genetic control of phenotypic variability. Proc. Natl. Acad. Sci. **1**.

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

Cao, Y., P. Wei, M. Bailey, J. S. K. Kauwe, and T. J. Maxwell, 2014 A versatile omnibus test for detecting mean and variance heterogeneity. Genet. Epidemiol. **38**: 51–59.

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.

Dudbridge, F. and B. P. C. Koeleman, 2004 Efficient computation of significance levels for multiple associations in large studies of correlated data, including genomewide association studies. Am. J. Hum. Genet. **75**: 424–35.

Hulse, A. M. and J. J. Cai, 2013 Genetic variants contribute to gene expression variability in humans. Genetics **193**: 95–108.

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

Martínez, O. and R. N. Curnow, 1992 Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. **85**: 480–488.

Mulligan, M. K., K. Mozhui, P. Prins, and R. W. Williams, 2017 Genenetwork: A toolbox for systems genetics. In *Methods Mol. Biol.*, volume 1488, pp. 75–120.

Paré, G., N. R. Cook, P. M. Ridker, and D. I. Chasman, 2010 On the use of variance per genotype as a tool to identify quantitative trait interaction effects: A report from the women's genome health study. PLoS Genet. **6**: 1–10.

Rönnegård, L. and W. Valdar, 2011 Detecting major genetic loci controlling phenotypic variability in experimental crosses. Genetics **188**: 435–447.

Rönnegård, L. and W. Valdar, 2012 Recent developments in statistical methods for detecting genetic loci affecting phenotypic variability. BMC Genet. **13**: 63.

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

Valdar, W., J. Flint, and R. Mott, 2006 Simulating the Collaborative Cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. Genetics **172**: 1783–1797.

Visscher, P. M., R. Thompson, and C. S. Haley, 1996 Confidence intervals in QTL mapping by bootstrapping.

Wang, R. J. and B. A. Payseur, 2017 Genetics of genome-wide recombination rate evolution in mice from an isolated Island. Genetics **206**: 1841–1852.

Wei, W.-H., J. Bowes, D. Plant, S. Viatte, A. Yarwood, *et al.*, 2016 Major histocompatibility complex harbors widespread genotypic variability of non-additive risk of rheumatoid arthritis including epistasis. Sci. Rep. **6**: 25014.

Wei, W.-H., S. Viatte, T. R. Merriman, A. Barton, and J. Worthington, 2017 Genotypic variability based association identifies novel non-additive loci DHCR7 and IRF4 in sero-negative rheumatoid arthritis. Sci. Rep. **7**: 5261.

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, J., R. Loos, M. Goddard, and P. M. Visscher, 2012 FTO genotype is associated with phenotypic variability of body mass index. Nature **490**: 267–272.

Yang, W., Z. Guo, C. Huang, L. Duan, G. Chen, *et al.*, 2014 Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. Nat. Commun. **5**: 5087.

APPENDIX

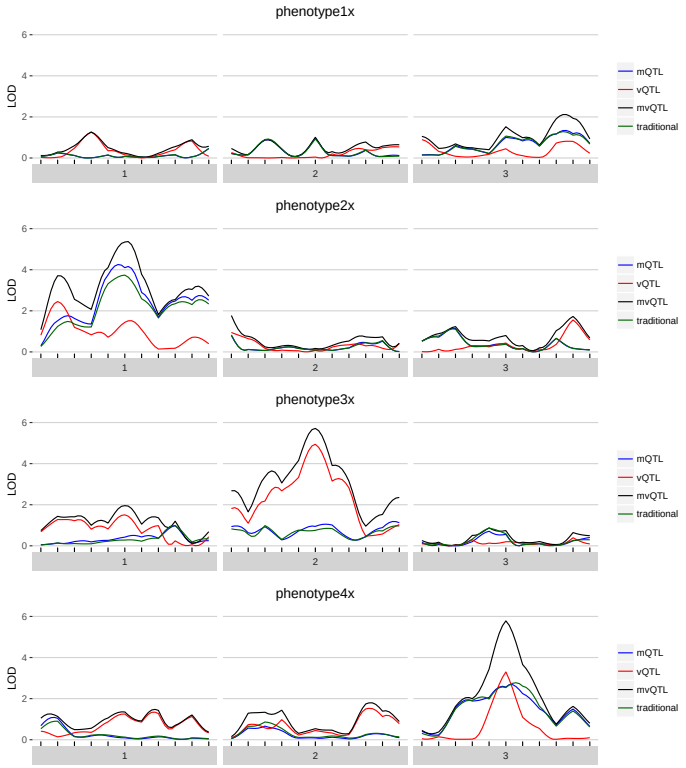

Figure A1 For each of the four simulated phenotypes with background variance heterogeneity, the genome scan shows the LOD score of each test – mean, variance, and joint – in blue, red, and black, respectively. The traditional test is in green and globally similar to the mean test.

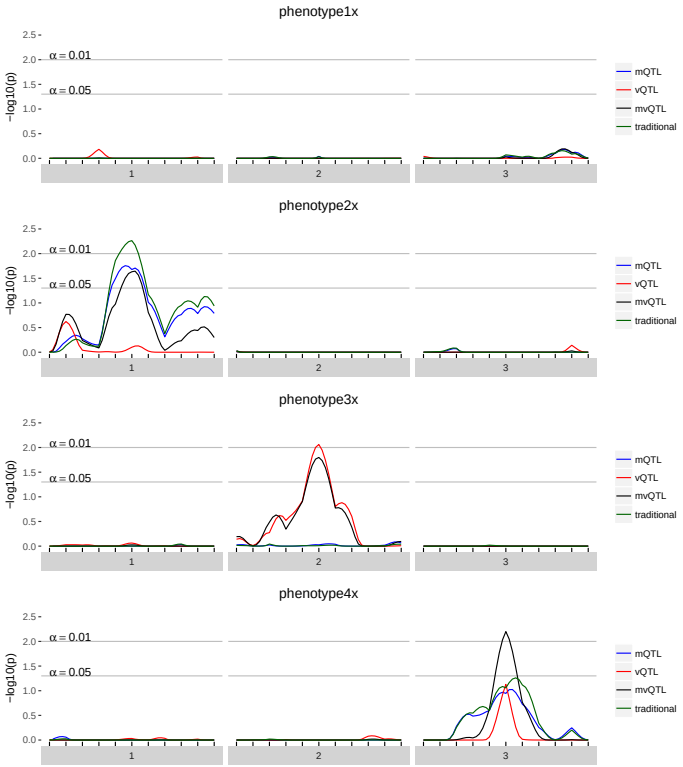

Figure A2 For each of the four simulated phenotypes with background variance heterogeneity, the genome scan shows the -log10 of the FWER-corrected p-value of each test – mean, variance, and joint – in blue, red, and black, respectively. Thus, a value of 3 implies that the quantity of evidence against 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.

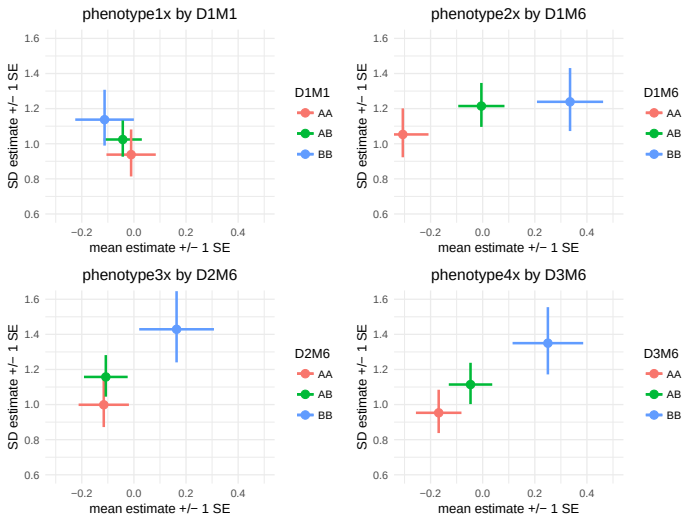

Figure A3 `mean_var_plots` show the estimated genotype effects at a locus, with mean effects on the horizontal axis and variance effects on the vertical axis. Horizontal lines indicate standard errors for mean effects and vertical lines indicate standard errors for variance effects. For `phenotype1x`, the pattern of overlapping estimates and standard errors is consistent with the fact that there are no genetic effects, and the p-value was not statistically significant at any locus. For `phenotype2x`, the pattern of horizontal, but not vertical, separation visually illustrates the identified mQTL on a background of variance heterogeneity. For `phenotype3x`, the pattern of vertical, but not horizontal, separation visually illustrates the identified vQTL on a background of variance heterogeneity. For `phenotype4x`, the pattern of two dimensional separation without either total horizontal or vertical separation illustrates an mvQTL with neither mean nor variance effect strong enough to define an mQTL or vQTL on a background of variance heterogeneity.